



Universidade do Minho
Escola de Engenharia

João Laureano Gonçalves Alves

Sistema de Business Intelligence no
Projeto Educativo de Guimarães

João Laureano Gonçalves Alves
Sistema de Business Intelligence no
Projeto Educativo de Guimarães

UMinho | 2015

outubro de 2015



Universidade do Minho
Escola de Engenharia

João Laureano Gonçalves Alves

Sistema de Business Intelligence no
Projeto Educativo de Guimarães

Dissertação de Mestrado
Ciclo de Estudos Integrados Conducentes ao Grau de
Mestre em Engenharia e Gestão de Sistemas de Informação

Trabalho efectuado sob a orientação do
Professora Doutora Maribel Yasmina Santos

DECLARAÇÃO

Nome: João Laureano Gonçalves Alves

Endereço electrónico: a61578@alunos.uminho.pt

Telefone: 916599169 / 258455052

Número do Bilhete de Identidade: 13938665

Título dissertação: Sistema de Business Intelligence no Projeto Educativo de Guimarães

Orientadora: Professora Doutora Maribel Yasmina Santos

Ano de conclusão: 2015

Designação do Mestrado: Mestrado Integrado em Engenharia e Gestão de Sistemas de Informação

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE/TRABALHO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE;

Universidade do Minho, ___/___/_____

Assinatura: _____

“A vida é a hesitação entre uma exclamação e uma interrogação. Na dúvida, há um ponto final.”
(Fernando Pessoa)

Agradecimentos

Solidão é uma condição interior do homem caracterizada por uma sensação de vazio. Uma dissertação de mestrado pode caracterizar-se como um percurso com vários obstáculos num processo solitário. No entanto, existem pessoas que nos rodeiam que amenizam e ajudam a afastar tais dificuldades e às quais devo imperativamente expressar os meus sinceros agradecimentos.

Em primeiro lugar, quero agradecer á minha orientadora, Professora Doutora Maribel Yasmina Santos, pela sua dedicação, entrega e profissionalismo apurado. Todo o empenho e os preciosos estímulos dados foram, sem dúvida, uma das alavancas mais importantes no processo de desenvolvimento desta dissertação de mestrado.

À Câmara Municipal de Guimarães, e em particular à Doutora Adelina Pinto pela oportunidade concedida, disponibilidade e colaboração nesta investigação. Agradeço também a todos os demais intervenientes em todo o processo de desenvolvimento.

Aos meus grandes amigos e parceiros desta jornada Carlos, Carina, Nuno, Rui Silva e Rui Peixoto. Obrigado por todos os momentos de alegria, solidariedade e entreaajuda, bem como o apoio nos momentos de sofrimento e dificuldade passados durante este percurso.

A todas as pessoas que convivem diariamente comigo e também a todos os meus amigos, principalmente os de infância, no qual eu me orgulho profundamente de os ter conhecido e que foram também responsáveis pela minha chegada até aqui.

Sendo a vida uma rápida passagem, nunca é tarde para se conhecer boas pessoas e maravilhosos seres humanos. Por isso, quero deixar o meu profundo obrigado aos grandes amigos Patricia e João Mendes que foram, a partir de uma parte do meu percurso académico, pessoas sensacionais na qual tive oportunidade de passar momento excecionais que dedicadamente e prontamente se mostraram sempre disponíveis.

Como um velho amigo consegue sempre tornar a saudade num sentimento agradável, quero, pela dedicação, estima, suporte e pelos inigualáveis momentos passados, agradecer á minha amiga de longa data e também “irmã de...” Luciana. Uma boa parte de toda esta caminhada deve-se também a ti.

Ao verdadeiro e fiel amigo de infância João Pedro, no qual as memórias dos nossos momentos não se esgotam e porque estás sempre presente há mais de vinte anos. Obrigado pelo apoio, preocupação e dedicação não só nesta fase, mas também em toda a minha existência como pessoa.

À Cátia a quem considero o grande pilar de todo o meu percurso académico, pois nos momentos mais difíceis e escuros estavas lá com aquela luz para indicar o caminho a seguir. Todo este caminho seria bem mais difícil e penoso sem o teu suporte, portanto quero agradecer todos os segundos que dedicas-te, o apoio incondicional, a paciência e a motivação dada. Grande parte de todo este ciclo que agora termina devo a ti, companheira de todas as horas.

A toda os meus familiares, sem exceção, que tiveram a sua cota de responsabilidade nesta longa caminhada. Obrigado por me ensinarem o que é humildade e pela educação dada.

Aos meus irmãos Pedro e Ana que os vi nascer e crescer com muito orgulho e esperança. Obrigado por me ajudarem a crescer em responsabilidade. Hoje sou eu e amanhã sereis vós.

Como amar os que se encontram longe é por vezes mais fácil, e nem sempre é fácil amar os que estão mesmo ao nosso lado, deixo o meu último agradecimento, e sem dúvida o maior agradecimento de todos, aos meus Pais. Para que fosse possível escrever estas palavras para eles, muitos sacrifícios e dificuldades tiveram de ser passados para que eu tivesse tal oportunidade. Por isso, esta dissertação espelha todo o esforço e sacrifício como forma de agradecimento por tudo neste marco da minha vida.

Simplemente Obrigado!

Resumo

Cada vez mais, a sociedade caminha para uma era dominada pela tecnologia e pela informação. Esta era, ao nível organizacional, caracteriza-se pelo crescimento exponencial da quantidade de dados que as organizações armazenam. Este fator, leva a que as organizações sintam a necessidade de adotar mecanismos que sejam capazes de extrair desses dados, a informação fundamental para o suporte à tomada de decisão e respetivo sucesso do negócio.

A Câmara Municipal de Guimarães, no contexto desta realidade, tem verificado que ao nível da gestão da sua rede escolar surge a necessidade de proceder a alterações que levem a uma adaptação da sua oferta ao nível da rede escolar, como consequência das alterações demográficas da população. A carta educativa é, para cada município, um dos instrumentos de gestão escolar que ajuda a definir planos e ações para colmatar os problemas existentes. Esta carta aglomera toda a informação proveniente das escolas. A par disto, associam-se também as alterações demográficas, com o número de jovens a decrescer, o que ano após ano tem obrigado à mudança de rumo por parte da gestão escolar. Dadas as necessidades anteriormente mencionadas, surge a proposta de conceção de um sistema de *Business Intelligence*, com a inclusão de um *Data Warehouse*, que irá possibilitar à Câmara Municipal de Guimarães o acesso à informação de gestão necessária para a elaboração da carta educativa.

Este documento tem como objetivo mostrar todo o trabalho desenvolvido na proposta e conceção de um protótipo de um sistema de *Business Intelligence* que seja capaz de auxiliar a Câmara Municipal de Guimarães na gestão da sua rede escolar. Este sistema inclui um *Data Warehouse* como repositório integrado dos dados, bem como mecanismos de processamento de dados geo-referenciados e processamento analítico dos mesmos através de *on-line analytical processing* (OLAP). Como resultado, apresenta-se uma plataforma que é capaz de disponibilizar mecanismos analíticos, através de um conjunto de *reports* e *dashboards*, que visam auxiliar a tomada de decisão ao nível da gestão municipal.

Palavras-chave: *Business Intelligence*; *Data Warehouse*; Educação; Carta Educativa; Dados espaciais; *Data Warehouse* espacial

Abstract

Increasingly, society is heading towards an era dominated by technology and information. This era, at the organizational level, is characterized by the exponential growth of the amount of data that organizations store. This factor leads the organizations to feel the need to adopt mechanisms be able to extract from such data, fundamental information to support the process of decision-making and respective business success.

The Municipality of Guimarães, in the context of this reality, has realized that in the management of its school system, there is the need to make changes in order to adapt its offer to the level of the school system, as a consequence of demographic population changes. The educational letter is, for each Municipality, one of the school management tools that helps define actions and plans to address existing problems. This letter integrates information from all the schools. Beside this, and due to the demographic changes, with the number of young people decreasing year after year, infrastructural adaptations need to be done.

Given the above requirements, this work proposes the design and implementation of a prototype for a Business Intelligence system, including a Data Warehouse, which will enable the Municipality of Guimarães to access and manage the necessary information for setting up the education letter.

This document aims to show the work done for the concretization of this Business Intelligent system, which is able to assist the Municipality of Guimarães in the management of its entire school system. The implemented prototype includes a Data Warehouse as an integrated repository of data and spatial data processing using online analytical processing (OLAP). As result, it is presented a platform that is capable of provide analytical mechanisms through a series of reports and dashboards, aimed to assist the process of decision-making at the level of local administration.

Keywords: *Business Intelligence; Data Warehouse; Education, Educational Letter; Spatial Data; Spatial Data Warehouse*

Índice

Agradecimentos	vii
Resumo	ix
<i>Abstract</i>	xi
Lista de Acrônimos e Abreviaturas	xxiii
1. Introdução	1
1.1. Enquadramento e Motivação	1
1.2. Objetivos e Resultados esperados	3
1.3. Metodologia de investigação.....	4
1.4. Estrutura do documento.....	6
2. Enquadramento Concetual.....	9
2.1. <i>Business Intelligence</i>	9
2.1.1. O conceito.....	9
2.1.2. Componentes de um Sistema de <i>Business Intelligence</i>	11
2.1.3. <i>Business Intelligence</i> na Área do Ensino	14
2.2. Sistemas de <i>Data Warehousing</i>	20
2.2.1. Definição.....	20
2.2.2. Dimensões e Factos	23
2.2.3. Modelação Multidimensional	28
2.2.4. Tipos de Servidores OLAP.....	32
2.2.5. Operações para Analisar um Cubo	33
2.2.6. Metodologia de Sistemas de <i>Data Warehousing</i>	35
2.2.7. Arquiteturas e Componentes de um Sistema de DW	37
2.2.8. Análise de Dados Espaciais	42
2.3. Carta Educativa	44
3. Caracterização das Tecnologias e dos Dados.....	51
3.1. Tecnologias a utilizar.....	51

3.2.	Proposta de Arquitetura do Sistema	56
3.3.	Caracterização e Análise Exploratória das Fontes de Dados	58
3.3.1.	Caracterização dos Dados Necessários para o Estudo.....	58
3.3.2.	Fonte de Dados Software JPM & Abreu.....	60
3.3.3.	Fonte de Dados Plataforma Edubox	65
4.	Implementação do Sistema de <i>Business Intelligence</i>	69
4.1.	Sistema de <i>Data Warehousing</i>	69
4.1.1.	Modelo de Dados do <i>Data Warehouse</i>	69
4.1.2.	Processo ETL	93
4.1.3.	Processamento Analítico de Dados	103
4.1.4.	Aplicações de <i>Front-End</i>	105
5.	Conclusão	113
5.1.	Síntese do Trabalho Realizado.....	113
5.2.	Dificuldades e Limitações.....	115
5.3.	Trabalho Futuro	116
	Referências Bibliográficas.....	119
	Anexos.....	123
	Anexo A – Atributos das Fontes de Dados.....	123
	A1. Fonte de Dados Software JPM & Abreu (AlunosGeralTotal.csv).....	123
	A2. Fonte de Dados Plataforma SIGA - Edubox	126
	Anexo B – <i>Scripts</i>	128
	B1. Criação da Tabela DimTempo.....	128
	B2. Script de Criação Base de Dados de Estágio	129
	B3. Script de Criação do <i>Data Warehouse</i>	130
	B4. Script Java Converter em Coordenadas Geográficas	138
	Anexo C – <i>Dashboards</i> das Aplicações <i>Front-End</i>	140
	C1. Vista Geral da <i>Dashboard</i> Caracterização Demográfica.....	140

C2. Vista Geral da <i>Dashboard</i> Enquadramento Geográfico	141
C3. Vista Geral da <i>Dashboard</i> Sistema Educativo.....	142
Anexo D – Relatórios nas Aplicações <i>Front-End</i>	144
D1. Vista Geral do Relatório Apoio Social Escolar	144
D2. Vista Geral do Relatório Aproveitamento Escolar.....	144
D3. Vista Geral do Relatório Caracterização por Agrupamento.....	145
D4. Vista Geral do Relatório Caracterização Económica	145
D5. Vista Geral do Relatório Caracterização por Escola	146
D6. Vista Geral do Relatório Densidade Populacional	146
D7. Vista Geral do Relatório Evolução Etária da População.....	147
D8. Vista Geral do Relatório Caracterização da População Estudantil	148
D9. Vista Geral do Relatório Geografia do Concelho	149
D11. Vista Geral do Relatório Hierarquia dos Aglomerados	150
D12. Vista Geral do Relatório Mortalidade/Natalidade.....	151
D13. Vista Geral do Relatório Setores de Atividade.....	152

Índice de Tabelas

Tabela 1 - Matriz de conceitos de <i>Business Intelligence</i> na área do Ensino	19
Tabela 2 - Bases de dados operacionais vs. <i>Data Warehouse</i> (adaptado de (Santos & Ramos, 2006))	21
Tabela 3 - Tabela de Factos.....	26
Tabela 4 - Tabela Dimensão: <i>Costumer</i>	26
Tabela 5 - Exemplo de tipo/fonte dos dados (adaptado de (Negash, 2004))	38
Tabela 6 - Intervenientes no processo de elaboração da Carta Educativa (Adaptado de(Oliveira et al., 2000))	48
Tabela 7 - Matriz das Tecnologias seleccionadas	55
Tabela 8 - Caracterização dos dados necessários.....	59
Tabela 9 - Conjuntos de dados extraídas do Software J.P.M & Abreu	61
Tabela 10 - Atributos da fonte de dados do Software J.P.M & Abreu	62
Tabela 11 - Atributos da fonte de dados da plataforma SIGA da Edubox	66
Tabela 12 - Tabela de factos FactNatalidadeMortalidade	72
Tabela 13 - Tabela de Factos FactAlunoEscola.....	74
Tabela 14 - Tabela de factos FactEnqGeograficoDemografico (a)	76
Tabela 15 - Tabela de factos FactEnqGeograficoDemografico (b)	77
Tabela 16 - Tabela de factos FactEnqGeograficoDemografico (c)	78
Tabela 17 - Tabela de factos FactSistemaEducativo	80
Tabela 18 - Tabela de factos FactCaractClimatica	81
Tabela 19 - Tabela de factos FactAnaliseNotas	83
Tabela 20 - Tabela de dimensão DimTempo	84
Tabela 21 - Tabela de dimensão DimTempoAno	84
Tabela 22 - Tabela de dimensão DimLocalizacao	85
Tabela 23 - Tabela de dimensão DimCurso.....	86
Tabela 24 - Tabela de dimensão DimBridgeCurso	86
Tabela 25 - Tabela de dimensão DimDisciplina	86
Tabela 26 - Tabela de dimensão DimBridgeDisciplina	87
Tabela 27 - Tabela de dimensão DimAgrupamento	87
Tabela 28 - Tabela de dimensão DimEscola.....	88
Tabela 29 - Tabela de dimensão DimAluno	89
Tabela 30 - Tabela de dimensão DimProfessor	89
Tabela 31 - Tabela de dimensão DimBridgeProf.....	89

Tabela 32 - Tabela de dimensão DimNaoDocente	90
Tabela 33 - Tabela de dimensão DimBridgeNaoDocente	90
Tabela 34 - Tabela de dimensão DimNivelEnsino	90
Tabela 35 - Tabela de dimensão DimNota	91
Tabela 36 - Tabela de dimensão DimPeriodo	91
Tabela 37 - Tabela de dimensão DimSituacao.....	91
Tabela 38 - Tabela de dimensão DimAnoEscolar	92
Tabela 39 - Tabela de dimensão DimEscalao	92
Tabela 40 - Tabela de dimensão DimTurma.....	92
Tabela 41 – Matriz de relação do estado dos indicadores (a)	111
Tabela 42 – Matriz de relação do estado dos indicadores (b)	112
Tabela 43 - Atributos da fonte de dados JPM & Abreu (a).....	123
Tabela 44 - Atributos da fonte de dados JPM & Abreu (b).....	124
Tabela 45 - Atributos da fonte de dados JPM & Abreu (c).....	125
Tabela 46 - Atributos da fonte de dados JPM & Abreu (d).....	126
Tabela 47 - Atributos da fonte de dados Edubox (a)	126
Tabela 48 - A tributos da fonte de dados Edubox (b)	127

Índice de Figuras

Figura 1 - Atividades do processo de Design Science Research	5
Figura 2 - Relação do BI com outros Sistemas de Informação (retirado de (Negash, 2004))	10
Figura 3 - Arquitetura de Alto Nível de um Sistema de BI (Adaptada de (Turban, Sharda, Aronson, & King, 2008))	13
Figura 4 - Arquitetura do sistema de DW da Universidade de Novi Sade (retirado de (Kazi et al., 2012))	15
Figura 5 - Modelo de <i>Data Mining</i> e Sistema de <i>Data Warehouse</i> (retirado de (Kurniawan & Halim, 2013))	15
Figura 6 - Arquitetura do SBIAES (retirado de (Nobre et al., 2014))	16
Figura 7 - Arquitetura do <i>Data Warehouse</i> (adaptado de ((Mickael Santos, 2014)))	18
Figura 8 - Cubo com 3 dimensões (retirado de (Rainardi, 2008))	23
Figura 9 - Hierarquia dos dados de 3 dimensões (retirado de (Rainardi, 2008))	28
Figura 10 - Esquema em Estrela (Moody & Kortink, 2003)	30
Figura 11 - Esquema em Floco de Neve (Moody & Kortink, 2003)	31
Figura 12 - Esquema em Constelação (adaptado de (Santos & Ramos, 2006))	32
Figura 13 - <i>Slice and dice</i> (Rainardi, 2008)	34
Figura 14 – Pivô (adaptado de (Santos & Ramos, 2006))	34
Figura 15 - Metodologia de Kimball (retirado de (Ralph Kimball & Ross, 2003))	36
Figura 16 - <i>Data Warehouse Framework</i> (adaptado de (Turban et al., 2008))	38
Figura 17 - Arquitetura de apoio ao <i>Business Intelligence</i> (retirado de (Santos & Ramos, 2006))	41
Figura 18 - Organograma do Sistema Educativo em Portugal (retirado de (M. Educação, 2007))	45
Figura 19 – Quadrante Mágico de plataformas de Análise e <i>Business Intelligence</i> (retirado de (Gartner, 2014))	54
Figura 20 - Arquitetura da Infraestrutura Tecnológica	56
Figura 21 - Registos do atributo N_BI	63
Figura 22 - Qualidade do atributo Nome	63
Figura 23 - Exemplo do Relatório de qualidade de dados - coluna TextoProfissaoPai	64
Figura 24 – Amostra dos Registos Duplicados	64
Figura 25 - Amostra dos dados extraídos da plataforma SIGA	67
Figura 26 - Análise aos valores do atributo Sexo	67
Figura 27 - Modelo Multidimensional	70
Figura 28 - Modelo de Dados da tabela FactNatalidadeMortalidade	71

Figura 29 - Modelo de dados da Tabela FactAlunoEscola	73
Figura 30 - Modelo de dados da Tabela FactEnqGeograficoDemografico.....	75
Figura 31 - Modelo de dados da tabela FactSistemaEducativo.....	79
Figura 32 - Modelo de dados da tabela FactCaractClimatica.....	80
Figura 33 - Modelo de dados da tabela FactAnaliseNota.....	82
Figura 34 - Metodologia do processo ETL.....	93
Figura 35 – Perspetiva Geral do Processo ETL implementado	94
Figura 36 - Componentes de criação das bases de dados	95
Figura 37 - Fluxo de Invocação e Extração dos Dados	95
Figura 38 - Carregamento das Tabelas de Dimensão	96
Figura 39 - Carregamento da dimensão Localização	97
Figura 40 - Exemplo do carregamento de uma Tabela de Dimensão.....	98
Figura 41 - Fluxo ETL da tabela de factos FactEnqGeograficoDemografico	99
Figura 42 - Fluxo ETL da tabela de factos FactAlunoEscola.....	100
Figura 43 - Parte do Fluxo que transforma os factos da FactAlunoEscola	101
Figura 44 - Fluxo ETL para transformação dos factos da FactMortalidadeNatalidade.....	102
Figura 45 - Hierarquia da tabela de dimensão DimTempo.....	104
Figura 46 - Hierarquia da tabela de dimensão DimLocalizacao.....	104
Figura 47 - Interface do Sistema de <i>Business Intelligence</i> Implementado	107
Figura 48 - Aplicação Desktop da ferramenta <i>PowerBI</i>	108
Figura 49 – Vista da <i>Dashboard</i> “Sistema Educativo” da plataforma <i>PowerBI</i>	108
Figura 50 - Relatório do Aproveitamento Escolar do Concelho	109
Figura 51 - Vistas das Dashboards na Aplicação Mobile	110
Figura 52 - Script de Criação da tabela DimTempo	128
Figura 53 - <i>Dashboard</i> Caracterização Demográfica (<i>PowerBI</i>)	140
Figura 54 - <i>Dashboard</i> Caracterização Demográfica 1 (Excel 2013).....	140
Figura 55 - <i>Dashboard</i> Caracterização Demográfica 2 (Excel 2013).....	141
Figura 56 - <i>Dashboard</i> Caracterização Demográfica 3 (Excel 2013).....	141
Figura 57 - <i>Dashboard</i> Enquadramento Geográfico	141
Figura 58 - <i>Dashboard</i> Enquadramento Geográfico 1 (Excel 2013).....	142
Figura 59 - <i>Dashboard</i> Enquadramento Geográfico 2 (Excel 2013).....	142
Figura 60 - <i>Dashboard</i> Sistema Educativo.....	142
Figura 61 - <i>Dashboard</i> Sistema Educativo 1 (Excel 2013)	143

Figura 62 - <i>Dashboard</i> Sistema Educativo 2 (Excel 2013)	143
Figura 63 - <i>Dashboard</i> Sistema Educativo 3 (Excel 2013)	143
Figura 64 - Relatório Apoio Social Escolar	144
Figura 65 - Relatório Aproveitamento Escolar	144
Figura 66 - Relatório Caracterização por Agrupamento	145
Figura 67 - Relatório Caracterização Económica.....	145
Figura 68 - Caracterização por Escola.....	146
Figura 69 - Relatório Densidade Populacional	146
Figura 70 - Relatório Evolução Etária da População 1	147
Figura 71 - Relatório Evolução Etária da População 2.....	147
Figura 72 - População Residente (Excel 2013)	148
Figura 73 - Relatório Evolução da População Estudantil 1.....	148
Figura 74 - Relatório Evolução da População Estudantil 2.....	149
Figura 75 - Relatório Geografia do Concelho	149
Figura 76 - Relatório Hierarquia dos Aglomerados.....	150
Figura 77 - Hierarquia dos Aglomerados (Excel 2013).....	150
Figura 78 - Relatório Mortalidade/Natalidade 1	151
Figura 79 - Mortalidade/Natalidade (Excel 2013).....	151
Figura 80 - Relatório Mortalidade/Natalidade 2.....	152
Figura 81 - Relatório Setores de Atividade	152

Lista de Acrônimos e Abreviaturas

CMG – Câmara Municipal de Guimarães

CET – Curso de Especialização Tecnológica

CSV – *Comma-Separated Values*

CTT- Correios e Telecomunicações de Portugal

BI – *Business Intelligence*

DW – *Data Warehouse*

ETL – *Extract, Transform, Load*

TI – Tecnologias de Informação

SI – Sistema de Informação

DSR - *Design Science Research*

DSRM - *Design Science Research Methodology*

DSA - *Data Staging Area*

MDB – *Multidimensional Data Bases*

OLTP - *Online Transaction Processing*

OLAP - *On-line Analytical Processing*

ROLAP – *Relational OLAP*

MOLAP – *Multidimensional OLAP*

HOLAP – *Hybrid OLAP*

SOLAP – *Spatial OLAP*

SAD – Sistemas de Apoio à Decisão

CRM - *Customer Relationship Management*

KPI - *Key Performance Indicators*

SQL - *Structured Query Language*

PK – *Primary Key*

FK – *Foreign Key*

SCD – *Slowly Changing Dimension*

IPC – Instituto Politécnico de Coimbra

SIG – Sistema de Informação Geográfica

SBIAES - Sistema de *Business Intelligence* para análise do Acesso ao Ensino Superior

DGES – Direção Geral de Ensino Superior

IBM – *International Business Machines*

ID – Identidade

ERP - *Enterprise Resource Planning*

POS - *Point of Sale*

URL - *Uniform Resource Locator*

XML - *Extensible Markup Language*

WSDL - *Web Service Description Language*

API – *Application Programming Interface*

1. Introdução

Neste primeiro capítulo do documento é elaborado um enquadramento dos temas abordados no âmbito desta dissertação. Menciona-se também a motivação, finalidade e objetivos propostos para a realização da mesma.

Na segunda parte deste capítulo será descrita a metodologia de investigação utilizada, suportando o trabalho de investigação desenvolvido, bem como a descrição da estrutura do documento.

1.1. Enquadramento e Motivação

Vivemos numa era em que a tecnologia e a informação estão cada vez mais presentes na vida dos cidadãos e das organizações, o que se traduz em mais e melhor conhecimento, quer ao nível pessoal como organizacional.

Hoje em dia, as organizações deparam-se com desafios ao nível da análise assertiva e atempada dos dados que estas geram ou possuem, como forma de produzir informação com melhor conteúdo que seja capaz de suportar os processos organizacionais.

Qualquer organização, para ser competitiva e alcançar um crescimento continuado, necessita de lidar com grandes volumes de informação. As bases de dados organizacionais são ótimas fontes de dados, que sendo devidamente exploradas podem proporcionar inúmeras vantagens competitivas.

Uma das primeiras dificuldades na obtenção de informação de qualidade prende-se com o estado das bases de dados organizacionais, pois estas apresentam graus de diversidade por vezes elevados e com uma grande falta de homogeneidade. Estes problemas são motivados, na maioria das vezes, pela falta de eficácia dos processos de desenvolvimento de *software*, pela falta de planeamento, estratégias na recolha e armazenamento de informação mal definidas e pelo défice ou não integração com os processos organizacionais. O elevado grau de diversidade e a falta de homogeneidade, referidos anteriormente, estão associados à existência de diversas bases de dados organizacionais. Estes problemas podem ser minimizados se, com base em critérios de seleção, extração, transformação e integração, os dados necessários para a tomada de decisão estiverem armazenados num único sistema de dados homogéneo.

Ao nível social, as alterações demográficas têm sido evidentes ao longo dos últimos anos, verificando-se cada vez mais uma população envelhecida e com uma diminuição contínua das taxas de natalidade de ano para ano. Portugal tem registado um crescimento da população segundo um ritmo fraco a tender para zero (Carrilho & Patricio, 2010).

À medida que estas alterações na demografia se vão acentuando cada vez mais, com o número de jovens a decrescer ano após ano, os desafios passam pela correta caracterização de cada região para melhor fundamentar as decisões daqueles que necessitam de as tomar.

Ao nível da gestão da rede escolar por parte dos municípios, surge também evidenciada a necessidade de se proceder a alterações que conduzam a uma adaptação das necessidades populacionais.

A carta educativa é, a nível municipal, um instrumento de planeamento e ordenamento prospetivo de edifícios e equipamentos educativos a localizar no concelho, de acordo com as ofertas de educação e formação que sejam necessárias satisfazer, tendo em vista a melhor utilização dos recursos educativos, no quadro de desenvolvimento demográfico e socioeconómico de cada município (Público, 2003).

Todos estes fatores levaram a que a Câmara Municipal de Guimarães sentisse a necessidade de utilizar um sistema de apoio à decisão, suportando os que no seu dia-a-dia necessitam de atuar e basear-se em indicadores precisos com um conhecimento aprofundado da realidade de cada região do concelho de Guimarães.

Sendo esta uma dissertação de mestrado que visa lidar com informação proveniente das escolas do concelho de Guimarães, esta será recolhida, armazenada e analisada num sistema de *Business Intelligence* a ser disponibilizado à Câmara Municipal de Guimarães, garantindo a proposta de um trabalho inovador e útil na gestão da rede escolar em Portugal. Neste aspeto, e no que diz respeito ao conteúdo científico, importa garantir que, de facto, existe uma problemática que carece de uma solução.

1.2. Objetivos e Resultados esperados

Esta dissertação tem como finalidade a conceção e implementação de um protótipo de um sistema de *Business Intelligence* que esteja preparado para receber dados provenientes de diversas fontes, integrando-os para que posteriormente sejam disponibilizados mecanismos de análise adequados. Este sistema deve incluir um *Data Warehouse* como repositório integrado dos dados, bem como mecanismos de processamento geo-espacial e processamento analítico dos mesmos através de cubos OLAP, por forma a suportar a tomada de decisão na gestão de toda a rede escolar da Câmara Municipal de Guimarães.

Como resultado final, prevê-se a criação de uma plataforma que seja capaz de disponibilizar um conjunto de mecanismos analíticos que visam substituir a carta educativa que atualmente se encontra disponível em papel. De futuro, e depois do protótipo passar a aplicação final, pretende-se a integração deste trabalho com o Website da Câmara Municipal de Guimarães. O objetivo é que a visualização e o acesso aos dados seja facilitado a qualquer utilizador, que a qualquer momento necessite de suportar a tomada de decisão com informação atualizada.

Para alcançar esta finalidade com sucesso, é de todo fundamental traçar e cumprir com um conjunto de objetivos:

- Rever a literatura associada ao estado da arte nesta área, de forma a conhecer as melhores práticas e os últimos desenvolvimentos neste âmbito;
- Identificar as fontes de dados necessárias;
- Caracterizar e analisar os dados disponíveis;
- Caracterizar quais os focos de análise, bem como as questões a que se pretende que o sistema responda;
- Definir a infraestrutura tecnológica adequada ao contexto organizacional;
- Definir o modelo de dados para o *Data Warehouse*;
- Implementar o *Data Warehouse* e todos os processos necessários à extração, transformação e carregamento dos dados;
- Concretizar os mecanismos analíticos através de cubos OLAP, *reports* e *dashboards*.

1.3. Metodologia de investigação

Os sistemas de informação são uma área de investigação onde, frequentemente, se aplicam teorias de muitas outras áreas do saber para resolver os problemas de integração das TI nas organizações (Peffer, Tuunanen, Rothenberger, & Chatterjee, 2007).

Nesse seguimento e para que esta dissertação possa ser concluída dentro da calendarização e com sucesso, foi selecionada a partir de uma pesquisa de várias metodologias aquela que melhor se enquadra nas práticas e procedimentos necessários para obter o resultado esperado.

A metodologia de investigação selecionada e aplicada foi a *Design Science Research*, sendo esta altamente relevante na investigação em Sistemas de Informação (SI), pois no domínio científico, as tecnologias e sistemas de informação são dotados de conhecimento interdisciplinar. Os contributos para o seu desenvolvimento têm origem num variado leque de áreas de saber que vão desde as ciências da computação, gestão, psicologia, medicina entre muitas outras (Ferreira, Ferreira, Silva, & Carvalho, 2012).

A investigação em SI tem sido dominada por estudos sobre os impactos das TI nas organizações, equipas e nos indivíduos, enquanto que *Design Research* foi considerado como um domínio de disciplinas mais técnicas, como por exemplo associadas às ciências da computação. No entanto, no início dos anos 90, a comunidade de SI reconheceu a importância desta metodologia (*Design Science Research*) no sentido de melhorar a eficácia e a utilidade das TI no contexto de problemas de negócio no mundo real (Hevner & Chatterjee, 2010).

Esta metodologia de investigação científica acarreta consigo um conjunto de princípios, práticas e procedimentos fundamentais para a realização de pesquisas científicas aplicadas a um ramo específico do conhecimento. Esta pretende proporcionar um modelo de processo para a realização da DSR através de um conjunto de regras práticas e fornecer um modelo mental para a apresentação e avaliação da mesma (Peffer et al., 2007).

Para os autores desta metodologia, foi necessário proceder à conceção de um processo *Design Science Research Methodology* (DSRM), realizando uma pesquisa à literatura já existente, tentando construir, baseado nisso, algo que fosse comumente aceite para a realização de investigação com base nos princípios da *Design Science Research*. Segundo Peffer (2007), do seu estudo resultou um modelo de processo que consiste numa sequência de seis atividades descritas de seguida (Figura 1).

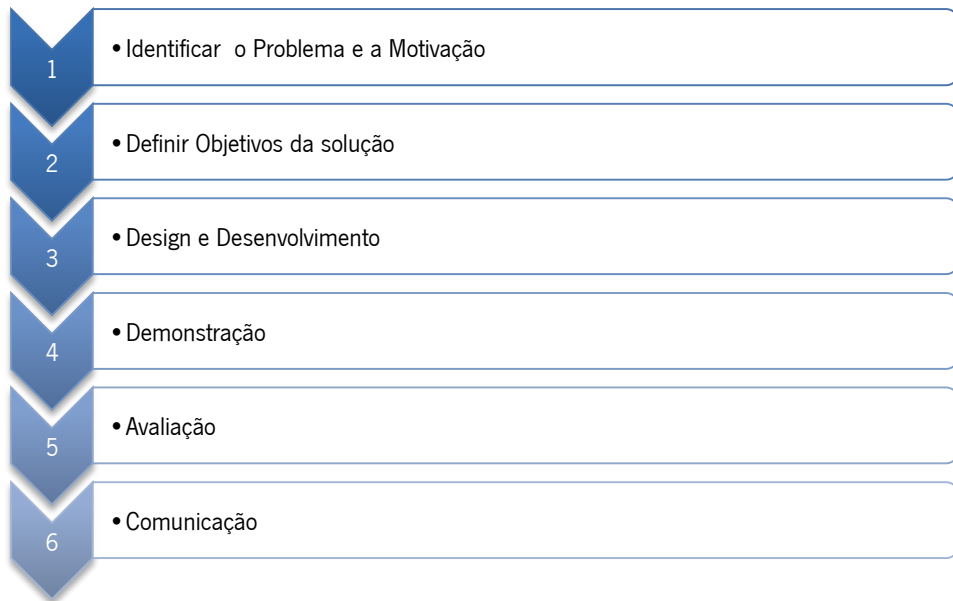


Figura 1 - Atividades do processo de Design Science Research

1) Identificar o problema e as motivações – Nesta primeira atividade pretende-se definir especificamente a problemática alvo de investigação e justificar devidamente a sua importância e o valor da solução. Para concluir esta fase é também necessário realizar a revisão da literatura de forma a focar nas temáticas principais do problema, bem como identificar o que já foi feito para a resolução do problema. Fazendo uma ponte com este projeto, a revisão de literatura incidirá sobre o tema de *Business Intelligence* bem como o conceito de *Data Warehouse* e os mecanismos analíticos subjacentes.

Para o desenvolvimento desta atividade é necessário ter o conhecimento do problema e da importância da solução.

2) Definir objetivos da solução – Após a definição do problema e sabendo o que é viável, devem inferir-se sobre essa análise os objetivos da solução apresentada. Para se desenvolver esta atividade é fundamental ter uma visão do que já existe e da sua eficácia através do estado da arte.

3) Design e Desenvolvimento – Com a aquisição do conhecimento acerca do estado da arte e feita a definição dos objetivos para a solução pretendida, procede-se à conceção da mesma. Nesta fase devem determinar-se as funcionalidades desejadas da solução bem como a sua arquitetura, procedendo-se finalmente ao

desenvolvimento da solução de *Business Intelligence* com o desenvolvimento do *Data Warehouse* e de todas as suas etapas de construção.

- 4) Demonstração** – Esta fase prende-se com a experimentação da solução desenvolvida, testando a mesma para resolver um ou mais dos problemas identificados inicialmente.
- 5) Avaliação** – Na fase final deste processo deve fazer-se uma reflexão a tudo o que já foi feito, desde a identificação do problema até ao desenvolvimento, fazendo-se uma medição da capacidade que o artefacto tem para suportar a solução apresentada para o problema. Esta atividade tem como finalidade comparar os objetivos definidos na atividade 2 com os resultados obtidos depois do seu desenvolvimento em concordância com as métricas definidas.
- 6) Comunicação** – A última atividade deste processo está relacionada com a publicação dos resultados obtidos. Como este projeto está a ser desenvolvido no âmbito de uma dissertação de mestrado, a comunicação será feita tal como regulamentado pela instituição de ensino através da elaboração de um documento científico e de uma apresentação pública avaliada por um conjunto de professores de instituições de ensino superior.

1.4. Estrutura do documento

Para que se compreenda melhor a estrutura deste documento, é explicado nesta secção o que se pretende em cada capítulo. O presente documento encontra-se dividido por capítulos que, por sua vez, se desdobram em secções e subsecções para que se obtenha uma melhor compreensão e organização dos conteúdos.

O primeiro capítulo diz respeito à introdução feita ao tema proposto com um enquadramento e as motivações que conduziram à realização do mesmo. Nas secções seguintes são definidos os objetivos e finalidade, bem como a metodologia de investigação utilizada para a concretização desta dissertação de mestrado.

No segundo capítulo é feito um enquadramento de uma perspetiva mais teórica dos temas relacionados. Este enquadramento conceptual foca-se nos conceitos de *Business Intelligence (BI)*, *Data Warehouse (DW)*, dados espaciais e mecanismos de processamento analítico.

O terceiro capítulo diz respeito ao enquadramento tecnológico e caracterização dos dados. É feita a definição das tecnologias a utilizar que irão suportar e adequar-se às necessidades da realização deste projeto. É apresentada uma proposta de arquitetura do sistema, através de uma abordagem concetual e tecnológica, que seja capaz de suportar o sistema pretendido. É ainda demonstrada a qualidade dos dados fornecidos e que servirão de base para o desenvolvimento do protótipo de um sistema de *Business Intelligence*. São expostos alguns dos resultados obtidos das análises exploratórias bem como tecidas algumas conclusões sobre a sua mais-valia para a sua utilização no projeto.

No quarto capítulo pretende-se expor a implementação de todo o sistema. São apresentadas todas as fases do desenvolvimento e implementação do sistema que foi idealizado para a concretização desta dissertação de mestrado. É feita a descrição do modelo multidimensional do *Data Warehouse*, seguindo-se a descrição de todo o processo ETL e respetiva descrição do desenvolvimento do processo analítico. Por fim, são apresentadas as aplicações *front-end* que possibilitam ao utilizador proceder à análise de toda a informação.

Por último, no capítulo quinto, apresentam-se as conclusões obtidas da realização desta dissertação de mestrado. Serão referidas as limitações e dificuldades sentidas bem como as propostas de trabalho futuro.

2. Enquadramento Concetual

Neste capítulo é realizada a revisão da literatura necessária para uma melhor compreensão dos temas a serem abordados na elaboração desta dissertação. Dada a dimensão deste projeto torna-se pertinente e fundamental realizar uma análise do estado atual dos conceitos de *Business Intelligence*, sistemas de *Data Warehouse*, mecanismos de processamento de dados bem como apresentar alguns casos de estudo realizados na área do ensino.

2.1. *Business Intelligence*

2.1.1. O conceito

O termo *Business Intelligence* já não é um conceito propriamente recente, tendo sido a empresa *Gartner* a responsável pela atribuição desta nomenclatura na década de 80. Este termo, normalmente abreviado como BI, é um conceito abrangente que inclui aplicações, infraestruturas, ferramentas e boas práticas que permitem o acesso e análise de informação para melhorar e otimizar o desempenho das organizações, bem como as decisões (Gartner, 2013b).

Os sistemas de BI alimentam-se dos dados disponíveis nas organizações e disponibilizam informação relevante para a tomada de decisão. Combinam um conjunto de ferramentas que sejam capazes de gerar relatórios para produzir informação, que será posteriormente utilizada pela gestão de topo das organizações no suporte à tomada de decisão (Santos & Ramos, 2006).

David Loshin (2012) define BI como todos os processos, tecnologias e ferramentas necessárias para transformar os dados em informação que posteriormente será transformada em conhecimento pertinente para todas as atividades do negócio. Além disto, considera que BI engloba o armazenamento de dados, as ferramentas de análise e de gestão do conhecimento.

Por sua vez, Hugh J. Watson (2007) considera um sistema de BI como um processo dividido em duas atividades principais, sendo que a primeira diz respeito à obtenção dos dados, ou seja, ao armazenamento dos dados que chegam de uma determinada origem num *Data Warehouse*. A segunda representa a forma como esses dados são depois extraídos, transformados, carregados, analisados e visualizados.

Como se pode verificar, os sistemas de BI colocam à disposição das organizações informação essencial como uma componente fundamental para o entendimento do negócio por parte dos gestores, que possibilita criar condições internas e externas que favorecem o sucesso da organização.

Segundo Santos e Ramos (2006), os sistemas de BI oferecem um conjunto de contributos importantes às organizações. Estes contribuem para aumentar a inteligência coletiva das organizações, a capacidade de aprendizagem das mesmas e a criatividade organizacional ao apoiar a produção de novas ideias, produtos ou serviços que permitem à organização adaptar-se de forma dinâmica.

Na perspetiva de Negash (2004), por vezes BI significa tomar decisões *on-line*, ou seja, decisões no instante. Este entende que, na maioria das vezes, estes sistemas pretendem diminuir o período de tempo entre a recolha e a disponibilização da informação, para quando chegar o momento da tomada de decisão esta seja, de facto, útil.

O mesmo refere que, com o despoletar da internet, o crescimento do *Data Warehouse* como um repositório, a evolução das técnicas de limpeza de dados e as cada vez maiores capacidades de *hardware e software*, levaram a que fosse possível, com a combinação de todas elas, criar sistemas de BI mais consistentes dos que existiam anteriormente. Para alimentar estes sistemas, Negash (2004) apresenta na Figura 2 os sistemas a partir dos quais o BI obtém informação.

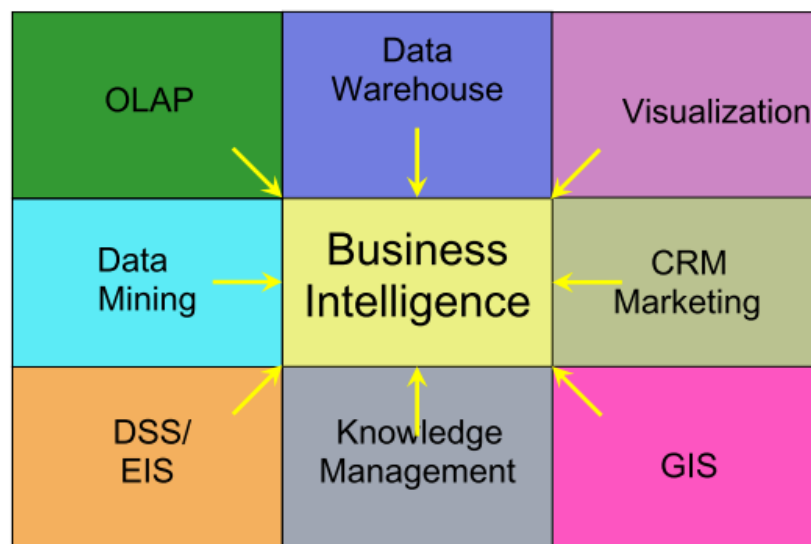


Figura 2 - Relação do BI com outros Sistemas de Informação (retirado de (Negash, 2004))

Como se pode verificar na Figura 2, o BI é um conceito amplo e versátil que possibilita que este seja suportado por um vasto conjunto de sistemas e tecnologias, cumprindo assim com as exigências e necessidades do meio em que está a ser implementado tornando-o mais eficiente e útil.

Estes sistemas, depois de implementados, têm como objetivo a elaboração de previsões baseadas em dados históricos, desempenhos passados e atuais da organização com a capacidade de criar cenários que evidenciam o impacto da alteração de várias variáveis. Estes permitem também o acesso *ad-hoc* (acesso não pré-definido) aos dados para responder a questões que não foram previamente definidas permitindo assim uma análise detalhada, obtendo um conhecimento mais aprofundado da organização (Negash, 2004).

Fazendo uma análise a cada uma das definições é possível constatar que BI é um tema amplamente bem definido, pois também já não é algo muito recente. De uma forma geral, todos os autores destacam a capacidade e a mais-valia destes sistemas para as organizações, no sentido em que estes proporcionam um conjunto de informações que se traduzirão em melhor conhecimento sobre uma organização e conseqüentemente permitirá aos seus responsáveis tomar melhores opções e de forma mais rápida.

É possível concluir que um dos principais benefícios do BI seja precisamente, como já referido anteriormente, a capacidade de criar mecanismos de análise precisos da informação histórica, como meio de auxílio para atingir o sucesso e os objetivos organizacionais.

2.1.2. Componentes de um Sistema de *Business Intelligence*

A implementação de um projeto de *Business Intelligence* é uma decisão algo complexa, pois desde o momento em que se pondera sobre as reais necessidades para a utilização deste tipo de sistemas até à sua implementação, são necessárias competências ao nível da gestão de projetos, de risco e da mudança que devem ser consideradas com bastante rigor para que o resultado seja, de facto, o esperado.

Posto isto, e quando se pensa na estratégia de implementação, surge instantaneamente uma das primeiras questões que se prende com: Qual será a melhor estratégia para se implementar um sistema de *Business Intelligence*?

Os sistemas de *Business Intelligence* têm aplicado a funcionalidade, escalabilidade e segurança dos atuais sistemas gestores de bases de dados para a construção de *Data Warehouses*, conforme explica Santos & Ramos (2006). Estes são analisados através de técnicas *On-Line Analytical Processing* e *Data mining* que visam a obtenção de informação relevante e precisa para o negócio.

Um sistema de BI é composto por quatro grandes componentes, tal como refere Turban (2007):

- *Data Warehouse;*
- *Business Analytics;*
- *Business performance management;*
- *Interface do utilizador.*

O primeiro componente, o *Data Warehouse*, é segundo Turban (2007) a “pedra angular” de qualquer sistema de BI. Este é considerado como um repositório de dados históricos, preparado para dar suporte às necessidades da tomada de decisão. A componente de *Business Analytics* corresponde a ferramentas que ajudam o utilizador a transformar os dados do DW em informação, incluindo, por exemplo, processamento analítico, multidimensionalidade e técnicas de análise avançada. A terceira componente, *Business Performance Management*, serve para monitorizar e comparar o desempenho da organização com metas e objetivos definidos, com aplicação de *balanced scorecards*. A última componente corresponde à interface do utilizador, ou seja, o modo como é disponibilizada a informação gerada.

Seguidamente é possível observar uma arquitetura de alto nível de um sistema de *Business Intelligence* (Figura 3) que é constituída por três níveis: ambiente *Data Warehouse*, *Business Analytics* e *Performance and Strategy*.

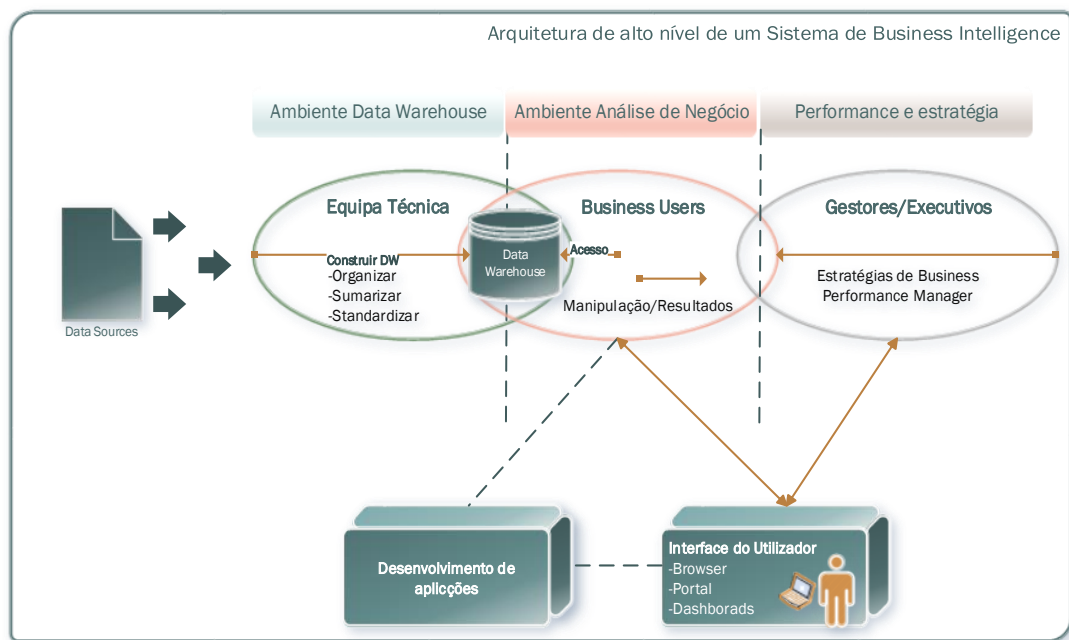


Figura 3 - Arquitetura de Alto Nível de um Sistema de BI (Adaptada de (Turban, Sharda, Aronson, & King, 2008))

Nesta arquitetura é possível verificar que existem responsáveis para cada um dos três níveis propostos. Para o ambiente de *Data Warehouse* existe uma equipa técnica responsável por construir e implementar o *Data Warehouse* que fazem com que os dados passem pelo processo de extração, transformação e carregamento. No segundo nível, os utilizadores responsáveis pelo negócio têm a tarefa de transformar os dados em informação útil usando ferramentas e técnicas como *queries*, *reports*, *data mining*, *web mining* e outras ferramentas matemáticas e estatísticas sofisticadas. O terceiro nível é da responsabilidade dos gestores de topo que devem utilizar a informação retirada do sistema para avaliar o estado e definir o rumo da organização consoante os objetivos traçados. Estes três níveis giram em torno do sistema de *Data Warehouse*, e apesar de cada um dos níveis ter um responsável, qualquer utilizador pode aceder ao sistema através da interface de utilizador, tendo assim disponível um conjunto de ferramentas de visualização, como por exemplo *dashboards*, capazes de fornecer uma visão abrangente do negócio através de indicadores (KPI's).

Mas qual será a melhor estratégia para se implementar um sistema de *Business Intelligence*?

A resposta está na construção de um *Data Warehouse* como um repositório integrado dos dados da organização. É um processo que permite a extração dos dados dos sistemas transacionais e operacionais, com capacidade para limpar e transformar esses dados permitindo a sua análise. As vantagens da implementação do *Data Warehouse* são:

- Possibilidade de criar um ambiente para integração e tratamento dos dados;
- Proporciona um ambiente cujo objetivo é único;
- Independência dos sistemas legados (transacionais);
- Repositório dos dados otimizado para realizar consultas;
- Suporte de ferramentas específicas para esse fim.

Este conceito será explicado em mais detalhe na secção 2.2 deste documento.

2.1.3. *Business Intelligence* na Área do Ensino

Nesta secção pretende-se apresentar alguns casos específicos da utilização de sistemas de *Business Intelligence* na área do ensino e da educação. Serão analisados os casos de estudo encontrados com uma descrição dos sistemas utilizados e qual o seu propósito.

Seguidamente serão apresentados os casos que foram encontrados quer a nível nacional e internacional:

- *Data Warehouse based evolution of students achievements in information systems education* (Kazi, Kazi, & Radulovic, 2012);
- *Use Data Warehouse and Data Mining to predict student academic performance in schools* (Kurniawan & Halim, 2013);
- *SBIAES – Business Intelligence systems for analysis of access higher education: the case of the polytechnic institute of Coimbra* (Nobre, Trigo, & Sanches, 2014);
- *School Performance Evaluation in Portugal: A Data Warehouse Implementation to Automate Information Analysis* (Castro, 2011);
- *Conceção e projeto de Data Warehouse: Olhar o futuro do ensino Básico e Secundário em Portugal* (Mickael Santos, 2014);

No primeiro artigo, “*Data Warehouse based evolution of students achievements in information systems education*”, é apresentada a implementação de um *Data Warehouse*, onde foram armazenados dados de quinze anos de avaliações de testes e trabalhos dos alunos dos cursos da área de sistemas de informação da Universidade de Novi Sade na Sérvia. O principal objetivo deste projeto era mostrar que os sistemas de *Data Warehouse* podem ser utilizados como instrumento de avaliação dos dados das avaliações dos alunos. Para a implementação deste *Data Warehouse* foram utilizadas duas fontes de dados com as avaliações dos testes e dos projetos de cada aluno, que depois de carregados para o *Data Warehouse* foram

submetidos a ferramentas OLAP para se proceder à sua análise tal como se pode ver na Figura 4.

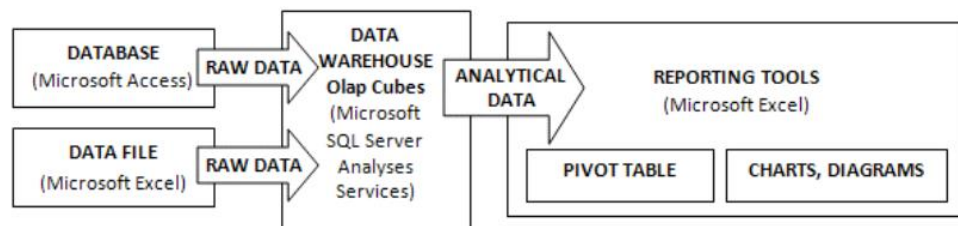


Figura 4 - Arquitetura do sistema de DW da Universidade de Novi Sade (retirado de (Kazi et al., 2012))

Foi então possível demonstrar as tendências nos dados ao longo do tempo bem como as variações dos resultados dos alunos influenciadas por certas mudanças no ambiente educacional.

Na mesma perspectiva segue o caso de estudo apresentado no artigo “*Use Data Warehouse and Data Mining to predict student academic performance in schools*” que visava analisar as avaliações dos alunos de uma escola da Indonésia (*BINUS International School*). Porém, este artigo propõe um modelo (Figura 5) que pode ser aplicado em sistemas de *Data Warehouse* e técnicas de *Data Mining* para prever o desempenho dos alunos.

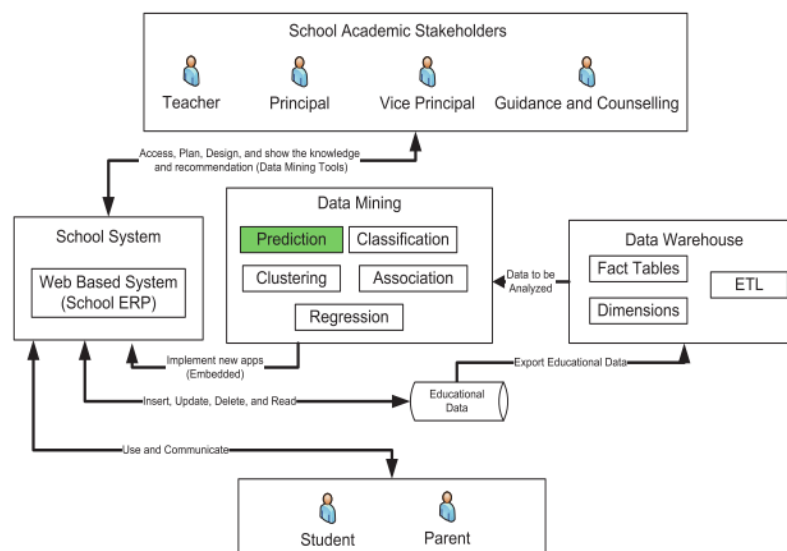


Figura 5 - Modelo de Data Mining e Sistema de Data Warehouse (retirado de (Kurniawan & Halim, 2013))

O modelo apresentado anteriormente foi projetado segundo a metodologia de Kimball e pensado para que seja possível utilizar sobre ele técnicas de *Data Mining*. O modelo permite carregar para o DW as notas dos testes e notas finais das disciplinas dos alunos bem como o seu registo de assiduidade. Com base nesse desempenho escolar, é possível realizar previsões através de técnicas de *Data Mining* quanto ao estado de avaliação de um aluno (aprovado ou reprovado). Com a aplicação deste modelo é possível ajudar os estudantes com menos rendimento, avaliar os cursos e prever medidas de intervenção para aumentar o desempenho académico dos alunos nas escolas.

O artigo sobre o caso do Instituto Politécnico de Coimbra (IPC) apresenta a implementação de um sistema de *Business Intelligence* que esta instituição decidiu desenvolver, o SBIAES, que tem como principal objetivo analisar a informação relativa aos candidatos ao Concurso Nacional de Acesso ao ensino superior nos últimos anos. Este projeto nasceu da necessidade do IPC em adquirir um conhecimento mais aprofundado das preferências em termos de cursos dos candidatos bem como as principais instituições concorrentes. Para isso, o SBIAES, assenta numa arquitetura de *Data Warehouse* e uma aplicação cliente em Microsoft Excel seguindo a metodologia proposta por Kimball tal como se pode observar na Figura 6.

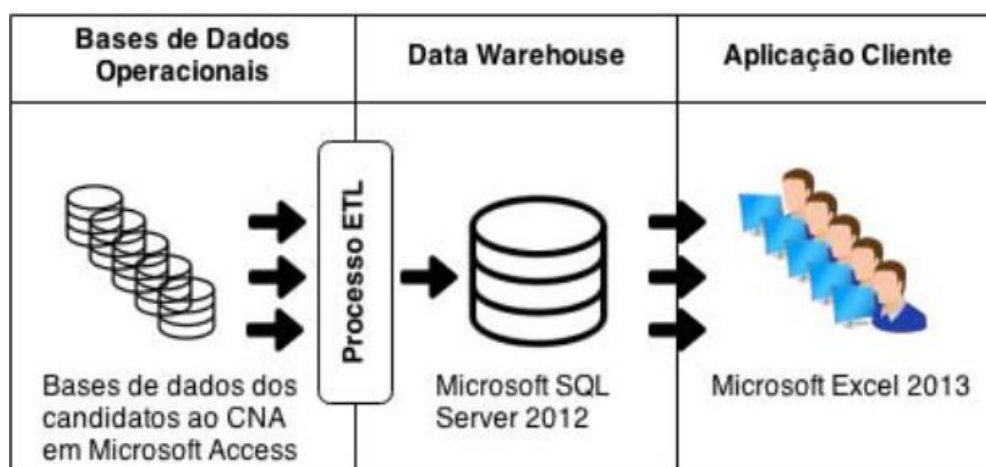


Figura 6 - Arquitetura do SBIAES (retirado de (Nobre et al., 2014))

Os dados utilizados para alimentar este sistema são provenientes das bases de dados do Concurso Nacional de Acesso cedidos pela Direção Geral do Ensino Superior (DGES), onde é possível obter os cursos e universidades preferidas, escolas e cursos secundários de onde são provenientes os alunos bem como o número de vagas de cada curso do ensino superior. A implementação deste sistema permitiu ao IPC aprofundar o conhecimento ao nível das

preferências dos candidatos ao concurso nacional de acesso, possibilitando à instituição posicionar-se competitivamente melhor. Desta forma foi possível dar suporte aos decisores quanto ao planeamento estratégico da instituição quer ao nível da sua promoção como ao nível da sua oferta formativa.

No artigo “*School Performance Evaluation in Portugal: A Data Warehouse Implementation to Automate Information Analysis*” é apresentada a implementação de um sistema de *Data Warehouse* com o objetivo de avaliar e analisar o desempenho interescolar em Portugal bem como o desempenho intraescola através de um modelo multidimensional adequado. Os dados carregados para o DW são provenientes das bases de dados do Ministério da Educação e de uma escola em particular, o colégio Paulo VI de Gondomar. Para a análise interescolar foram utilizados os dados referentes aos exames nacionais e para a análise intraescola foram utilizados dados dos alunos, professores, escola e disciplinas. A implementação deste sistema permitiu analisar o desempenho escolar a partir de diferentes pontos de vista, trazendo como vantagens o aumento da confiabilidade e integridade da informação, segurança da informação e maior acesso a uma grande quantidade de informação.

Na tese de mestrado, “Conceção e projeto de *Data Warehouse*: Olhar o futuro do ensino Básico e Secundário em Portugal” é feita uma proposta de um sistema de *Data Warehouse* projetado e planeado para uma empresa privada de consultadoria que presta serviços a escolas. O principal objetivo neste projeto foi possibilitar à empresa, após a implementação do projeto, analisar os dados das escolas de uma forma integrada, para que fosse possível encontrar os fatores que mais influenciam o desempenho escolar dos alunos. Para isso foram utilizados dois ficheiros em *Excel* de duas escolas (Figura 7) de forma a analisar o desempenho com base nas notas e percentagens de aprovação tendo em conta as componentes escola, disciplina, escalão social dos alunos, sexo, idade e as habilitações dos pais dos alunos.

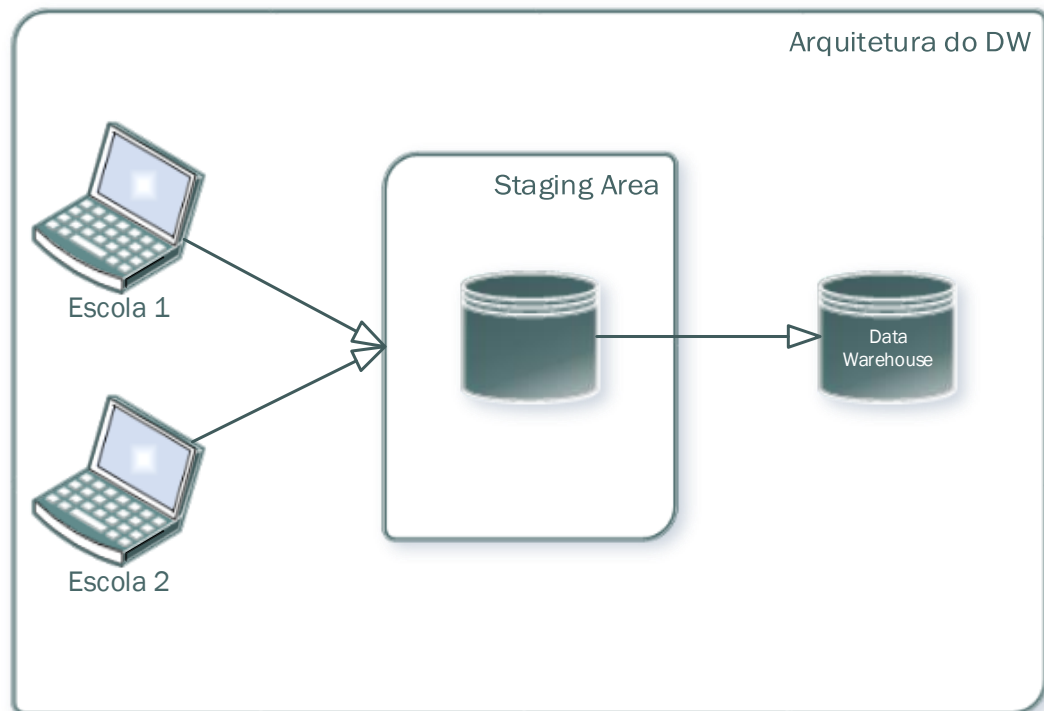


Figura 7 - Arquitetura do *Data Warehouse* (adaptado de ((Mickael Santos, 2014)))

No modelo proposto, os dados das duas escolas são extraídos para a *staging área* onde se realiza o tratamento dos mesmos para que posteriormente sejam devidamente carregados para o DW. Neste projeto é feita a proposta de um modelo multidimensional que não permite verificar a evolução dos alunos ao longo do tempo. A proposta feita nesta tese seguiu a metodologia de Kimball. Com isto, a empresa poderá fornecer às escolas um serviço que suporte a tomada de decisão para que seja possível melhorar o desempenho dos alunos nas escolas.

Por forma a salientar que o trabalho proposto é algo de inovador, é apresentado na Tabela 1 uma matriz dos conceitos abordados em cada um dos trabalhos apresentados anteriormente. Na matriz da Tabela 1 é feito o cruzamento dos trabalhos já existentes com alguns conceitos de *Business Intelligence* e conceitos da área de ensino em Portugal.

Tabela 1 - Matriz de conceitos de *Business Intelligence* na área do Ensino

Trabalho/Conceitos	DW	Data Mining	Data Mart	Ensino Obrigatório	Ensino Superior	Carta Educativa	Metodologia Kimball
<i>Data Warehouse based evolution of students achievements in information systems education</i>	X				X		
<i>Use Data Warehouse and Data Mining to predict student academic performance in schools</i>	X	X			X		X
<i>SBIAES – Business Intelligence systems for analysis of access higher education: the case of the polytechnic institute of Coimbra</i>	X				X		X
<i>School Performance Evaluation in Portugal: A Data Warehouse Implementation to Automate Information Analysis</i>	X			X			
Conceção e projeto de <i>Data Warehouse</i> : Olhar o futuro do ensino Básico e Secundário em Portugal	X			X			X

Na revisão de conceitos efetuada foram encontrados mais exemplos de desenvolvimento de sistemas de BI nas instituições de ensino como é o caso do sistema desenvolvido por Turkmen (2007) na Universidade de Atilim para analisar o desempenho escolar dos estudantes daquela instituição. Outro exemplo da implementação destes sistemas foi realizado nas Universidades da Pensilvânia (Pennsylvania, 2015) e de Miami (Miami, 2015) onde é permitido o acesso a

determinados dados do *Data Warehouse* ou *Data Marts* por parte dos alunos, professores e funcionários.

Após o levantamento dos vários casos de estudo, constata-se que a implementação deste tipo de sistemas em Portugal ainda não é uma realidade, verificando-se poucos casos registados. Porém, em outros países, sobretudo nos Estados Unidos da América, já existem bastantes exemplos da implementação deste tipo de sistemas no ensino superior sendo já uma realidade cada vez mais vincada.

2.2. Sistemas de *Data Warehousing*

2.2.1. Definição

As origens do armazenamento de dados e dos sistemas de apoio à decisão (SAD) remontam aos primórdios da computação e dos sistemas de informação (Inmon, 2002).

As bases de dados e as teorias associadas às mesmas já existem há bastante tempo. Inicialmente e na maioria dos casos, o foco principal destes sistemas de base de dados era operacional, usual em processamento transacional (Inmon, 2002). Nos últimos anos, uma conceção mais sofisticada das bases de dados tem emergido, designada de *Data Warehouse*, capaz de servir as necessidades informacionais/analíticas.

Segundo Inmon (2002), a divisão das bases de dados em operacionais e informacionais ocorreu por diversas razões. Uma delas prende-se com o facto de os dados que servem as necessidades operacionais serem fisicamente diferentes dos dados que servem as necessidades informacionais/analíticas. Outras duas razões que levaram a esta divisão estão relacionadas com as características de processamento e as tecnologias de suporte, pois estas são fundamentalmente diferentes do ambiente operacional para o ambiente informacional/analítico.

Kimball (1998) afirma que um dos ativos mais importantes de uma organização é a informação que esta possui. Tal como Inmon, Kimball refere que esse ativo, a informação, é quase sempre mantido por uma organização de duas formas, ou em sistemas operacionais ou em sistemas informacionais designados de *Data Warehouse*, em que ambos usam servir diferentes necessidades.

Para mostrar as diferenças existentes entre bases de dados operacionais e sistemas de *Data Warehouse*, Santos e Ramos (2006) salientam essas diferenças devidamente sistematizadas na Tabela 2, evidenciando assim as principais características destes dois tipos de bases de dados.

Tabela 2 - Bases de dados operacionais vs. *Data Warehouse* (adaptado de (Santos & Ramos, 2006))

Bases de dados Operacionais	<i>Data Warehouse</i>
Objetivos Operacionais	Registo Histórico
Acessos de leitura/escrita	Acessos só de leitura
Acesso por transações predefinidas	Acesso por questões <i>ad-hoc</i> e relatórios periódicos
Acesso a poucos registos de cada vez	Acesso a muitos registos de cada vez
Dados atualizados em tempo real	Carregamentos periódicos de mais dados
Estrutura otimizada para atualizações	Estrutura otimizada para processamento de questões

Delvin e Murphy, dois investigadores da IBM, utilizaram pela primeira vez em 1988 a expressão “*Information Warehouse*” que é considerado o termo anterior ao termo *Data Warehouse* e que lhe deu origem (Sá, 2010). O termo *Data Warehouse* foi empregue pela primeira vez em 1991, por Inmon, quando este publicou o seu livro “*Building the Data Warehouse*”.

Sá (2010) afirma que este termo, designado de *Data Warehouse*, surge no século XX na segunda metade da década de 80 e era anunciado como:

- O fornecedor da “versão única da verdade”;
- A forma de melhorar o processo de gestão;
- O suporte fundamental para as iniciativas de melhoria do negócio com o cliente (*Business to Customer*) e negócio com outros negócios (*Business to Business*);
- A gestão do relacionamento com os clientes (CRM);

Santos e Ramos (2006) definem um *Data Warehouse* como “um armazém de dados, um repositório integrado que permite o armazenamento de informação relevante para a tomada de decisão. Estes repositórios podem ser analisados utilizando ferramentas *On-Line Analytical Processing* e/ou ferramentas de *Data Mining*”.

Um *Data Warehouse*, segundo Inmon (2002), é orientado ao assunto, integrado e não volátil que varia temporalmente e é capaz de apoiar na tomada de decisão. Este contém dados corporativos que podem ser usados para várias finalidades.

Da definição dada anteriormente por Inmon, importa clarificar alguns conceitos que o mesmo considera fundamentais.

- Orientado ao assunto – a informação armazenada é organizada por assuntos chave de uma organização;
- Integrado – os registos informacionais são selecionados e armazenados no *Data Warehouse* provenientes de fontes distintas onde se juntam para fornecer uma visão coerente;
- Não volátil – Num *Data Warehouse* pode ser sempre adicionada nova informação, mas não pode ser removida do mesmo;
- Varia temporalmente – Toda a informação armazenada no *Data Warehouse* deve estar catalogada temporalmente através de uma dimensão tempo para que seja possível obter um registo informacional histórico.

Noutra perspetiva, Kimball (1998) entende que um *Data Warehouse* é uma cópia de registos informacionais provenientes de uma transação estruturada para que, sobre eles se possam realizar interrogações e análises. Este refere também que estes sistemas visam um conjunto de objetivos, como a capacidade de fazer com que a informação da organização esteja acessível bem como consistente. O DW é uma fonte de informação adaptável e resiliente, é como um “guardião” da informação, e por fim, Kimball (1998) assume que o DW é a base para a tomada de decisão. Importa também nesta definição clarificar os objetivos que o autor enumera, nomeadamente:

- Informação acessível – os dados armazenados são compreensíveis, navegáveis e o seu acesso caracteriza-se por um desempenho rápido;
- Informação consistente – A informação que é carregada para o DW é de alta qualidade, o que significa que esta é categorizada, bem rotulada e completa;
- Adaptável e resiliente – o DW é projetado para a mudança contínua, portanto quando novas questões são colocadas aos dados, as tecnologias que suportam não são alteradas ou interrompidas. O mesmo acontece quando nova informação é adicionada;
- Segurança da informação – O DW não controla apenas o acesso aos dados, oferece também aos seus administradores uma boa visibilidade do uso dos dados;
- Base para a tomada de decisão – A grande finalidade deste tipo de sistemas é o apoio aos gestores na sua tomada de decisão, sendo este o seu verdadeiro *output*. Quando se implementa um sistema destes está-se a construir um sistema de apoio à decisão.

Analisando cada uma das definições apresentadas, pode verificar-se que é destacada a capacidade de estes sistemas conseguirem transformar os dados operacionais em informação útil. Esta encontra-se integrada de forma consistente, apresentando a capacidade de apresentar temporalmente um registo. É assim um conceito definido de forma algo idêntica, no sentido em que todos os autores referem que os sistemas de *Data Warehouse* são uma ótima fonte de informação, útil e eficaz no apoio à tomada de decisão, sendo este um dos principais objetivos enumerados.

Um *Data Warehouse* é portanto um sistema construído para uniformizar e consolidar a informação de uma organização, sendo este uma base de dados mantida de forma autónoma em relação às bases de dados operacionais. Toda a informação armazenada é etiquetada temporalmente, ou seja, os dados não se alteram ao longo do tempo, permitindo armazenar dados de vários anos para que os utilizadores possam extrair e analisar os dados de uma forma seletiva.

2.2.2. Dimensões e Factos

Nesta seção do documento serão abordados os conceitos de tabelas de dimensão e de factos. Para que melhor se clarifiquem estes conceitos, serão abordados também os conceitos de cubo OLAP e de hierarquias dos dados.

2.2.2.1. Cubos OLAP

Os sistemas de *Data Warehouse* têm a capacidade de proporcionar o acesso e análise de diversas fontes de dados através de Cubos OLAP. Estes são compostos por dimensões que lhe conferem diferentes perspetivas e formas de visualização da informação como se pode observar na Figura 8 (Elmasri & Navathe, 2000; Jensen & Pedersen, 2010).

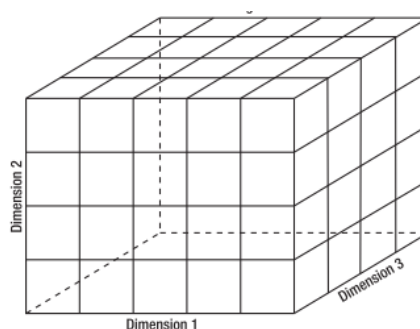


Figura 8 - Cubo com 3 dimensões (retirado de (Rainardi, 2008))

Na Figura 8 é apresentado como exemplo a constituição de um cubo com três dimensões, embora o termo “Cubo” implique três dimensões, um cubo pode ter mais dimensões. É possível verificar que através da interseção entre dimensões se formam diversas células que constituem o cubo. A essas células, sendo elas não vazias, dá-se o nome de factos (Jensen & Pedersen, 2010).

Para que tudo se conjugue e se obtenha a informação desejada, é possível definir hierarquias nas dimensões que constituem um cubo. Essa hierarquia confere-lhe a possibilidade de análise de dados com o nível de detalhe desejado, por forma a cumprir com os indicadores de negócio considerados na tabela de factos.

2.2.2.2. Dimensões

As dimensões são um conceito essencial em bases de dados multidimensionais. Estas permitem a análise da informação através de diferentes perspetivas. As tabelas de dimensão são utilizadas para selecionar os dados e agrupa-los de acordo com o nível de detalhe desejado (Jensen & Pedersen, 2010).

As tabelas de dimensão têm a capacidade para integrar um conjunto diversificado de atributos, ou colunas, que contêm, normalmente, descrições que permitem a contextualização das métricas em análise. Cada tabela de dimensão contém uma chave primária única que corresponde exatamente a uma das partes da chave composta da tabela de factos associada (Kimball, 1997; Santos & Ramos, 2006).

Um *Data Warehouse* é um repositório para consulta de dados que não permite ao utilizador realizar operações de escrita para possíveis alterações aos dados armazenados nas várias dimensões. A operação de escrita nas diversas dimensões (*slowly changing dimensions*) está reservada apenas ao momento de carregamento e refrescamento do *Data Warehouse*. Só é possível realizar alterações aos valores disponíveis nas dimensões nos períodos de refrescamento do DW, que por norma não são muito frequentes e devem ser bem planeadas e definidas (Santos & Ramos, 2006). Para se proceder a essas alterações nas tabelas de dimensões existem três alternativas:

- Escrever por cima: este tipo de atualização é o mais rápido e simples, que tem como desvantagem o facto de não ser possível manter o histórico do registo que está a ser alterado.
- Inserir um novo registo na dimensão: Esta opção é a mais utilizada, pois permite guardar todas as alterações efetuadas mantendo desta forma todo o histórico. Em contrapartida esta é a alternativa mais complexa de realizar. Para que seja implementada normalmente é utilizada uma chave primária composta por dois componentes. A primeira componente corresponde ao registo já existente e a segunda componente diz respeito ao registo substituto.
- Prever atributos adicionais nas dimensões: A última alternativa consiste em prever colunas que podem ser adicionadas cujos atributos podem ser alterados ao longo tempo. A desvantagem desta opção é o facto de apenas ficar guardada uma parte do histórico, pois quando uma alteração acontece, o atributo que vai ser alterado é guardado na coluna adicional substituindo o valor histórico que lá se encontrava. Este problema acontece, pois não é possível prever quantas alterações um determinado atributo será sujeito.

2.2.2.3. Tabela de factos

As tabelas de factos constituem um dos componentes principais dos modelos multidimensionais, pois permitem armazenar os assuntos que se pretendem analisar. Os factos são objetos que representam determinado assunto das análises pretendidas, ou seja, algo relevante para a organização que deve ser analisado para um melhor entendimento desse assunto (Jensen & Pedersen, 2010; Santos & Ramos, 2006).

As tabelas de dimensão e de factos estão relacionadas, uma vez que na maioria dos modelos multidimensionais os factos são implicitamente definidos pela combinação das dimensões. Desta combinação de dimensões, formam-se células não vazias a que se dá o nome de factos. A tabela de factos tem uma chave primária composta por duas ou mais chaves estrangeiras, sendo sempre expressa por uma relação de um para muitos com as tabelas de dimensão (Jensen & Pedersen, 2010; Kimball, 1997).

Os factos são valores numéricos e cada linha da tabela representa uma determinada métrica ou medida do processo de negócio (Tabela 3). Cada um dos acontecimentos (factos) deve

estar representado recorrendo à mesma granularidade dos dados (Santos & Ramos, 2006). Essa granularidade representa o nível de detalhe da informação armazenada nas tabelas de dimensão que será explicado no ponto seguinte.

Para que sejam mais claros os conceitos abordados e a relação existente entre as tabelas de factos (Tabela 3) e as tabelas de dimensão (Tabela 4) é apresentado, seguidamente, um exemplo simples da venda de produtos.

Tabela 3 - Tabela de Factos

TimeID	ProductID	CustomerID	UnitSold
6	20	5	5
10	40	7	6
3	33	6	6

A Tabela 3 apresenta o processo de negócio das vendas de três produtos. Cada linha representa uma determinada métrica desse processo, que é composta apenas por valores numéricos (*ID's*) que fazem referência às respetivas dimensões.

Tabela 4 - Tabela Dimensão: *Costumer*

CustomerID	Name	Gender	Region
5	João	M	3
6	Luís	M	4
7	Mariana	F	7

Na Tabela 4 é apresentada a dimensão *Costumer* com os dados dos clientes (nome, sexo e região). Na tabela de factos está registado que o cliente 7 (Mariana) comprou 6 unidades de um produto (id 40) numa determinada data com o id=10. Tal como é possível verificar na dimensão *Costumer*, é permitido saber mais detalhadamente os dados dos clientes, o mesmo acontece para as restantes dimensões no caso de se pretender saber, por exemplo, o dia da compra ou os detalhes do produto comprado.

Quando se pretende fazer uma análise dos factos, na maioria das vezes procede-se à agregação ou sumarização dos valores que lhe estão associados. Mas nem todos os factos apresentam as mesmas características, o que implica que não podem ser analisados da mesma maneira (Jensen & Pedersen, 2010; Santos & Ramos, 2006). Posto isto, é comum distinguir entre três tipos de factos (Kimball, 2013; Santos & Ramos, 2006):

- Factos aditivos: Este tipo de factos são os mais flexíveis e mais usados, pois possibilitam a agregação e sumarização por parte de todas as dimensões que estão associadas à tabela de factos. A agregação acontece, normalmente, através da soma de vários acontecimentos associados a um facto.
- Factos semi-aditivos: Neste tipo de factos, a agregação só pode ser feita através de algumas dimensões e não por todas. Pode ser apenas uma ou mais dimensões, mas nunca todas o que lhe confere essa particularidade em relação aos factos aditivos, pois haverá, pelo menos uma dimensão que não fará sentido ser agregada a um determinado facto.
- Factos não-aditivos: Existem certos factos que são não aditivos, que tal como o nome indica não podem ser agregados por nenhuma das dimensões do modelo multidimensional. Como a agregação remete para um valor total, a abordagem passa por armazenar, se possível, os componentes aditivos do facto não-aditivo e proceder à soma dessas componentes, caso contrário esse valor total, semanticamente não irá corresponder à realidade.

2.2.2.4. Hierarquia dos dados

Nas tabelas de dimensão pode existir uma estrutura hierarquizada. Esta hierarquia dos dados nas dimensões é importante porque lhes confere caminhos que possibilitam a análise dos dados. Nas tabelas de dimensão, por vezes um atributo (coluna) é um subconjunto de outro atributo, o que significa que é possível agrupar esse atributo por outro (Rainardi, 2008).

Por exemplo, na Figura 9, na dimensão *Store* se o atributo *Division* é um subconjunto de *Region*, então *Store* pode ser agrupado por *Region* e este pode ser agrupado por *Division*. Pode dizer-se que *Division* está a um nível inferior de *Region*.

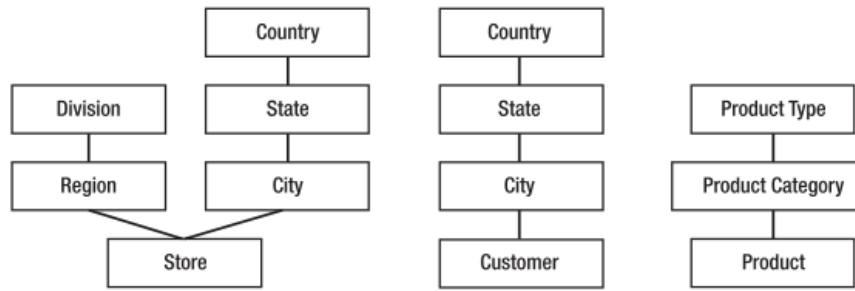


Figura 9 - Hierarquia dos dados de 3 dimensões (retirado de (Rainardi, 2008))

A definição das hierarquias é parte integrante do processo de construção das dimensões (Rainardi, 2008). Para se aplicar a hierarquia é necessário, em primeiro lugar, olhar para os atributos existentes nas tabelas de dimensão e identificar quais os conjuntos e subconjuntos. Após definir as hierarquias, é necessário verificar os dados para saber se os atributos que estão no nível mais baixo podem ser agrupados pelo atributo com o nível superior.

As tabelas de dimensão, por norma, são tabelas não normalizadas o que lhes permite conter uma ou mais hierarquias. No caso de existirem vários caminhos na mesma hierarquia, estes devem ser identificados.

2.2.3. Modelação Multidimensional

As bases de dados multidimensionais, ao contrário do que se possa pensar, não tiveram origem nas tecnologias de bases de dados, mas resultaram da álgebra matricial multidimensional que vinha sendo utilizada para a análise de dados desde o século IX (Jensen & Pedersen, 2010).

A modelação multidimensional é uma técnica utilizada para projetar bases de dados, que foi desenvolvida para a construção de *Data Warehouses* colmatando as limitações de métodos tradicionais de *design* de bases de dados utilizadas para tais finalidades (Moody & Kortink, 2003).

Este tipo de modelação é utilizado com o objetivo de produzir estruturas de dados fáceis de compreender para o utilizador com a capacidade de otimizar o desempenho no processamento de consultas (interrogações ao sistema), ao contrário da otimização do processamento de

atualizações, como acontece no modelo relacional de dados (Moody & Kortink, 2003; Santos & Ramos, 2006).

Esta abordagem tornou-se a mais predominante e viável para modelar dados para uso multidimensional, sendo um dos grandes avanços na construção de estruturas de bases de dados, possíveis de serem compreendidas e utilizadas diretamente pelos utilizadores finais (Kimball, 1997; Moody & Kortink, 2003).

As tabelas de factos e de dimensão estão relacionadas, isto é, estão ligadas através de um modelo de dados e são apresentadas em forma de esquema (Kazi et al., 2012). Os esquemas em estrela (*star schema*) e floco de neve (*snowflake schema*) são os dois modelos utilizados com mais frequência. Porém a modelação multidimensional pode ser construída através da implementação de três esquemas com características diferentes. Os esquemas em estrela, em floco de neve e em constelação serão apresentados de seguida.

2.2.3.1. Esquema em estrela

Os esquemas em estrela são a forma mais comum de se proceder à modelação dos dados. Na verdade este tipo de modelação é bastante simples pois baseia-se numa única estrutura de dados que é altamente regular por forma a facilitar as consultas. Este esquema consiste numa única tabela de factos que está rodeada por uma série de tabelas de dimensão (Moody & Kortink, 2003).

Este esquema foi “batizado” de estrela pois a sua disposição apresenta a forma de uma estrela estando a tabela de factos no centro, enquanto que as tabelas dimensões que a rodeiam formam os pontos da mesma tal como é possível visualizar na Figura 10.

Como já referido anteriormente, as tabelas de factos correspondem, normalmente, a processos de negócio que no exemplo apresentado, na Figura 10, corresponde ao processo de vendas. As tabelas de dimensão que rodeiam o esquema fornecem a informação necessária para permitir calcular as métricas definidas na tabela de factos.

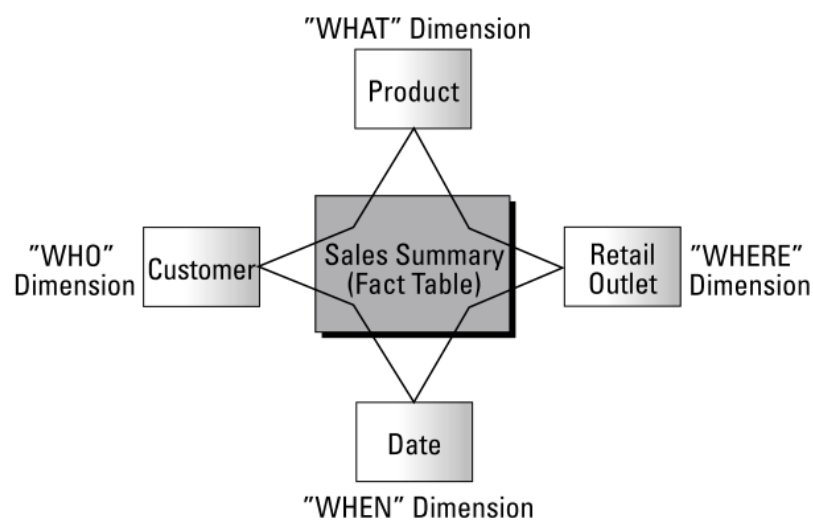


Figura 10 - Esquema em Estrela (Moody & Kortink, 2003)

O esquema em estrela apresentado na Figura 10 é constituído apenas por quatro dimensões, mas este tipo de esquemas pode ter um número qualquer de tabelas de dimensão. A partir delas é possível analisar a informação através de várias perspetivas, sendo possível responder a questões do tipo “quem?”, “quando?”, “onde?”, “porquê?” e “como?”.

O conceito de modelo multidimensional referido no ponto anterior provém deste esquema (estrela), pelo facto de este poder ser visualizado como um cubo, onde cada uma das suas dimensões representa cada uma das faces do cubo.

2.2.3.2. Esquema em floco de neve

Um outro modelo comumente utilizado é o esquema em floco de neve que surgiu da variação do esquema em estrela explicado anteriormente. Este esquema é semelhante ao esquema em estrela no que toca à estrutura, porém as tabelas de dimensão podem ser organizadas numa hierarquia e estão completamente normalizadas (Elmasri & Navathe, 2000).

Estes esquemas em floco de neve apresentam uma estrutura mais complexa que o esquema anterior. Estes deixam de ter uma estrutura regular, podendo cada um dos ramos do floco ter tamanhos e comprimentos diferentes, como se pode verificar no exemplo da Figura 11.

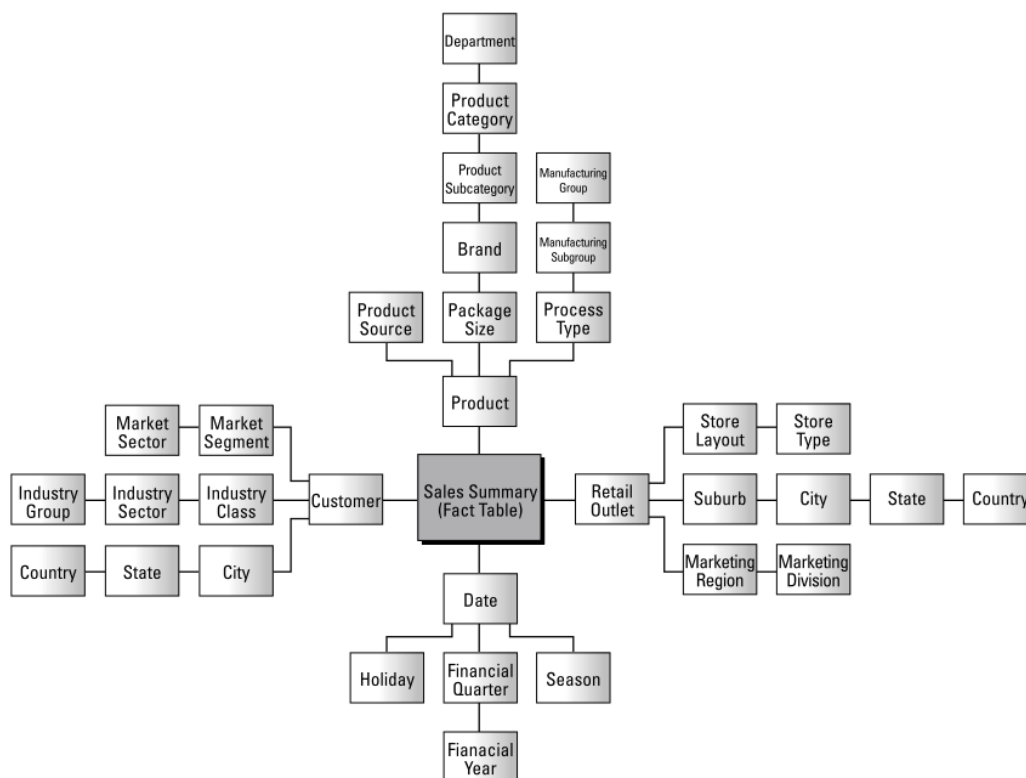


Figura 11 - Esquema em Floco de Neve (Moody & Kortink, 2003)

O esquema apresentado na Figura 11 apresenta uma maior complexidade em comparação ao esquema apresentado na Figura 10. Os atributos redundantes foram removidos tornando o modelo mais homogêneo, o que na perspectiva de Kimball (2003) acrescenta uma complexidade desnecessária, reduz o desempenho e não diminui substancialmente o espaço de armazenamento.

Como vantagens, este esquema permite indicar claramente a estrutura de cada uma das suas dimensões, evitando que a informação redundante seja armazenada permitindo aos utilizadores entender melhor as hierarquias dos dados e serem analisados de forma mais coerente (Ralph Kimball & Ross, 2003; Santos & Ramos, 2006).

2.2.3.3. Esquema em constelação

Por último, os esquemas em constelação apresentam uma estrutura formada por um conjunto de tabelas de factos que partilham entre si algumas das tabelas de dimensão do esquema (Elmasri & Navathe, 2000).

Na Figura 12 é apresentado um exemplo de um esquema em constelação que pode ser visto com um conjunto integrado de esquemas em estrela.

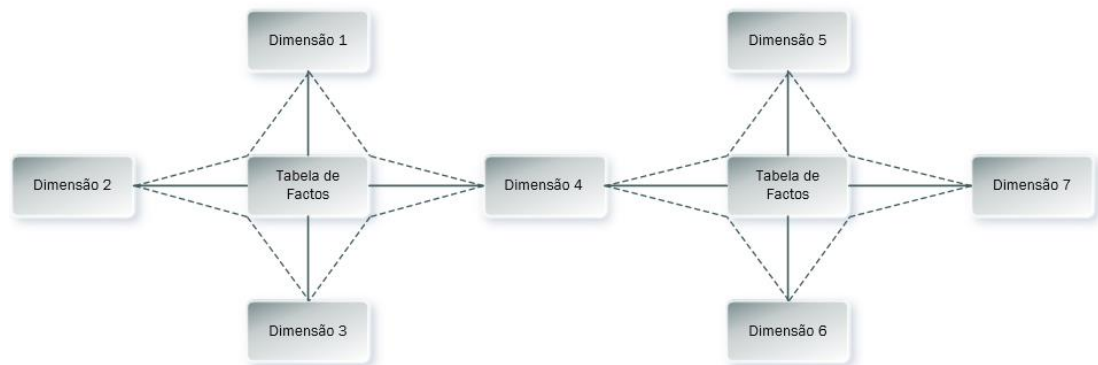


Figura 12 - Esquema em Constelação (adaptado de (Santos & Ramos, 2006))

A particularidade destes esquemas é a possibilidade de as várias estrelas poderem estar interligadas por uma ou mais dimensões (Santos & Ramos, 2006). No exemplo apresentado na Figura 12 apenas a dimensão 4 está a interligar as duas estrelas do esquema, mas poderiam existir mais dimensões a serem partilhadas pelas duas tabelas de factos.

2.2.4. Tipos de Servidores OLAP

A tecnologia OLAP, ou *on-line analytical processing*, é uma das atividades que faz parte do BI. Ferramentas OLAP são uma forma interativa de analisar a informação armazenada nas dimensões de um *Data Warehouse* para que os gestores possam tomar decisões estratégicas para o seu negócio (Rainardi, 2008).

Para Santos e Ramos (2006), os servidores OLAP possibilitam a análise multidimensional dos dados independentemente do local onde os mesmos estão armazenados e podem ser divididos em:

- ROLAP (*Relational OLAP*)
- MOLAP (*Multidimensional OLAP*)
- HOLAP (*Hybrid OLAP*)

Importa agora clarificar estes três conceitos em que os três autores (Rainardi, 2008; Santos & Ramos, 2006) entram em concordância.

Os sistemas ROLAP operam como intermediários entre a base de dados relacional e as ferramentas do tipo cliente para a análise dos dados. Os dados resumidos, chamados de agregados, são armazenados em tabelas que dependem de instruções SQL. Ao receber as consultas, os sistemas ROLAP escolhem as tabelas a ser usadas para satisfazer as instruções indicadas e, em seguida, complementam os resultados com instruções *select* para as tabelas base. Um exemplo de um sistema ROLAP é o *MicroStrategy OLAP Services*.

No caso dos sistemas MOLAP, estes caracterizam-se por utilizarem bases de dados multidimensionais, o que permite vistas multidimensionais sobre os dados. Um exemplo deste tipo de sistemas é o *Cognos BI & Analysis* e *ProClarity Analytics*.

Por fim, o caso híbrido (HOLAP), combina as duas tecnologias anteriores. Estes sistemas beneficiam da alta escalabilidade do ROLAP e da velocidade de processamento do MOLAP. É possível combinar estes dois sistemas e, por exemplo, armazenar todos os dados numa base de dados relacional e manter separadamente os agregados num motor MOLAP. Um exemplo da utilização deste tipo de sistema é o *Microsoft SQL Server* que também suporta ROLAP e MOLAP.

2.2.5. Operações para Analisar um Cubo

A forma como os dados são organizados pelo modelo multidimensional permite a análise dos dados através de diversas perspetivas. Essas análises aos dados são executadas sobre o cubo através de várias operações OLAP. Essas operações de manipulação dos cubos são (Chaudhuri & Dayal, 1997; Santos & Ramos, 2006):

- ***Drill-down***. Este é dos tipos de operação de análise mais antigo e mais comum num *Data Warehouse*. A operação de *Drill-down* tem como objetivo dar uma visão mais detalhada sobre os dados em análise (Ralph Kimball et al., 1998; Santos & Ramos, 2006). Por exemplo, adicionar uma linha de cabeçalho a um pedido SQL. Neste tipo de operação é possível navegar desde os dados mais generalizados para os dados com mais detalhe.
- ***Roll-up***. Esta operação, também denominada de *drill-up* por muitos autores, é precisamente a operação contrária ao *drill-down*. O objetivo é permitir a agregação dos dados mais detalhados para dados mais generalizados, isto é, permite agregar

dinamicamente esses dados utilizando uma determinada hierarquia (Santos & Ramos, 2006).

- **Slice and dice:** Esta operação tem a capacidade de aceder ao *Data Warehouse* através de qualquer dimensão (Ralph Kimball & Ross, 2003). Ela divide-se em duas operações (corte e redução) e permite restringir, de um conjunto de dados, a informação a visualizar (Figura 13).

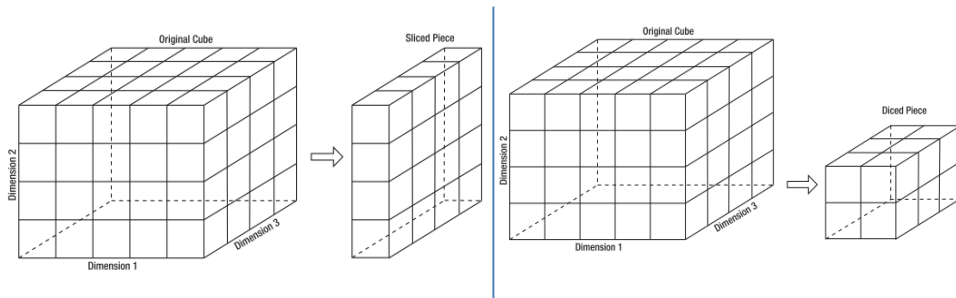


Figura 13 - Slice and dice (Rainardi, 2008)

O corte (*slice*) permite restringir os dados do cubo a um subconjunto de dados, filtrando uma dimensão segundo uma determinada condição. A redução (*dice*), permite definir a partir do cubo um subcubo filtrando duas ou mais dimensões (Rainardi, 2008; Santos & Ramos, 2006).

- **Pivot:** Nesta operação é possível rodar os eixos de visualização dos dados do cubo (Figura 14), o que disponibiliza diferentes perspectivas de visualização dos mesmos (Santos & Ramos, 2006).

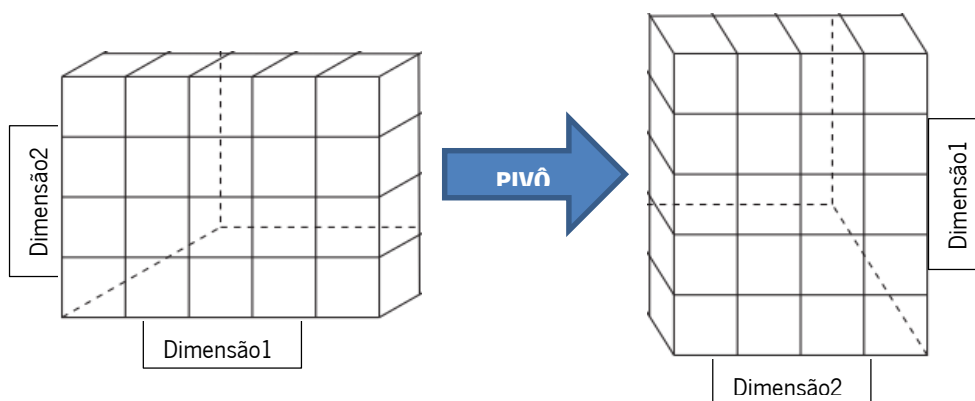


Figura 14 – Pivô (adaptado de (Santos & Ramos, 2006))

2.2.6. Metodologia de Sistemas de *Data Warehousing*

Nesta secção pretende-se clarificar quais os aspetos a considerar na construção de um sistema de *Data Warehouse* bem como a metodologia a seguir no âmbito deste projeto.

Uma metodologia tem por objetivo descrever o processo de construção e implementação de um determinado projeto, ou seja, quais os resultados a obter das várias atividades e técnicas adotadas para construir e implementar um sistema de *Data Warehouse*.

Segundo Sá (2010), um dos aspetos que mais gera confusão em sistemas deste tipo é a informação contraditória que existe sobre arquiteturas e metodologias de implementação. Por exemplo, questões como: devem ser os registos normalizados ou estar em esquemas em estrela? Necessidade de um *Data Warehouse* central ou serve um *Data Warehouse* composto por *Data Marts* devidamente integrados? Deve-se seguir uma abordagem *top-down* ou *bottom-up*?

A metodologia a seguir no âmbito desta dissertação será a de Kimball. Este é um dos métodos mais referenciados em termos académicos e também profissionais.

Kimball *et al.* (Ralph Kimball & Ross, 2003), no seu livro sobre esta metodologia, utiliza metaforicamente o termo “peças de xadrez”, dizendo que a implementação de sistemas de *Data Warehouse* são como jogos de xadrez, onde é preciso saber qual a importância estratégica das peças e como manuseá-las para se obter o sucesso desejado.

Neste contexto metafórico, utilizado pelos autores do livro, as “peças” estão associadas às atividades os elementos básicos necessários para a implementação e sucesso destes sistemas, que tal como o jogo de xadrez, necessita que quem o implementa tenha que pensar no que pode acontecer no futuro.

Seguidamente, serão apresentadas as atividades constituintes da metodologia de Kimball que podem ser visualizadas na Figura 15 e que serão explicadas e detalhadas posteriormente

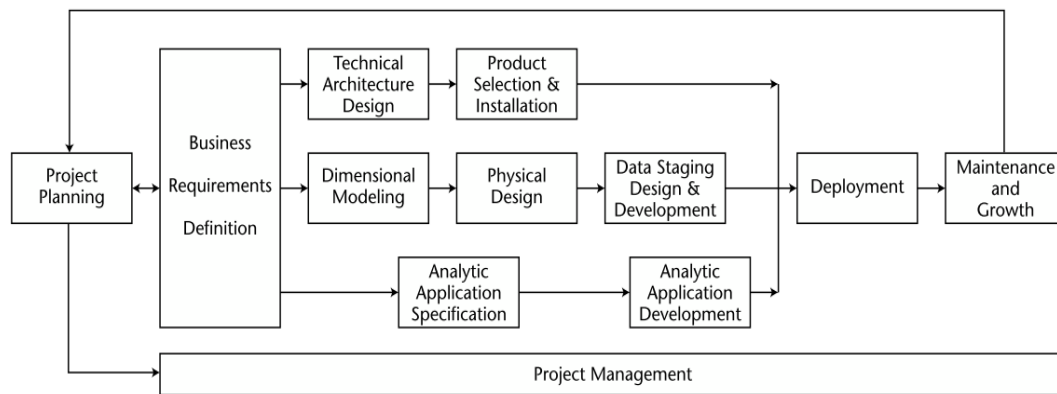


Figura 15 - Metodologia de Kimball (retirado de (Ralph Kimball & Ross, 2003))

Na primeira atividade, planeamento do projeto, é avaliado e identificado o nível de maturidade da organização para adquirir um sistema de *Data Warehouse*, quais os riscos associados, a motivação e cultura organizacional. Como *output* desta atividade pretende-se obter uma versão provisória do âmbito do projeto, os critérios de sucesso e o retorno do investimento no mesmo.

Após esta primeira atividade, e tendo sido o projeto aprovado, segue-se a atividade de definição dos requisitos de negócio onde se pretende identificar os requisitos iniciais para o sistema a desenvolver. Estes requisitos devem ser recolhidos por intermédio de reuniões e entrevistas aos executivos, gestores, analistas e técnicos de TI. Deve, nesta fase, ter-se já a preocupação de identificar quais as fontes de dados à disposição.

Depois de realizadas estas especificações iniciais, encontra-se aberto o caminho para a realização de três atividades que irão decorrer em simultâneo, sendo elas a definição da arquitetura, a modelação multidimensional e o desenho da arquitetura aplicacional de BI.

Na atividade de definição da arquitetura, pretende-se modelar as várias áreas de retenção de dados, o esquema de conteúdo do repositório do *Data Warehouse* e o ambiente aplicacional. São também definidos nesta atividade os critérios para a seleção das ferramentas que implementem esta arquitetura, tais como *hardware*, bases de dados, sistemas de carregamento e ferramentas de acesso aos dados. Esta atividade dá então origem à seleção, propriamente dita, das ferramentas bem como a sua instalação.

A atividade de modelação multidimensional desenrola-se depois da junção da análise feita às fontes de dados e aos requisitos de negócio. Pretende-se identificar qual a granularidade das tabelas de factos, dimensões associadas, atributos e as hierarquias. Este conjunto de

atividades termina com o desenvolvimento do mapeamento dos dados desde a fonte até ao destino.

Na atividade do desenho da arquitetura aplicacional de BI é onde se projeta como e quais as aplicações que serão utilizadas para aceder ao sistema de *Data Warehouse*.

Após estas três atividades, desenvolve-se o modelo físico onde se identificam os índices e as tabelas agregadas.

Seguidamente desenrola-se a atividade de ETL, conceção e desenvolvimento, onde se define o processo de carregamento e refrescamento. É importante definir estratégias que garantam a conformidade e integridade das tabelas de dimensão quando estas são partilhadas por vários esquemas.

Na atividade de entrega é onde se disponibiliza o sistema que foi construído, garantindo que este está devidamente carregado com os dados necessários e que as aplicações de acesso aos dados estão prontas para o utilizador final. Uma das tarefas mais importantes a delinear é a estratégia de formação e suporte aos utilizadores.

Por fim, a última atividade, a de manutenção e crescimento visa criar uma equipa de técnicos que devem gerir e definir alterações importantes bem como apresentar novas propostas de desenvolvimento ao sistema desenvolvido.

2.2.7. Arquiteturas e Componentes de um Sistema de DW

Um *Data Warehouse* é um repositório físico, onde os dados relacionais são devidamente organizados e limpos para depois serem disponibilizados num formato padronizado para que seja possível obter análises relevantes destinadas a apoiar a tomada de decisão.

Para a construção de um sistema de *Data Warehouse* é necessário identificar os requisitos que este necessita, pois nem todas as implementações carecem das mesmas necessidades. Isto remete que para diferentes sistemas de *Data Warehousing* existem diferentes estruturas. Enquanto umas, por exemplo, podem ter apenas um repositório de dados operacional, outras podem apresentar vários *Data Marts*. Algumas podem apresentar um pequeno número de fontes de dados, enquanto algumas podem apresentar dezenas delas (1keydata, 2015).

Um sistema de *Data Warehouse* não se restringe apenas à construção do *Data Warehouse* como repositório de dados, mas sim à construção de um sistema que integra diversos componentes fundamentais (Figura 16). Por norma estes sistemas são constituídos pelas seguintes componentes (Sá, 2010):

- Fontes de dados;
- Processo de extração, transformação e carregamento (ETL);
- Repositórios de dados;
- Mecanismos de análise da informação;

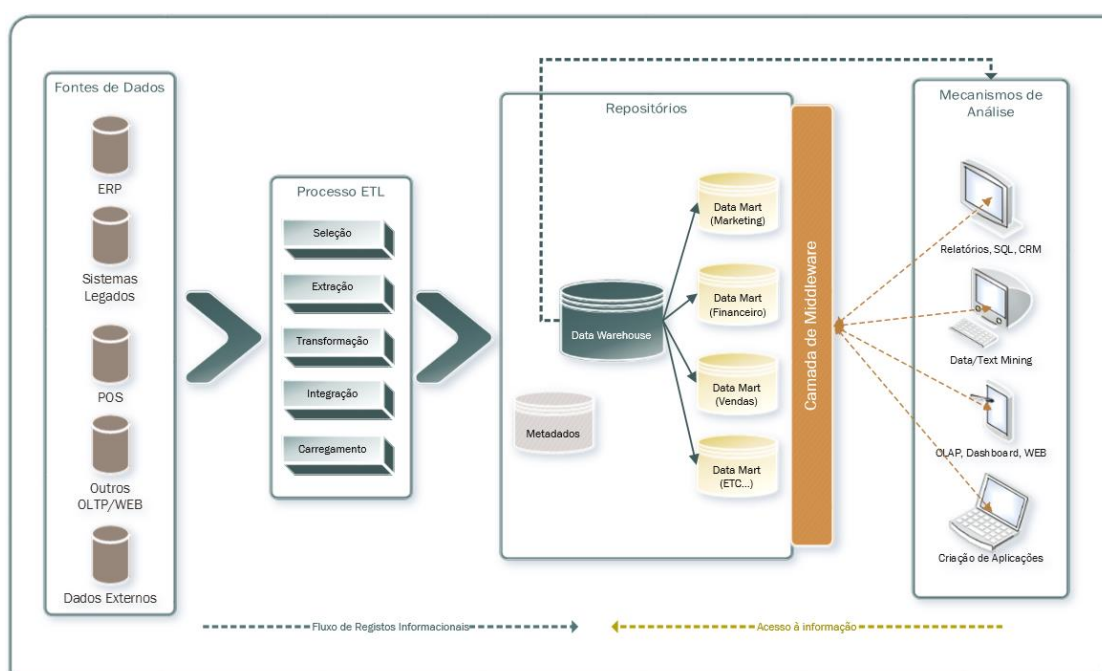


Figura 16 - Data Warehouse Framework (adaptado de (Turban et al., 2008))

Fontes de dados

As fontes de dados podem ser analisadas quanto à sua origem e tipo. Estas podem ser internas quando são provenientes das atividades diárias da organização ou externas quando a origem é externa à organização. Na Tabela 5 são apresentados dois exemplos de fontes de dados que também podem conter dados estruturados e semiestruturados.

Tabela 5 - Exemplo de tipo/fonte dos dados (adaptado de (Negash, 2004))

Tipo \ Fonte	Interna	Externa
Estruturada	ERP	CRM
Semiestruturada	Processo de Negócio	Notícias/ emails

Enquanto nos dados estruturados é possível verificar que estes encaixam perfeitamente numa estrutura bem definida, os dados semiestruturados é possível reconhecer que apenas alguns apresentam alguma estrutura como por exemplo os *emails* (Negash, 2004).

Processo ETL

Um DW é utilizado como instrumento para a tomada de decisão, para tal é importante que os dados que são armazenados estejam corretos (Chaudhuri & Dayal, 1997). As ferramentas de extração, transformação e carregamento (ETL) são uma das componentes fundamentais para detetar e corrigir anomalias nos dados, tornando-os homogéneos para que posteriormente sejam carregados para o *Data Warehouse*. Os dados provenientes das diversas fontes são extraídos através de mecanismos adequados que depois serão armazenados numa *Staging Area* (DSA). É nessa área que os dados passam pelo processo de transformação e limpeza, ficando devidamente corrigidos e homogeneizados para que depois transitem dessa área para o *Data Warehouse* através do processo de carregamento (Ralph Kimball & Ross, 2003; Santos & Ramos, 2006).

Repositórios

Nesta componente, a organização pode optar por uma arquitetura baseada num *Data Warehouse* organizacional, *Data Marts* independentes ou por uma implementação de *Data Marts* dependentes. Se optar pela conceção de um *Data Warehouse* organizacional, todos os dados da organização serão integrados e o DW será alimentado por um ou mais sistemas operacionais. No caso dos *Data Marts* independentes é feita a integração de subconjuntos de subconjuntos de dados da organização como forma de satisfazer as necessidades de unidades específicas da organização. Tal como no *Data Warehouse*, os *Data Marts* independentes são alimentados a partir de sistemas operacionais ou fontes externas. No caso de uma implementação com *Data Marts* dependentes a diferença reside na forma como estes são alimentados, pois estes recorrem ao *Data Warehouse* organizacional. Tanto na construção de um *Data Mart* como na construção de um *Data Warehouse* são usados os mesmos princípios e metodologias (Santos & Ramos, 2006).

Mecanismos de Análise

A última componente, que representa os mecanismos de visualização e análise da informação integra um conjunto de ferramentas de análise que possibilitam o acesso aos dados armazenados no repositório. Estas consultas aos dados podem ser realizadas de várias maneiras, como por exemplo: linguagem de consulta estruturada (SQL), geração de relatórios, ferramentas OLAP, *customer relationship management* (CRM), navegadores web ou identificar padrões e tendências nos dados através de técnicas de *Data Mining* (Sá, 2010; Santos & Ramos, 2006).

Turban (2008) afirma que existem dez fatores que afetam potencialmente o processo de seleção de uma arquitetura de *Data Warehouse*. Esses fatores podem afetar cada um dos componentes apresentados anteriormente e como tal dar origem a arquiteturas diferentes. Os fatores que Turban (2008) menciona são:

- 1) Informação interdependente entre várias unidades organizacionais;
- 2) Necessidade de obter informação da administração;
- 3) Necessidade urgente de um *Data Warehouse*;
- 4) Natureza das tarefas do utilizador final;
- 5) Constrangimentos sobre os recursos disponíveis;
- 6) Visão estratégica do *Data Warehouse* antes da sua implementação;
- 7) Compatibilidade com os sistemas já existentes;
- 8) Ter uma equipa de TI com capacidades;
- 9) Questões técnicas;
- 10) Fatores sociais e políticos.

Quando uma organização pensa em implementar um sistema de *Data Warehouse*, esta pode optar por fazê-lo de três formas diferentes no que toca à componente de repositórios de dados (Figura 17). A primeira alternativa é através de um *Data Warehouse* organizacional com todos os dados da organização integrados e armazenados no mesmo repositório. A segunda e terceira alternativa estão associadas à implementação de *Data Marts* que podem ser dependentes ou independentes respetivamente (Santos & Ramos, 2006).

Na Figura 17 é apresentada a arquitetura proposta por Han e Kamber que é tomada como exemplo no livro de Santos e Ramos (2006) pela sua abrangência e diversidade de conceitos.

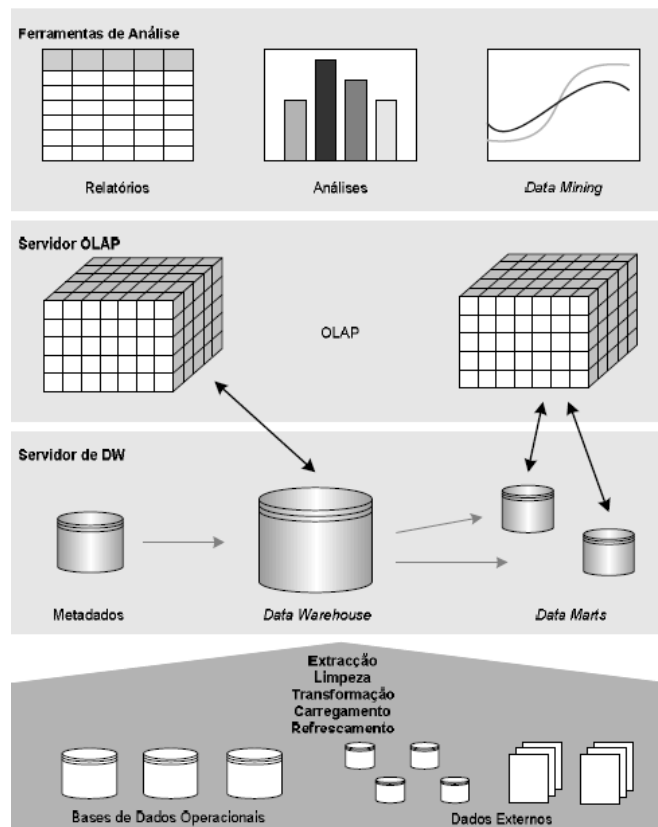


Figura 17 - Arquitetura de apoio ao *Business Intelligence* (retirado de (Santos & Ramos, 2006))

A arquitetura da Figura 17 apresenta duas camadas distintas. A primeira camada diz respeito às fontes de dados e respetivo processo ETL com a extração, limpeza, transformação, carregamento e refrescamento dos dados. A segunda camada é composta por três níveis: o nível do servidor de DW, o nível do servidor OLAP e por fim o nível das ferramentas de análise.

As arquiteturas de DW têm vindo ao longo do tempo a evoluir. Esta evolução é motivada não só pela necessidade de lidar com dados não-estruturados, mas também pela exigência cada vez maior ao desempenho das bases de dados tradicionais. Um novo conceito, designado de *Big Data*, tem surgido nos últimos anos e tem levantado algumas questões no que toca ao modo como vemos e lidamos com a informação. *Big Data* não é algo que seja só simplesmente grande quantidade de dados, tendo sido definido pela Gartner (2013a) como algo muito volumoso, de alta velocidade e de uma grande variedade de dados. O objetivo é permitir uma boa gestão da informação para se atingir o quarto “V” de valor. O conceito de *Big Data* e as tecnologias associadas ao *Data Warehouse* e *Business Intelligence* estão a transformar por completo o mundo dos negócios (GRT, 2015).

2.2.8. Análise de Dados Espaciais

As bases de dados multidimensionais (MDB) são utilizadas pelos Sistemas de Apoio à Decisão e por Sistemas de Informação Geográfica (SIG). Estes SIG têm por objetivo localizar dados espaciais sobre a superfície terrestre estudando a sua evolução ao longo do tempo (Taniar, 2009). Na prática, um SIG consiste num conjunto de camadas, em que cada uma apresenta uma característica bidimensional particular, que podem ser sobrepostas umas em cima das outras (Negash, 2004). Segundo Zimányi (2003), estes sistemas por si só não são eficientes para serem aplicados no apoio à tomada de decisão, pois na maioria das vezes apresentam interfaces de consulta muito complexas com tempos de resposta muito lentos.

Com a ampla utilização dos sistemas de *Data Warehousing* e as ferramentas OLAP associadas para a realização de análises de dados não espaciais, surgiu a necessidade de estender este conceito para que fosse possível analisar dados espaciais (Zimányi et al., 2003). Um *Data Warehouse* espacial é um repositório com capacidade de analisar não só dados tabulares mas também dados espaciais. Estes últimos são representados através de mapas para facilidade de compreensão e perceção por parte dos agentes decisores (Morgado, 2013). Esta variante de *Data Warehouse* foi criada porque apesar de os *Data Warehouses* convencionais terem a capacidade de lidar com vários tipos de dados, estes não têm a capacidade de processar dados espaciais não sendo possível, em alguns cenários, extrair toda a informação armazenada pelas organizações.

Os dados espaciais são referidos como entidades espaciais ou geográficas tal como pontos, linhas, polígonos e superfícies. A utilização de dados espaciais no processo de tomada de decisão permite a visualização dos dados no espaço o que permite aos utilizadores identificar padrões que são difíceis de descobrir de outra forma. Este tipo de dados é representado normalmente por objetos geográficos, ou seja, objetos localizados na superfície da terra, como por exemplo cidades ou localidades (Wang, 2008). Estes dados não podem ser tratados da mesma forma que os restantes, uma vez que possuem características específicas. Os mesmos dados espaciais podem ser analisados e tratados de diferentes formas e tamanhos (Taniar, 2009).

O agrupamento de dados espaciais é idêntico ao processo de agrupamento tradicional. Tal como acontece no agrupamento tradicional, a função do agrupamento espacial é juntar os objetos de acordo com as suas semelhanças, porém o que diferencia este tipo de agrupamento

é o facto de ter de se considerar não as semelhanças entre objetos, mas também a posição que estes ocupam no espaço (Dinis, 2011).

Para que fosse possível analisar estes dados, a solução passou pela combinação dos pontos fortes dos SIG com os pontos fortes das ferramentas OLAP, surgindo desta forma as tecnologias SOLAP (*spatia*/OLAP). A tecnologia SOLAP pode ser definida como uma plataforma visual desenvolvida para a análise/exploração fácil de dados espaciais e temporais. Esta segue uma abordagem multidimensional composta por níveis de agregação visíveis de forma cartográfica e tabular (Rivest, Bédard, Proulx, & Nadeau, 2003).

Este tipo de ferramentas suportam as estruturas de bases de dados multidimensionais descritas na secção 2.2.3, porém as capacidades de manipulação dos dados espaciais levaram a inclusão de três tipos de dimensões e medidas espaciais que podem ser usadas.

Dimensões Espaciais

Na construção de um *Data Warehouse* espacial, a partir de um modelo multidimensional, deve-se considerar para além da semântica usual e das dimensões temporais, a existência de três tipos de dimensões espaciais (Bédard, Merrett, & Han, 2001):

- Dimensão Espacial Não geométrica: Esta dimensão é constituída unicamente por dados não geométricos, ou seja, dados apenas nominais. O facto de estes não serem dados geométricos não implica que não seja possível retirar informação que permita a localização no espaço (por exemplo, nomes de cidades ou localidades).
- Dimensão Espacial Geométrica para não geométrica: Este tipo de dimensão é composto por dados geométricos, que a partir de uma generalização (nível hierárquico mais baixo para o mais alto), tornam a dimensão geométrica numa dimensão não geométrica.
- Dimensão totalmente Geométrica: Nesta dimensão os dados, para qualquer nível granular são sempre geométricos, ou seja, mesmo que se procedam a generalizações dos dados estes mantêm a sua tipologia não alterando a dimensão.

Medidas espaciais

Para além das medidas que existem num *Data Warehouse* tradicional, é possível definir, no contexto do modelo multidimensional, três tipos de medidas espaciais (Bédard et al., 2001):

- Medidas numéricas: Constitui um tipo de medida que contém apenas dados numéricos;
- Medidas espaciais: É uma medida que contém um conjunto de pontos (coordenadas) para um objeto espacial;
- Medidas completamente espaciais: Constitui a combinação de um valor numérico e objeto espacial associado.

Tal como nas arquiteturas de sistemas OLAP, que são compostas por uma base de dados multidimensional estruturada, as arquiteturas dos sistemas SOLAP são compostas por bases de dados espaço-temporais multidimensionais estruturadas. Estas bases de dados têm também por objetivo armazenar os dados das tabelas de dimensão e de factos que podem incluir componentes geométricas (Rivest et al., 2003).

2.3. Carta Educativa

“Nas últimas décadas, Portugal tem feito um enorme esforço de qualificação escolar da população, que se traduziu em progressos substanciais em matéria de educação. Contudo, o país continua a apresentar um défice estrutural de formação e qualificação da população que exige uma aposta clara e persistente na resolução dos problemas que têm impedido a convergência com os atuais padrões da União Europeia, nomeadamente os níveis de insucesso e abandono escolares e o défice de qualificações da população ativa” (M. Educação, 2007).

Para além de várias medidas que foram implementadas, como por exemplo a escolaridade obrigatória, é preciso agora colocar as escolas no centro da política educativa, qualificando-as e melhorando o seu funcionamento para que os alunos sejam beneficiados.

O sistema educativo em Portugal está dividido em quatro escalões. Na base do sistema educativo está o ensino pré-escolar seguindo-se o ensino básico, secundário e por fim o ensino superior tal como é possível verificar na Figura 18.

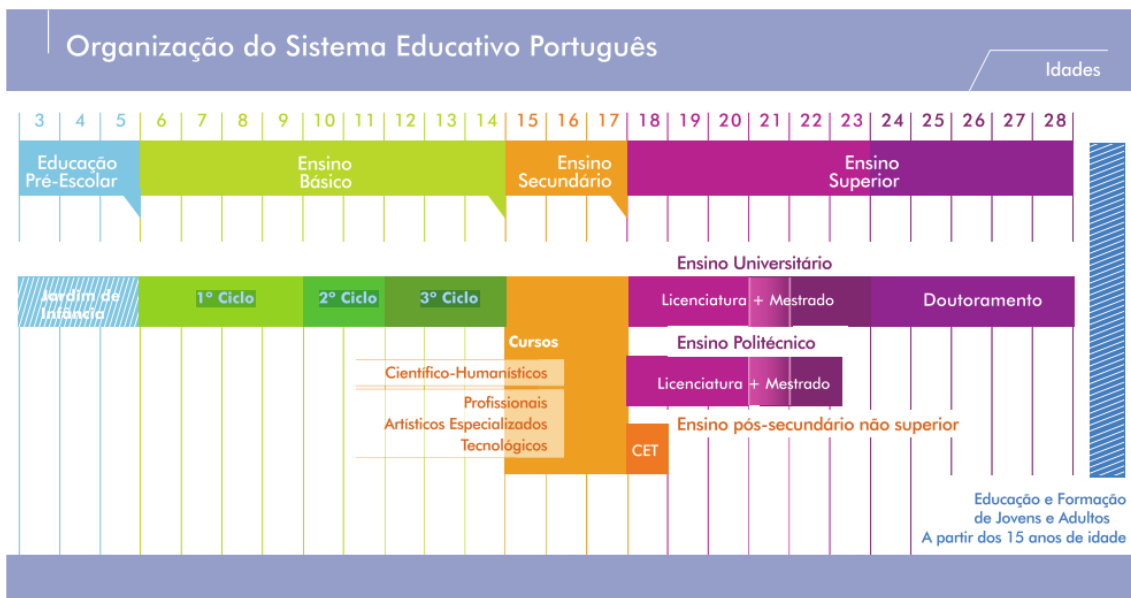


Figura 18 - Organograma do Sistema Educativo em Portugal (retirado de (M. Educação, 2007))

A educação pré-escolar não é de cariz obrigatório, destinando-se a crianças com idades entre os 3 anos e a entrada na escolaridade obrigatória. No ensino básico inicia-se a escolaridade obrigatória, com uma duração de nove anos e que se divide em três ciclos. O 1º ciclo refere-se ao período entre o 1º e 4º anos de escolaridade, o segundo entre o 5º e 6º ano e o 3º ciclo engloba os anos escolares desde o 7º ao 9º ano.

No ensino secundário pode verificar-se que existem formas diferenciadas, com cursos orientados tanto para prosseguir nos estudos como cursos profissionais que permitem ao aluno especializar-se numa área mais focada para o mundo de trabalho. O ensino secundário tem um período de três anos para a sua conclusão e está dividido nas seguintes áreas de ensino:

- Cursos Científico Humanísticos;
- Cursos Profissionais;
- Cursos artísticos especializados;
- Cursos Tecnológicos;

O ensino pós-secundário não superior está organizado em cursos de especialização tecnológica (CET), que têm por objetivo oferecer a possibilidade de obter o nível 4 de formação profissional. Por fim, o ensino superior, que já não faz parte da escolaridade obrigatória e que se encontra estruturado segundo o que foi definido na Declaração de Bolonha, é ministrado em institutos

politécnicos e universidades, sejam elas de natureza pública, privada, cooperativa ou concordatária.

“A carta educativa é, a nível municipal, um instrumento de planeamento e ordenamento prospetivo de edifícios e equipamentos educativos a localizar no concelho, de acordo com as ofertas de educação e formação que sejam necessárias satisfazer, tendo em vista a melhor utilização dos recursos educativos, no quadro de desenvolvimento demográfico e socioeconómico de cada município” (M. Público, 2003).

Inicialmente a carta educativa era designada por Carta Escolar (Decreto-Lei 159/99), tendo sido instituído o nome atual pelo Decreto-lei 7/2003 que deve assentar numa dimensão prospetiva que possibilita responder à procura da educação a nível municipal tendo em conta a política educativa nacional. Até ao decreto-lei em vigor, a Carta Escolar era entendida apenas como um documento em que se registavam os edifícios escolares existentes bem como os que faltavam construir.

A nível nacional, a grande maioria das Cartas Educativas em vigor tiveram o seu período de homologação entre 2006 e 2008 (V.Educação, 2015).

Relativamente à gestão da rede escolar por parte dos municípios, a carta educativa é entendida como o principal instrumento de apoio à decisão por parte de quem tem a responsabilidade de gerir os caminhos da educação e formação num determinado território. Trata-se de um instrumento de planeamento que, para além de analisar a realidade existente, em termos educativos, agrega também aspetos sociais visando desta forma detetar as respostas mais eficientes e adequadas às necessidades da educação (CMGuimarães, 2006). De acordo com o artigo 11º do decreto-lei 7/2003, este instrumento de planeamento pretende assegurar a adequação da rede escolar dos estabelecimentos pré-escolares, ensino básico e secundário para que a cada momento se garantam as ofertas educativas necessárias.

A carta educativa, enquanto instrumento de gestão, não pode ser considerado como um documento acabado, mas sim como uma (re)configuração da rede escolar, que conforme as evoluções socioeconómicas e do sistema educativo que vão ocorrendo durante os anos, necessita de uma revisão e adequação constante.

Objetivamente, a carta educativa deverá permitir aos responsáveis pela gestão escolar (Oliveira, Coragem, & Martins, 2000):

- Orientar a expansão do sistema educativo de um território em função do desenvolvimento económico e sociocultural;
- Orientar a tomada de decisão quanto à construção ou encerramento de escolas bem como a reconversão e adaptação do parque, otimizando a funcionalidade da rede existente e respetiva expansão;
- Permitir a definição de prioridades;
- Permitir a otimização da utilização dos recursos consagrados à educação;
- Evitar ruturas e desadequações da rede escolar em relação à realidade social e desenvolvimento urbanístico existente.

Para que estes objetivos e este tipo de análises sejam conseguidos, a carta educativa deverá conter, para além de outros pontos que eventualmente sejam de interesse, alguns capítulos fundamentais como:

- Enquadramento Legislativo;
- Caracterização Socioeconómica;
- Diagnóstico estratégico (Evolução do sistema educativo);
- Projeção de cenários de desenvolvimento (Reconfiguração/Reordenamento);
- Proposta de intervenção;

Sendo a educação uma área bastante abrangente, a elaboração da carta educativa diz respeito não só a cada município, mas também a uma série de intervenientes que têm a sua cota de participação no desenvolvimento deste documento. Na Tabela 6 são apresentados os intervenientes no processo de construção bem como o tipo de participação que estes têm nos conteúdos em que a carta educativa se baseia.

Tabela 6 - Intervenientes no processo de elaboração da Carta Educativa (Adaptado de(Oliveira et al., 2000))

Atribuições	Câmara Municipal	Conselho Local de Educação	Comissão de Coordenação Regional	M. Educação	
				Serviços Centrais	Direção Regional de Educação
Análise da Política Educativa				3	
Definição de Normativos Técnicos				3	
Adequação dos critérios de reordenamento a nível regional	2		3		1
Caracterização Socioeconómica					
Atividades Económicas e a sua Localização	3	1	2		
Demografia	3		1		
Hierarquização dos Aglomerados	3		2		
Caracterização e Evolução do Sistema Educativo					
Procura da educação e do ensino	3	1			2
Oferta (Parque Escolar)	3	1			2
Diagnóstico da Situação Escolar	3	1			2
Reconfiguração/Reordenamento da Rede	3	1			2
Monitorização/Avaliação	3			2	2

Legenda: **1** – Participante; **2** – Participante/Executor; **3** - Executor

A carta educativa do concelho de Guimarães, elaborada e homologada em 2006, foi desenvolvida em conformidade com as normas definidas, apresentando cinco capítulos essenciais. O primeiro está relacionado com o enquadramento geográfico à escala nacional, regional e local bem como a caracterização socioeconómica do território. No segundo capítulo é feita uma caracterização demográfica da população do concelho onde é possível, através de um conjunto de indicadores, avaliar a evolução da população residente. O terceiro capítulo é onde se apresenta a análise detalhada ao sistema educativo do concelho de Guimarães. Através de um conjunto de indicadores, traduzidos em quadros e figuras, é possível analisar e avaliar o sistema escolar (abandono escolar, aproveitamento, taxa de analfabetismo, entre outros) bem como todos os equipamentos escolares, quer do ponto de vista da mobilidade e proveniência dos alunos quer da sua caracterização. O quarto capítulo diz respeito às

previsões, ou seja, à projeção de cenários futuros tendo em conta as análises efetuadas. Por fim, no quinto capítulo são feitas as propostas de intervenção e medidas a tomar tendo em conta as previsões efetuadas no capítulo anterior.

3. Caracterização das Tecnologias e dos Dados

Neste capítulo serão apresentadas as tecnologias utilizadas e que estão disponíveis para o suporte e desenvolvimento do protótipo do sistema de BI que se propõe nesta dissertação de mestrado. Para além das tecnologias utilizadas e a forma como as mesmas são integradas na arquitetura proposta, este capítulo apresenta a caracterização dos dados disponíveis bem como a análise à sua qualidade.

3.1. Tecnologias a utilizar

Nesta secção, é realizada a análise das ferramentas que serão utilizadas, bem como descritas as funcionalidades integrantes de cada uma para cumprir com as necessidades deste desenvolvimento. Espera-se que na condução desta dissertação, em todas as suas etapas e, especialmente, na construção do protótipo da solução do sistema de BI sejam utilizadas ferramentas que contribuam para que a solução consiga responder aos objetivos definidos com qualidade e eficácia.

Para a construção do protótipo da solução foram utilizadas ferramentas Microsoft, tais como o *SQL Server Business Intelligence 2014*, o *Microsoft Excel 2013* e o *Tableau* por serem ferramentas cuja licença é disponibilizada pela Universidade do Minho e pelo conhecimento já adquirido das mesmas. Foi também utilizada a ferramenta de desenvolvimento *Talend Open Studio* bem como o *Microsoft PowerBI* para se obter um conjunto de análises mais completas. Seguidamente, será feita a descrição de cada uma das ferramentas e explicadas quais as suas funcionalidades bem como a parte do projeto onde foram utilizadas.

- **Microsoft SQL Server Business Intelligence 2014 (Microsoft, 2015)**

O *SQL Server Business Intelligence* proporciona uma plataforma abrangente, que permite a criação e implementação de soluções de BI com funcionalidades avançadas de análise e geração de relatórios. É possível efetuar a gestão e controlo das análises dos dados através de *dashboards* no *SQL Server* e no *SharePoint* que monitorizam a atividade e a utilização do armazenamento de dados dos utilizadores finais, recolhendo métricas de desempenho dos servidores. Os utilizadores desta plataforma podem obter análises preditivas através da ferramenta *Microsoft Excel*, com os suplementos *SQL Server Data Mining* para *Microsoft Office*.

Os recursos do *SQL Server* são parte integrante da plataforma *Microsoft Business Intelligence* e incluem os componentes de *Analysis Services*, *Integration Services*, *Master Data Services*, *Reporting Services* e várias aplicações cliente úteis para a criação e análise de dados (Tabela 7). Todos os componentes integrantes desta ferramenta conferem-lhe um conjunto de capacidades como:

- Exploração e visualização rápida dos dados;
- Análises de elevado desempenho;
- Mecanismos de extração, transformação e carregamento dos dados;
- Gestão de dados mestre;
- Funcionalidades básicas de disponibilidade, segurança e gestão de dados.

- **Talend Open Studio for Data Integration 5.6.2.20150508 (Talend, 2015a)**

O *Talend* foi fundado em 2005 com intuito de fornecer *software open source* para integração de dados. Esta é uma ferramenta que oferece uma integração robusta de dados numa arquitetura aberta e escalável que permite ao utilizador um conjunto de mecanismos de extração, transformação e de carregamentos dos dados.

Esta ferramenta funciona como um gerador de código, capaz de produzir scripts de transformação dos dados e programas subjacentes em java através da criação de “*jobs*” e arrastando as várias componentes disponibilizadas. Ao utilizador, o Talend Open Studio for Data Integration oferece a possibilidade de:

- Sincronização ou replicação de vários tipos de bases de dados;
- Mecanismo de ETL para análise;
- Migração de dados;
- Transformações e carregamentos complexos de dados.

É importante referir que foi utilizada a versão 5.6.2.20150508, uma vez que até á data em que foi utilizada era a última versão estável disponibilizada.

- **Talend Open Studio for Data Quality (Talend, 2015b)**

Esta ferramenta é uma componente *open source* do pacote *open studio* do *Talend* para analisar a qualidade de dados. É uma ferramenta que permite avaliar a qualidade dos dados, usando diversos parâmetros, possibilitando tomar decisões no que toca às

ações a aplicar para corrigir possíveis erros. Ao utilizador, o *Talend Open Studio for Data Quality* oferece a possibilidade de:

- Identificar a estrutura das bases de dados bem como a definição de metadados;
- Definição de estatísticas e métricas para análise dos dados;
- Visualização dos resultados através de gráficos;
- Resolver inconsistências através da normalização e duplicação dos dados.

- **Microsoft Excel 2013**

O Microsoft Office Excel é um editor de folhas de cálculo desenvolvido pela Microsoft. Esta é uma ferramenta bastante poderosa com uma interface intuitiva e capacitada, com ferramentas de cálculo e construção de gráficos que fazem desta uma das aplicações mais populares até aos dias de hoje.

Esta ferramenta será utilizada para a criação de *reports e dashboards*. Serão adicionados os recursos *PowerView* e *PowerPivot* que estão disponíveis nas edições do *Office Professional Plus*, *Office 365 Professional Plus* bem como na versão autónoma do Microsoft Excel 2013. O *PowerPivot* é um *add-in* do Excel que permite realizar análises de dados.

Com a integração destes componentes é possível criar relatórios e visualizações analíticas através do *PowerView* com tabelas e gráficos interativos que permitem explorar e apresentar os dados visualmente no Excel. Também é possível, com a componente *Power Map*, explorar dados geoespaciais através de experiências em mapas 3D.

- **Tableau (Tableau, 2015)**

O *Tableau* é uma ferramenta que permite realizar análises complexas através de um conjunto de *dashboards* com a possibilidade de integração com a maioria das plataformas de BI. Esta ferramenta também permite realizar várias visualizações e análises geoespaciais (Gartner, 2014).

O *Tableau* é uma ferramenta que permite que o utilizador compreenda, através de uma grande diversidade de visualizações, os seus dados. Esta foi a ferramenta líder de mercado no ano de 2014, atribuída pela Gartner (2014) na sua análise anual aos fornecedores de *Business Intelligence* e plataformas de análise existentes (Figura 19).








Figura 19 – Quadrante Mágico de plataformas de Análise e *Business Intelligence* (retirado de (Gartner, 2014))

Na Figura 19 é ilustrado o quadrante mágico elaborado pela *Gartner* e sendo bem visível que esta se destaca no quadrante dos líderes de mercado pela sua qualidade reconhecida.

Na Tabela 7 são apresentadas as tecnologias a utilizar bem como a distribuição das suas componentes pelas necessidades da implementação deste projeto. A tabela encontra-se dividida pelas componentes de gestão de base de dados, análise exploratória dos dados, processo de extração, limpeza, transformação e carregamento, servidor OLAP e, por fim, *reports* e *dashboarding*. São também descritas as funcionalidades/características das ferramentas a utilizar.

Tabela 7 - Matriz das Tecnologias selecionadas

Etapa	Ferramenta	Componente	Funcionalidades
Sistemas de Gestão de Base de Dados 	❖ Microsoft SQL Server 2014	✓ Management Studio	<ul style="list-style-type: none"> ➤ Acesso e integração dos dados ➤ Gestão das bases de dados ➤ Escalabilidade e desempenho ➤ Segurança
Análise da Qualidade dos dados 	❖ Talend	✓ Open Studio Data Quality	<ul style="list-style-type: none"> ➤ Definição de metadados ➤ Definição de métricas de análise ➤ Visualização gráfica
	❖ Tableau	-	<ul style="list-style-type: none"> ➤ Suporte a grande quantidade de dados ➤ Interativo ➤ Novas formas de Visualização ➤ Análises Geo-Espaciais
Processo ETL 	❖ Talend	✓ Open Studio Data Integration	<ul style="list-style-type: none"> ➤ Integração de Dados ➤ Arquitetura aberta e escalável ➤ Geração de código native ➤ Job Designer ➤ Modelador de Negócio
Cubos OLAP 	❖ Microsoft SQL Server Business Intelligence 2014	✓ Analysis Services	<ul style="list-style-type: none"> ➤ Serviços espaciais e de localização ➤ Ferramentas de Desenvolvimento ➤ Tratamento de dados ➤ Flexível e fácil de utilizar
Reports e Dashboarding 	❖ Microsoft Excel 2013	<ul style="list-style-type: none"> ✓ Power View ✓ Power Pivot ✓ Power Map 	<ul style="list-style-type: none"> ➤ Exploração e visualização rápida dos dados ➤ Interativo ➤ Combinação de dados de qualquer origem ➤ Análise Geo-espacial ➤ Integração com várias plataformas
	Microsoft PowerBI	-	<ul style="list-style-type: none"> ➤ Rápido, simples, interativo e de fácil compreensão ➤ Partilha das análises efetuadas ➤ Aplicação para dispositivos móveis ➤ Análise Geo-espacial ➤ Integração com várias plataformas

3.2. Proposta de Arquitetura do Sistema

Depois de revista a literatura acerca dos conceitos de *Business Intelligence*, *Data Warehousing*, componentes e arquiteturas associadas, é possível nesta fase propor uma arquitetura do sistema que se propõe desenvolver. Na Figura 20 é apresentada a arquitetura proposta.

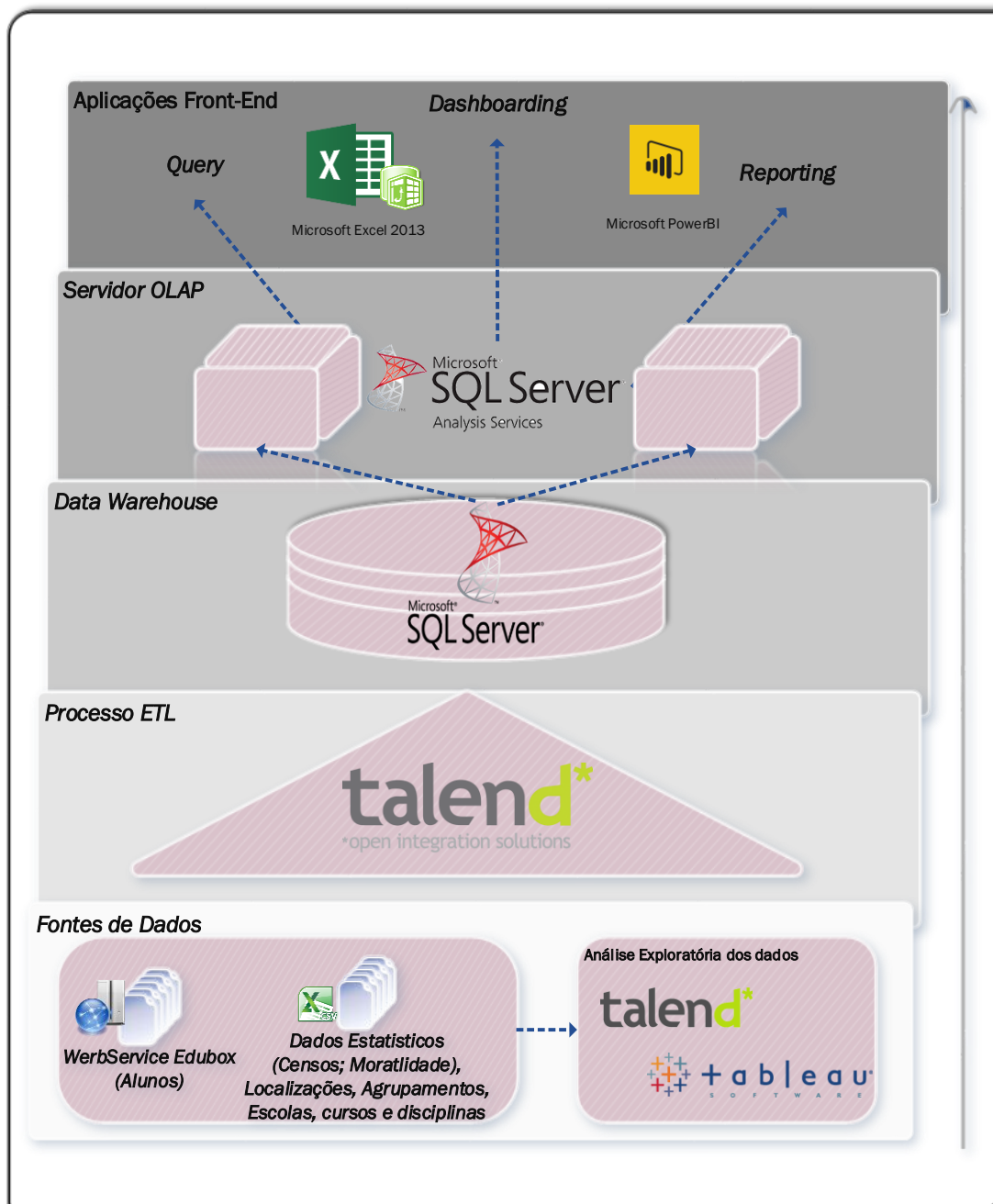


Figura 20 - Arquitetura da Infraestrutura Tecnológica

Analisando a arquitetura proposta, é possível observar que a mesma reflete a abordagem conceptual e tecnológica do sistema pretendido, sendo esta composta por cinco níveis.

O primeiro nível representa as fontes de dados que alimentarão o sistema e que podem ser de dois tipos. O primeiro diz respeito aos dados operacionais provenientes do *WebService Edubox*, que são referentes aos alunos provenientes da plataforma *Edubox* (<http://www.edubox.pt/>) para a gestão municipal dos alunos do pré-escolar e do 1º ciclo de escolaridade. Por outro lado, como fontes de dados externas o sistema recebeu, por exemplo, os dados estatísticos do número de nados-vivos e mortos e dos censos (<http://censos.ine.pt>), bem como de outras fontes que não estão ligadas diretamente com as instituições de ensino. Estas fontes de dados contêm dados estruturados.

Quando se fala num projeto de *Business Intelligence*, e se cruzam os conceitos de *Data Warehouse*, é natural que se associe a eles as técnicas de extração, limpeza, transformação e carregamento dos dados. Este processo corresponde ao segundo nível da arquitetura proposta que diz respeito ao ETL, onde se pretende corrigir todas as anomalias nos dados, tornando-os homogéneos para que posteriormente possam ser carregados para o *Data Warehouse*. A plataforma utilizada para o desenvolvimento nesta fase será o *Talend Open Studio for Data Integration*, pois incorpora um conjunto de funcionalidades que cumprem com as necessidades desta fase.

Após todo o tratamento aos dados, estes serão carregados para o *Data Warehouse* passando assim para o terceiro nível da arquitetura da Figura 20. A componente *Management Studio do SQL Server Business Intelligence* será a responsável pela gestão das bases de dados deste sistema. Depois de serem devidamente catalogados e carregados os dados, é possível passar ao próximo nível com criação dos cubos OLAP para que posteriormente se façam as análises pretendidas com o auxílio da componente de *Analysis Services*.

Por último, após todos os processos descritos anteriormente, será possível visualizar os dados através de *Dashboards*, *query's* e *reports* utilizando para isso as aplicações cliente Microsoft Excel e Microsoft PowerBI. Os mecanismos de visualização e análise da informação, como já referido anteriormente, possibilitam o acesso aos dados armazenados no repositório,

¹ A *Edubox* presta serviços à Câmara Municipal de Guimarães com a sua plataforma SIGA (Sistema Integrado de Gestão e Aprendizagem) para a gestão de alunos do pré-escolar e do 1º ciclo de escolaridade.

componente importante para que a organização possa ter uma visão mais abrangente quanto ao negócio através de indicadores e regras de negócio.

De uma forma geral, esta plataforma deve, de forma rápida e automatizada, permitir aos responsáveis pela gestão da rede escolar, o acesso de forma flexível a toda a informação.

3.3. Caracterização e Análise Exploratória das Fontes de Dados

Nesta secção, pretende-se mostrar a qualidade dos dados fornecidos e que serviram de base para o desenvolvimento do protótipo do sistema de *Business Intelligence*, capaz de permitir um conjunto alargado de análises sobre a realidade associada à “carta educativa” no conselho de Guimarães. São, então, apresentadas as fontes de dados disponibilizadas para a execução deste projeto, bem como alguns dos resultados obtidos da análise efetuada às mesmas. Numa primeira subsecção é feita a caracterização dos dados que serão necessários para realizar o estudo. De seguida, são apresentados os resultados da análise efetuada a um *backup* do software de gestão de alunos presente num dos agrupamentos do concelho. Na terceira, e última subsecção, são apresentados os resultados das análises realizadas aos dados da plataforma *Edubox* que presta serviços ao município de Guimarães. Estas análises exploratórias foram realizadas através das ferramentas *Talend Open Studio for Data Quality* e *Tableau* que contemplaram análises ao nível do número de registos, registos nulos, valores distintos, registos únicos, duplicados e em branco.

3.3.1. Caracterização dos Dados Necessários para o Estudo

Antes de se prosseguir com a análise exploratória dos dados para perceber qual a sua qualidade, importa proceder-se previamente a uma devida caracterização de quais os dados necessários para implementação do projeto.

A carta educativa é um documento que visa não só analisar toda a rede escolar (alunos, escolas e agrupamentos) mas também toda a envolvente socioeconómica da região e dos seus habitantes. Sendo assim, este é um documento que carece de alguma diversidade de fontes de dados para a sua elaboração.

No concelho de Guimarães, a carta educativa encontra-se dividida em três capítulos essenciais para esta dissertação de mestrado. Na Tabela 8 é possível verificar a divisão feita pelos diferentes capítulos tentando desta forma mostrar, para as diferentes análises, quais os dados e as fontes necessárias para a sua concretização neste projeto.

Tabela 8 - Caracterização dos dados necessários

Enquadramento Geográfico		
Indicador	Fonte	Observações
Caracterização Física (Clima)	IPMA	Temperaturas, Precipitação, Humidade, etc..
Caracterização Económica	INE (censos)	Desemprego, Ramos de Atividade, etc.
Setores de Atividade	INE (censos)	-
Rede Viária	CMG	Dados das Estradas do Concelho
Hierarquia dos Aglomerados	CMG	-
Caracterização Demográfica		
População Residente	INE (censos)	Censos de 2001 e 2011
Densidade Populacional	INE	-
Estrutura Etária da População	INE (censos)	Censos de 2001 e 2011
Natalidade e Mortalidade	INE	Nados-vivos e Nados-mortos
Envelhecimento da população	INE	Censos de 2001 e 2011
Sistema Educativo		
Distribuição da População	INE (censos)	Censos de 2001 e 2011
Abandono escolar, saída antecipada e precoce	Escolas/INE	Dados dos Alunos
Sucesso Escolar	Escolas	Dados dos Alunos
Classificação dos Exames Nacionais	DGES	Resultados dos Alunos do Concelho
Proveniência dos Alunos Para a escola	Escolas	Dados dos Alunos
Caracterização das IPSS's	IPSS's	Instituições do Concelho
Distância média percorrida pelos alunos	Escolas	Dados dos Alunos
Caracterização dos Agrupamentos de Escola	CMG/Escolas	-
Caracterização das Escolas	CMG/Escolas	-
Escolas com Excesso de Lotação	Escolas	-
Evolução do número de alunos	Escolas	Dados dos Alunos
Alunos por ciclo de estudo/Agrupamento/Escolas	Escolas	Dados dos Alunos
Alunos com educação especial	Escolas	Dados dos Alunos
Alunos deslocados	Escolas	Dados dos Alunos
Distribuição pela oferta formativa	Escolas	Dados dos Alunos
Alunos subsidiados	Escolas	Dados dos Alunos
Corpo Docente	Escolas	Dados dos Alunos
Infraestruturas de ensino existentes	CMG/Escolas	-

Como é possível verificar pela Tabela 8, para o desenvolvimento de um projeto com as características e necessidades que este apresenta, é notória a diversidade de fontes de dados

que é preciso obter para possibilitar todas as análises pretendidas. De todas as fontes de dados identificadas, as que mais se destacam pela sua importância são as que contêm os dados referentes aos alunos das escolas do concelho de Guimarães. Nas secções seguintes, as duas fontes de dados que foram obtidas, referentes aos dados dos alunos, serão analisadas em detalhe.

3.3.2. Fonte de Dados Software JPM & Abreu

As fontes de dados fornecidas para esta análise são provenientes de um *software* de gestão de alunos desenvolvido pela empresa J.P.M & Abreu, LDA que fornece cerca de 95% das escolas/agrupamentos do ensino público português (Abreu, 2004). No caso do concelho de Guimarães, esta empresa apenas não fornece o seu *software* a um agrupamento de escolas, dos 14 agrupamentos existentes no concelho. O *software* em questão contempla todos os ciclos de estudos sendo que o módulo de Alunos permite verificar avaliações, assiduidade, matrículas, listagem de turmas e de alunos, entre outros.

Neste sentido, houve a necessidade de contactar com a empresa de forma a obter o acesso a uma versão experimental do módulo de alunos e assim, poder visualizar e extrair os conjuntos de dados existentes. Desta forma, foi disponibilizada uma versão experimental do módulo de alunos do Software e cedido, por parte da Câmara Municipal de Guimarães, um *backup* dos dados de um dos agrupamentos de escolas do concelho. Não foi possível o acesso a todos os dados de todos os agrupamentos, pois as bases de dados que suportam este *software* não se encontram centralizadas, mas sim instaladas localmente em cada um dos estabelecimentos de ensino.

Para este estudo foi realizada uma análise detalhada à qualidade dos dados através da plataforma de análise *Talend Open Studio for Data Quality* e o *Tableau*, onde foram avaliados os resultados das análises tendo em conta os seguintes fatores:

- Número de registos;
- Número de registos nulos;
- Número de valores distintos;
- Número de registos únicos;
- Número de registos duplicados;
- Número de registos em branco.

A partir do *backup* disponibilizado pelo município de Guimarães, foi possível extrair as fontes de dados presentes na Tabela 9. A aplicação permite descarregar ficheiros do tipo *Comma-Separated Values* (csv) com os dados referentes aos vários componentes do módulo de alunos.

Tabela 9 - Conjuntos de dados extraídas do Software J.P.M & Abreu

Fonte de dados	Descrição
AlunosGeralTotalFinal.csv	Conjunto de dados principal referente aos dados dos alunos do agrupamento em questão
Cursos2ºe3ºCiclo.csv	Dados dos cursos lecionados no 2º e 3º ciclo do agrupamento
CursosProfissional.csv	Dados dos cursos lecionados no ensino Profissional do agrupamento
CursosRecorrente.csv	Dados dos cursos lecionados no ensino Recorrente do agrupamento
CursosSecundário.csv	Dados dos cursos lecionados no ensino Secundário do agrupamento
Disciplina2ºe3ºCiclo.csv	Dados das disciplinas lecionadas no 2º e 3º ciclo do agrupamento
Disciplinas1ºciclo.csv	Dados das disciplinas lecionadas no 1ºciclo do agrupamento
DisciplinasProfissional.csv	Dados das disciplinas lecionadas no ensino profissional do agrupamento
DisciplinasRecorrente.csv	Dados das disciplinas lecionadas no ensino recorrente do agrupamento
DisciplinasSecundário.csv	Dados das disciplinas lecionadas no ensino secundário do agrupamento
Escolas.csv	Dados das escolas que pertencem ao agrupamento
SituacaoFinalAlunos1º.csv	Situação dos alunos do 1º ciclo quanto ao aproveitamento nos anos letivos em que frequentaram o agrupamento
SituacaoFinalAlunos2º3º.csv	Situação dos alunos do 2º e 3º ciclo quanto ao aproveitamento nos anos letivos em que frequentaram o agrupamento

De uma forma geral, é possível afirmar que as fontes de dados existentes carecem de uma grande falta de valores retirando-lhes desta forma bastante qualidade. Tendo em conta as análises efetuadas, fica presente a ideia de uma constante falta de dados em todos os conjuntos anteriormente referidos. Foi, portanto, registado um elevado número de colunas com valores omissos relevantes para este estudo. Para além disso é impossível determinar informação importante como qual a escola a que um determinado aluno pertence.

Dos conjuntos de dados apresentados na Tabela 9, o mais relevante para este estudo, e que se esperava que possuísse informação com qualidade, é o **AlunosGeralTotalFinal.csv**. Na Tabela 10 é possível ver um excerto dos atributos mais relevantes que foram obtidas. A versão completa com todos os atributos pode ser visualizada em anexo (Anexo A1. Fonte de dados Software JPM & Abreu (AlunosGeralTotal.csv))

Tabela 10 - Atributos da fonte de dados do Software J.P.M & Abreu

Atributo	Descrição
N_Processo	Número do processo do aluno
IdMatricula	Identificador do número de matrícula do aluno
Nome	Nome do aluno
Nacionalidade	Nacionalidade do aluno
Sexo	Género do aluno
Morada1	Morada da residência do aluno
0000	Código postal de 4 dígitos
000	Código postal de 3 dígitos
CodigoPostal	Código postal
N_BI	Número do bilhete de identidade ou cartão de cidadão
ViveComPais	Indicação se o aluno vive com os pais
NecEnsiEspecial	Indicação se o aluno necessita de ensino especial
Subsidiado	Tipo de subsídio que o aluno usufrui
Deficiencia	Indicação se o aluno é portador de algum tipo de deficiência
TransporteUtilizado	Transporte utilizada na deslocação para a escola
ProfissaoPai	Profissão do pai
ProfissaoMae	Profissão da mãe
HabilitacoesPai	Habilitações literárias do pai
SituacaoSEPAi	Situação socioeconómica do pai
ParentescoED	Grau de parentesco do encarregado de educação
TipoCartao	Tipo de cartão utilizado na escola pelo aluno
LocalidadeMae	Local de residência do mãe
Localidade	Local de residência do aluno
NºElementosAgrFami	Número de elementos do agregado familiar
Deslocado	Indicação se o aluno se encontra deslocado da sua região geográfica
Curso	Curso que o aluno frequenta
Abandono	Indicação se o aluno abandonou a escola
AbandonoData	Data em que o aluno abandonou a escola
NIF	Número de identificação fiscal do aluno
PercursoCurricularAlternativo	Indicação se o aluno teve um percurso curricular alternativo

A qualidade que se esperava não se verificou e, num total de 113 colunas, quase todas apresentam mais de 25% dos dados em falta, o que se apresenta como um valor bastante elevado para os 4434 registos que o conjunto de dados contém. A par deste problema, foram ainda identificados muitos outros como: maioria dos códigos postal e moradas inválidas, não

uniformização dos dados inseridos no sistema (Figura 21), números de BI inválidos (Figura 21), NIF inválidos, atributo que seria utilizado como identificador único do aluno, bem como atributos já obsoletos e vários registos duplicados (Figura 22).

Nºbi	Count
0	54
2	1
6	1
13	1
13/2007	1
17/1997	1
17de2003	1
22/2002	1
27/2000	1
29/96	1
29/1998	1
30/2006	1
36	1
39/2000	1
43/2005	1
44/99	1
48/2001	1
48de2005	1
49	1
52/96	1
52/2003	1
55/99	1
55/2000	1
56/1998	1
58/99	1
59/2002	1
59de2001	1
60/2007	1
62/98	1
66/2008	1

Figura 21 - Registos do atributo N_BI

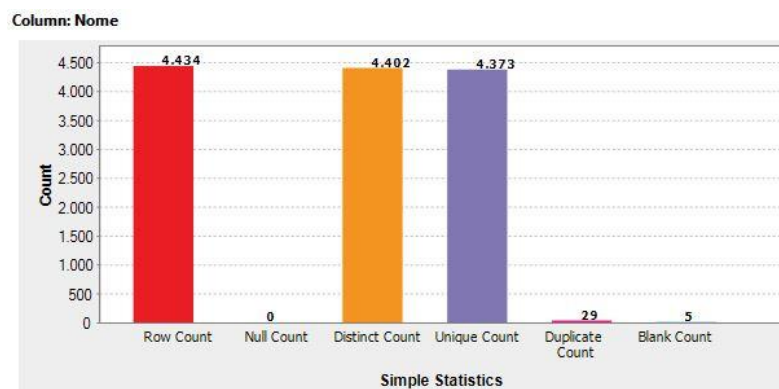


Figura 22 - Qualidade do atributo Nome

Para além de todos estes problemas, neste conjunto de dados era impossível associar um determinado aluno a uma escola pois essa informação não estava presente nos ficheiros de suporte.

Como exemplo, é apresentada na Figura 23 a coluna TextoProfissaoPai do conjunto de dados em questão, e onde fica evidenciado que, apesar de estarem preenchidos, os registos mais frequentes são “Profissão Desconhecida” ou “Sem profissão”, o que inviabiliza estudos futuros que possam necessitar desta informação.

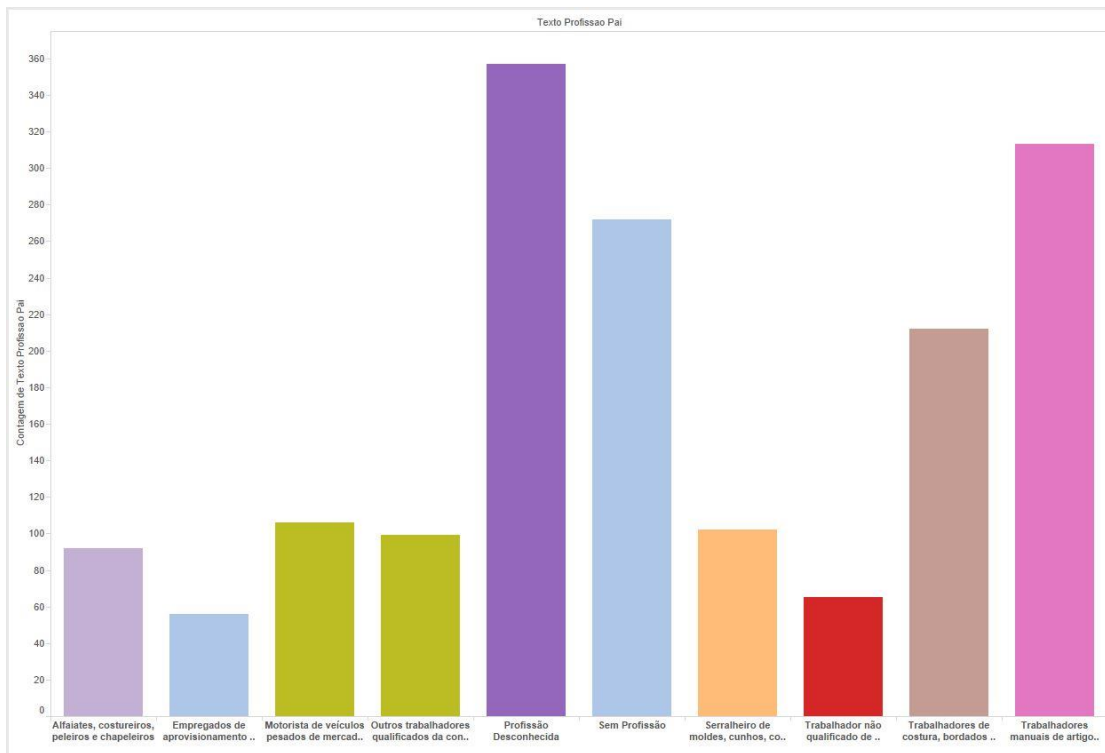


Figura 23 - Exemplo do Relatório de qualidade de dados - coluna TextoProfissaoPai

Para além do problema referido anteriormente, evidenciou-se também que não existe para este atributo, bem como tantos outros, uniformização dos dados inseridos, tendo sido detetado para a mesma profissão diversas designações (ex. Operário Fabril, Operário de Fábrica, operário).

Como é possível verificar na Figura 24, existem registos duplicados que se referem ao mesmo aluno. Nestes casos, um dos registos duplicados apresenta normalmente menos informação, sendo também visível que a informação não é sempre concordante entre registos.

1	NºProcess	Nome	PaisOrigem	Sexo	DataNascimr	LinguaEstr	Morada1	Morada2	0000	000	CodigoPostal	Telefone	NºBI	ValidadeE	NomePai	NomeMae
2937	13987		PT	Masculinc	11/12/2000	Inglês		Conde			conde GMR			8/31/2015 J		
3158	14208		PT	Masculinc	11/12/2000											
3399	15080		PT	Masculinc	7/15/1978	Inglês		Vila das A			Santo Tirso					
3795	15607		PT	Masculinc	7/15/1978	Inglês		Vila das A			Vila das Aves					
4421	16256		PT	Feminino	02/01/2007			Roriz			Roriz		mar/07			
4429	16264		PT	Feminino	02/01/2007						Roriz		0			

Figura 24 – Amostra dos Registos Duplicados

Por exemplo, nas duas primeiras linhas da Figura 24, para além de se verificar que o número de processo é diferente, dizem respeito ao mesmo aluno, em que na primeira linha o registo encontra-se com as colunas todas preenchidas contrastando com a segunda que só apresenta as cinco primeiras colunas preenchidas.

Relativamente aos conjuntos de dados com informação das disciplinas e dos cursos, verificam-se menos debilidades que o anterior, sendo possível a sua utilização no decorrer do projeto. Quanto aos conjuntos com os dados das escolas e da situação dos alunos, foram também identificadas uma série de problemas que também inviabilizam a sua utilização em estudos futuros. Todas estas anomalias estão devidamente identificadas no relatório de qualidade de dados que pode ser acedido através do seguinte link ([Relatório de qualidade](#)).

3.3.3. Fonte de Dados Plataforma Edubox

A *Edubox*S.A, *spin-off* da Universidade de Aveiro, é uma empresa de base tecnológica fundada em 2010 a partir de um projeto de investigação com objetivo de desenvolver software educativo. De entre várias soluções globais de educação desenvolvidas, a *Edubox* presta serviços à Câmara Municipal de Guimarães com a sua plataforma SIGA (Sistema Integrado de Gestão e Aprendizagem) para a gestão de alunos do pré-escolar e do 1º ciclo de escolaridade.

Os dados presentes nesta plataforma foram recolhidos através da invocação de um *Web Service*² disponibilizado pela *Edubox*. Os dados foram obtidos da plataforma SIGA referentes ao sistema educativo do concelho de Guimarães. Na Tabela 11 é possível ver a estrutura da fonte de dados e alguns dos atributos que foram obtidas através do *web service*. A versão completa com todos os atributos pode ser visualizada em anexo (Anexo A2. Fonte de dados Edubox)

² *Web Service* é uma forma de criar uma linha de comunicação entre aplicações para que estas possam comunicar entre si. É uma forma padronizada de integrar aplicações *web-based* usando protocolos web como XML, SOAP, WSDL e UDDI.

Tabela 11 - Atributos da fonte de dados da plataforma SIGA da Edubox

Atributo	Descrição
BIAluno	Número do documento de identificação do aluno
BIEncEdu	Número do documento de identificação do encarregado de educação
BIMae	Número do documento de identificação da mãe
BIPai	Número do documento de identificação do pai
CodPostalAluno	Código postal (4 dígitos) do aluno
DataNascimento	Data de nascimento do aluno
EmailAluno	Endereço de email do aluno
EscalaoAluno	Escalão de apoio social escolar do aluno
ExtensaoCodPostalAluno	Código postal (3 dígitos) do aluno
GrauParentescoEncEdu	Grau de parentesco do encarregado de educação em relação ao aluno
LocalidadeAluno	Localidade do aluno
NifAluno	Número de identificação fiscal do aluno
NomeAluno	Nome do aluno
NomeAnoEscolaridade	Ano de escolaridade que o aluno frequenta
NomeEncEdu	Nome do encarregado de educação
NomeEscola	Nome da escola que o aluno frequenta
NomeTurma	Indicação da turma que aluno frequenta
TelefoneAluno	Número de telefone do aluno
moradaAluno	Morada do aluno
nomeAgrupamento	Nome do agrupamento de escolas que o aluno frequenta
sexo	Indicação do sexo do aluno
InitialDate	Data de entrada do aluno numa turma
EndDate	Data de saída do aluno da turma
IdUser	Identificador único de um determinado aluno

Os dados presentes nesta fonte de dados apresentam, de uma forma geral, boa qualidade. Os dados que são necessários para realização do estudo, nomeadamente a morada do aluno, o código postal, o escalão, sexo, data de nascimento, escola, agrupamento, turma, ano de escolaridade e localidade apresentam-se praticamente preenchidos na sua totalidade. Todavia, foram detetados alguns registos com algumas incoerências ao nível dos códigos postais após cruzamento dos mesmos com os dados dos CTT (Correios e Telecomunicações de Portugal). Também foram detetados alguns números de identificação fiscal e datas de nascimento inválidos.

É importante referir que foram seguidos os mesmos procedimentos de análise da fonte de dados apresentada anteriormente recorrendo à plataforma de análise *Talend Open Studio for Data Quality* e o *Tableau*.

Na Figura 25 é possível verificar uma amostra dos dados extraídos da plataforma SIGA da Edubox num total de 14968 registos.

BtNum	BtEncEdu	BtMe	BtPa	CodPostalNum	CodPostalEncEdu	CodPostalMe	CodPostalPa	DataNascimento	EnaNum	EnaEncEdu	EnaMe	EnaPa	EscalaNum	ExtensaoCodPostalNum	ExtensaoCodPostalEncEdu	ExtensaoCodigoPostalMe	ExtensaoCodigoPostalPa	GrauFaltasEncEdu
1	4810			4810	4810	4810	4810						escala B	041	041	041	041	Outro
2	4810			4810	4810	4810	4810						escala B	220	220	220	220	Outro
3	4810			4810	4810	4810	4810						escala B	006	006	006	006	Fa
4	4780			4780	4780	4780	4780						sem escala	106	106	106	106	Mse
5	4610			4610	4610	4610	4610						sem escala	270	270	270	270	Outro
6	4810			4810	4810	4810	4810						sem escala	026	026	026	026	Outro
7	4810			4810	4810	4810	4810						sem escala	009	009	009	009	Outro
8	4810			4810	4810	4810	4810						sem escala	009	009	009	009	Outro
9	4810			4810	4810	4810	4810						escala A	220	220	220	220	Mse
10	4810			4810	4810	4810	4810						sem escala	233	233	233	233	Mse
11	4810			4810	4810	4810	4810						sem escala	009	009	009	009	Mse
12	4810			4810	4810	4810	4810						escala A	222	222	222	222	Mse
13	4810			4810	4810	4810	4810						sem escala	246	246	246	246	Outro
14	4800			4800	4800	4800	4800						escala B	018	018	018	018	Outro
15	4805			4805	4805	4805	4805						sem escala	128	128	128	128	Outro
16	4810			4810	4810	4810	4810						sem escala	008	008	008	008	Mse
17	4810			4810	4810	4810	4810						sem escala	220	220	220	220	Mse
18	4810			4810	4810	4810	4810						escala B	006	006	006	006	Outro
19	4810			4810	4810	4810	4810						sem escala	008	008	008	008	Mse
20	4810			4810	4810	4810	4810						escala B	009	009	009	009	Mse
21	4810			4810	4810	4810	4810						escala A	562	562	562	562	Outro
22	4810			4810	4810	4810	4810						escala A	505	505	505	505	Fa
23	4810			4810	4810	4810	4810						escala A	225	225	225	225	Fa
24	4810			4810	4810	4810	4810						sem escala	133	133	133	133	Mse
25	4800			4800	4800	4800	4800						sem escala	018	018	018	018	Outro
26	4810			4810	4810	4810	4810						escala B	233	233	233	233	Outro
27	4810			4810	4810	4810	4810						sem escala	517	517	517	517	Mse
28	4810			4810	4810	4810	4810						escala A	009	009	009	009	Mse
29	4810			4810	4810	4810	4810						sem escala	225	225	225	225	Outro
30	4810			4810	4810	4810	4810						sem escala	009	009	009	009	Outro
31	4810			4810	4810	4810	4810						sem escala	008	008	008	008	Outro
32	4810			4810	4810	4810	4810						sem escala	246	246	246	246	Outro
33	4810			4810	4810	4810	4810						sem escala	006	006	006	006	Outro
34	4810			4810	4810	4810	4810						escala A	218	218	218	218	Outro
35	4800			4800	4800	4800	4800						sem escala	240	240	240	240	Mse
36	4800			4800	4800	4800	4800						sem escala	441	441	441	441	Mse
37	4810			4810	4810	4810	4810						sem escala	006	006	006	006	Mse
38	4810			4810	4810	4810	4810						sem escala	275	275	275	275	Outro
39	4810			4810	4810	4810	4810						sem escala	008	008	008	008	Mse
40	4810			4810	4810	4810	4810						sem escala	006	006	006	006	Fa

Figura 25 - Amostra dos dados extraídos da plataforma SIGA

Para além dos problemas detetados, no atributo sexo seria expectável somente dois tipos de valores que definiam o género feminino e masculino. Contudo foi verificado um terceiro valor que é o “não definido”, como é possível verificar na Figura 26. Optou-se por manter estes valores pela dificuldade de automatizar o processo e definir os valores reais para cada registo.

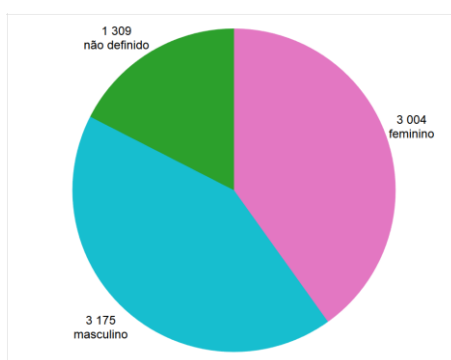


Figura 26 - Análise aos valores do atributo Sexo

Todas estas anomalias estão devidamente identificadas no relatório de qualidade de dados, desenvolvido no âmbito de outra dissertação de mestrado associada à caracterização geo-espaial da população estudiantil de Guimarães e que pode ser acedido através do seguinte [link \(Relatório qualidade\)](#).

4. Implementação do Sistema de *Business Intelligence*

Neste capítulo são apresentadas todas as fases do desenvolvimento e implementação do sistema que foi proposto para a concretização deste trabalho. O capítulo inicia-se com a descrição da conceção do *Data Warehouse*, seguida da descrição individual de cada uma das tabelas de factos e de dimensão presentes no modelo de dados apresentado. Segue-se a descrição pormenorizada de todo o processo ETL realizado para a extração, transformação e carregamento dos dados no modelo. Para finalizar este capítulo, é feita a descrição de todo o processo analítico desde a criação do modelo de cubos OLAP bem como a criação e análises feitas através de *reports e dashboards*. Neste âmbito são também apresentadas as aplicações *front-end* que possibilitam ao utilizador proceder à análise exploratória de toda a informação.

4.1. Sistema de *Data Warehousing*

Nesta secção é apresentado e descrito o modelo de dados do *Data Warehouse* bem como o processo ETL que foi desenvolvido após a análise e elaboração do relatório da qualidade dos conjuntos de dados.

4.1.1. Modelo de Dados do *Data Warehouse*

A modelação multidimensional é uma técnica utilizada para projetar bases de dados, que foi desenvolvida para a construção de *Data Warehouses* colmatando as limitações dos métodos tradicionais. Nesta fase fulcral para o sucesso da solução, idealizou-se um modelo que fosse capaz de cobrir todos os aspetos e análises efetuadas na carta educativa mesmo não tendo acesso a determinadas fontes de dados para a sua realização. Nesse sentido, é importante referir que a modelação concebida foi definida de acordo e com base nas necessidades identificadas para a realização de uma carta educativa completa, que seja capaz de conduzir a uma fundamentada tomada de decisão no processo de análise da rede escolar do município de Guimarães.

O modelo multidimensional apresentado na Figura 27 está assente num esquema em Constelação que tem como particularidade a possibilidade de interligar as várias estrelas por uma ou mais dimensões. Mais à frente, são apresentadas cada uma das estrelas de forma individual para que seja possível a sua leitura e análise.

Na Figura 27 é possível visualizar todas as estrelas que constituem a constelação, as respetivas tabelas de dimensão e de factos, e como se relacionam entre si. Este modelo é composto por seis tabelas de factos que se encontram unidas por vinte e uma dimensões e pretende-se, de seguida, detalhar todas as estrelas que o compõem.

As tabelas de factos presentes no modelo apresentado na Figura 27 são representadas pelas seguintes designações:

- **FactSistemaEducativo;**
- **FactEnqGeograficoDemografico;**
- **FactAnaliseNotas;**
- **FactMortalidadeNatalidade;**
- **FactAlunoEscola;**
- **FactCaractClimatica;**

A tabela de factos **FactMortalidadeNatalidade**, representada na Figura 28, possibilita o armazenamento de informação relativa ao número de nados vivos e mortos ocorridos por localização e por ano, de acordo com o relacionamento existente com as tabelas de dimensão (**DimLocalizacao** e **DimTempoAno**) que lhe conferem estas perspetiva.

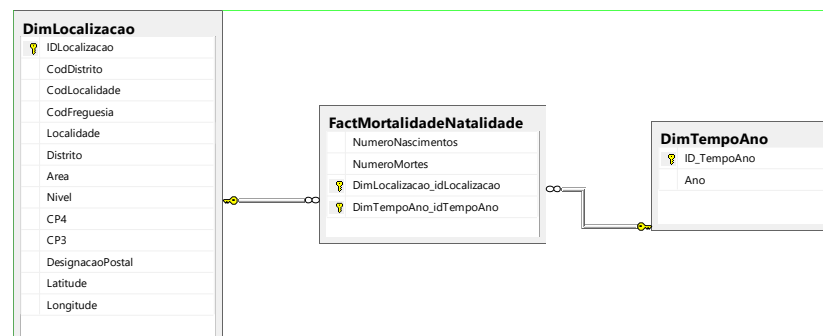


Figura 28 - Modelo de Dados da tabela FactNatalidadeMortalidade

Na Tabela 12 é possível visualizar com mais detalhe todos os atributos que integram a tabela de factos apresentada na Figura 28.

Tabela 12 - Tabela de factos FactNatalidadeMortalidade

FactMortalidadeNatalidade			
Atributo	Tipo	Chave	Descrição
NumeroNascimentos	INT	-	Número de Nados Vivos
NumeroMortes	INT	-	Número de Nados Mortos
DimLocalizacao_idLocalizacao	INT	PK e FK	ID da DimLocalizacao
DimTempoAno_idTempoAno	INT	PK e FK	ID da DimTempo

Nesta tabela de factos destacam-se as chaves estrangeiras (FK) das dimensões Localização e TempoAno. A chave primária (PK) é composta pelas duas chaves estrangeiras e permite identificar uma situação em específico. É importante referir que, para qualquer uma das tabelas de factos, na implementação do sistema OLAP podem ser calculadas várias medidas. A tabela apresentada anteriormente contém dois factos (**NumeroNascimentos** e **NumeroMortes**) que foram carregados através de um ficheiro Excel, com os dados relativos ao número de nados-vivos e mortos, ao longo dos últimos anos, no concelho de Guimarães. Estes dados foram obtidos através do *website* do instituto nacional de estatística.

A tabela de factos **FactAlunoEscola**, representada na Figura 29, permite armazenar a informação referente à ligação de um determinado aluno à sua respetiva escola. É possível analisar esta informação a partir de várias perspetivas, pois esta tabela de factos está relacionada com várias tabelas de dimensão (**DimSituacao**, **DimAnoEscolar**, **DimLocalizacao**, **DimEscola**, **DimAluno**, **DimEscalao**, **DimTempo**, **DimTurma**, **DimAgrupamento**).

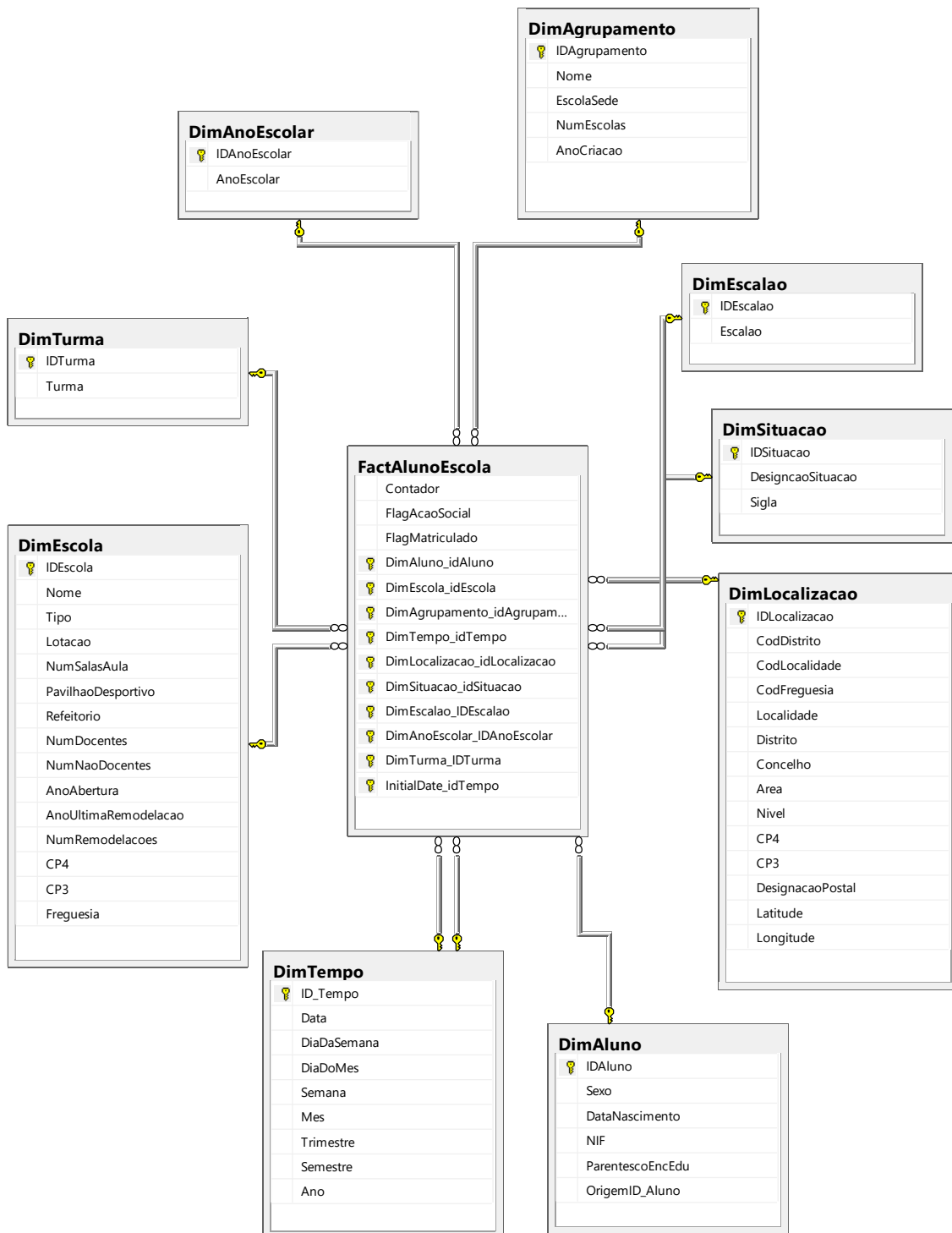


Figura 29 - Modelo de dados da Tabela FactAlunoEscola

Na Tabela 13 é possível visualizar com mais detalhe todos os atributos que integram a tabela de factos apresentada na Figura 29.

Tabela 13 - Tabela de Factos FactAlunoEscola

FactAlunoEscola			
Atributo	Tipo	Chave	Descrição
Contador	INT	-	Contador de registos
FlagAcaoSocial	INT	-	Indica se o aluno esta sujeito a apoio social
FlagMatriculado	INT	-	Indica se o aluno está matriculado
DimAluno_idAluno	INT	PK e FK	ID da DimAluno
DimEscola_idEscola	INT	PK e FK	ID da DimEscola
DimAgrupamento_idAgrupamento	INT	PK e FK	ID da DimAgrupamento
DimTempo_idTempo	INT	PK e FK	ID da DimTempo
DimLocalizacao_idLocalizacao	INT	PK e FK	ID da DimLocalizacao
DimSituacao_idSituacao	INT	PK e FK	ID da DimSituacao
DimEscalao_IDEscalao	INT	PK e FK	ID da DimEscalao
DimAnoEscolar_IDAnoEscolar	INT	PK e FK	ID da DimAnoEscolar
DimTurma_IDTurma	INT	PK e FK	ID da DimTurma
InitialDate_idTempo	INT	PK e FK	ID da DimTempo

A tabela de factos FactAlunoEscola contém todos os registos que associam um aluno a uma determinada escola. É possível verificar todas as chaves estrangeiras que relacionam esta tabela com as respetivas dimensões, de notar também que existem duas ligações à tabela de dimensão **DimTempo**, isto pela necessidade de, para o mesmo registo, relacionar a data de nascimento do alunos bem como data de entrada numa determinada turma. É importante referir que esta é uma versão “reduzida” desta tabela de factos, pois a falta de dados não permitiu uma implementação mais abrangente. Apesar de se ter desenvolvido um modelo o mais próximo possível da realidade da carta educativa, nesta situação teve de se optar por reduzir o número de factos para que esta tabela fosse possível de ser implementada. Numa revisão futura do modelo, esta tabela poderá evoluir para uma versão mais completa contemplando outros atributos que possam ser úteis para outras análises mais extensivas.

A tabela de factos **FactEnqGeograficoDemografico**, representada na Figura 30, possibilita o armazenamento de informação relativa ao enquadramento geográfico e demográfico da população. Esta informação é proveniente dos Censos³ que podem ser analisados por

³ Os Censos são considerados como um processo normalizado de recolha, tratamento, avaliação, análise e difusão de dados referenciados num determinado momento temporal e que dizem respeito a todas as unidades estatísticas (indivíduos, famílias, alojamentos e edifícios) de uma determinada zona geográfica (INE, 2013).

localização e por ano, de acordo com os relacionamentos existentes com as tabelas de dimensão (**DimLocalizacao** e **DimTempoAno**) que lhe conferem estas perspetivas.

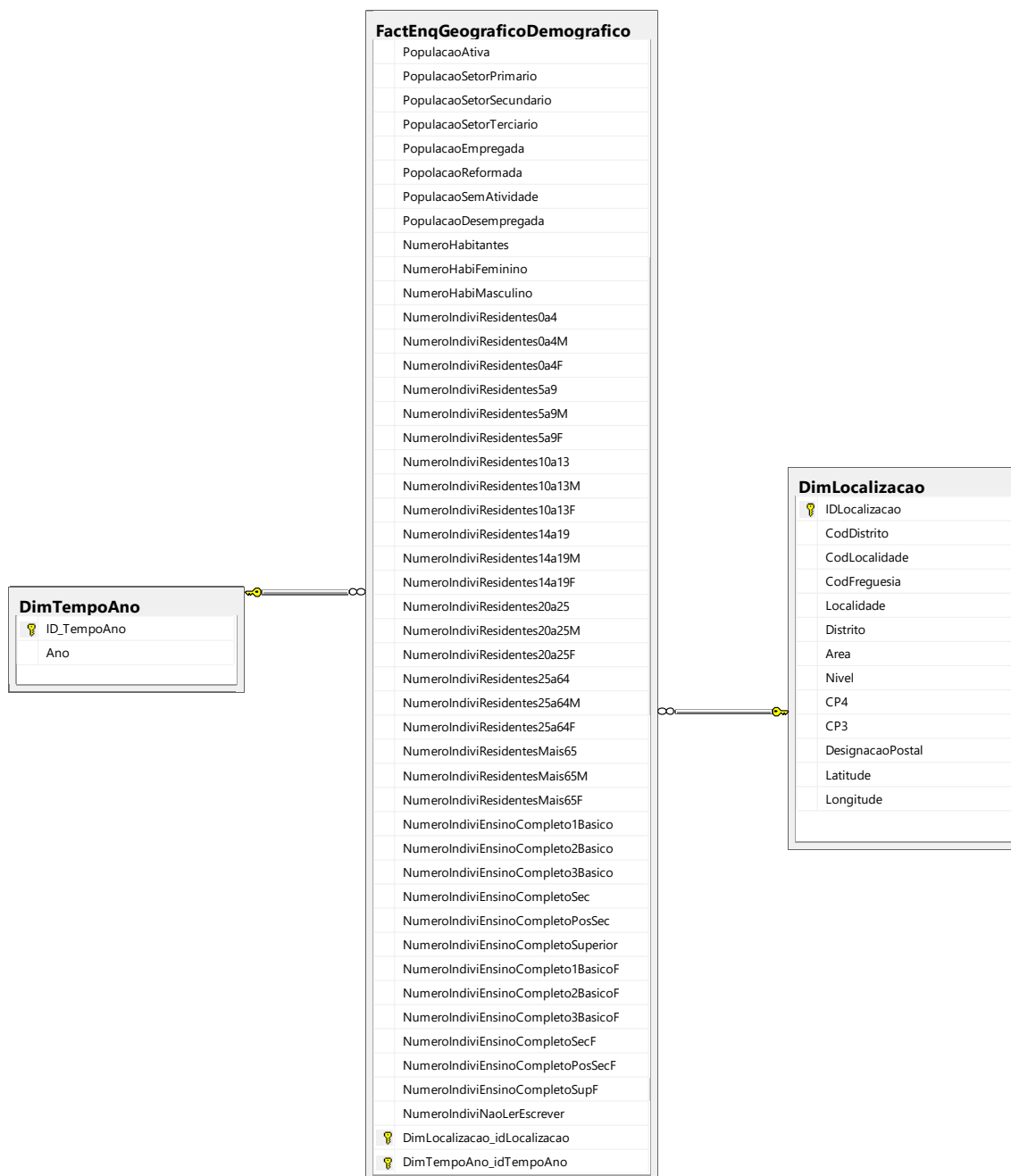


Figura 30 - Modelo de dados da Tabela FactEnqGeograficoDemografico

Nas Tabela 14, Tabela 15 e Tabela 16 é possível visualizar com mais detalhe todos os atributos que integram a tabela de factos apresentada na Figura 30.

Tabela 14 - Tabela de factos FactEnqGeograficoDemografico (a)

FactEnqGeograficoDemografico			
Atributo	Tipo	Chave	Descrição
PopulacaoAtiva	INT	-	População com atividade económica
PopulacaoSetorPrimario	INT	-	População a laborar no setor primário
PopulacaoSetorSecundario	INT	-	População a laborar no setor secundário
PopulacaoSetorTerciario	INT	-	População a laborar no setor terciário
PopulacaoEmpregada	INT	-	Total de indivíduos empregados
PopulacaoReformada	INT	-	Total de indivíduos reformados
PopulacaoSemAtividade	INT	-	Total de indivíduos sem atividade económica
PopulacaoDesempregada	INT	-	Total de indivíduos desempregados
NumeroHabitantes	INT	-	Número de indivíduos residentes
NumeroHabiFeminino	INT	-	Número de indivíduos residentes do sexo feminino
NumeroHabiMasculino	INT	-	Número de indivíduos residentes do sexo masculino
NumeroIndiviResidentes0a4	INT	-	Número de indivíduos residentes entre os 0 e os 4 anos de idade
NumeroIndiviResidentes0a4M	INT	-	Número de indivíduos residentes entre os 0 e os 4 anos de idade do sexo masculino
NumeroIndiviResidentes0a4F	INT	-	Número de indivíduos residentes entre os 0 e os 4 anos de idade do sexo feminino
NumeroIndiviResidentes5a9	INT	-	Número de indivíduos residentes entre os 5 e os 9 anos de idade
NumeroIndiviResidentes5a9M	INT	-	Número de indivíduos residentes entre os 5 e os 9 anos de idade do sexo masculino
NumeroIndiviResidentes5a9F	INT	-	Número de indivíduos residentes entre os 5 e os 9 anos de idade do sexo feminino
NumeroIndiviResidentes10a13	INT	-	Número de indivíduos residentes entre os 10 e os 13 anos de idade
NumeroIndiviResidentes10a13M	INT	-	Número de indivíduos residentes entre os 10 e os 13 anos de idade do sexo masculino
NumeroIndiviResidentes10a13F	INT	-	Número de indivíduos residentes entre os 10 e os 13 anos de idade do sexo feminino
NumeroIndiviResidentes14a19	INT	-	Número de indivíduos residentes entre os 14 e os 19 anos de idade
NumeroIndiviResidentes14a19M	INT	-	Número de indivíduos residentes entre os 14 e os 19 anos de idade do sexo masculino

Tabela 15 - Tabela de factos FactEnqGeograficoDemografico (b)

FactEnqGeograficoDemografico			
Atributo	Tipo	Chave	Descrição
NumeroIndiviResidentes14a19F	INT	-	Número de indivíduos residentes entre os 14 e os 19 anos de idade do sexo feminino
NumeroIndiviResidentes20a25	INT	-	Número de indivíduos residentes entre os 20 e os 25 anos de idade
NumeroIndiviResidentes20a25M	INT	-	Número de indivíduos residentes entre os 20 e os 25 anos de idade do sexo masculino
NumeroIndiviResidentes20a25F	INT	-	Número de indivíduos residentes entre os 20 e os 25 anos de idade do sexo feminino
NumeroIndiviResidentes25a64	INT	-	Número de indivíduos residentes entre os 25 e os 64 anos de idade
NumeroIndiviResidentes25a64M	INT	-	Número de indivíduos residentes entre os 25 e os 64 anos de idade do sexo masculino
NumeroIndiviResidentes25a64F	INT	-	Número de indivíduos residentes entre os 25 e os 64 anos de idade do sexo feminino
NumeroIndiviResidentesMais65	INT	-	Número de indivíduos residentes com mais de 65 anos de idade
NumeroIndiviResidentesMais65M	INT	-	Número de indivíduos residentes com mais de 65 anos de idade do sexo masculino
NumeroIndiviResidentesMais65F	INT	-	Número de indivíduos residentes com mais de 65 anos de idade do sexo feminino
NumeroIndiviEnsinoCompleto1Basico	INT	-	Número de indivíduos residentes com o 1º ciclo de escolaridade completo
NumeroIndiviEnsinoCompleto2Basico	INT	-	Número de indivíduos residentes com o 2º ciclo de escolaridade completo
NumeroIndiviEnsinoCompleto3Basico	INT	-	Número de indivíduos residentes com o 3º ciclo de escolaridade completo
NumeroIndiviEnsinoCompletoSec	INT	-	Número de indivíduos residentes com ensino secundário completo
NumeroIndiviEnsinoCompletoPosSec	INT	-	Número de indivíduos residentes com ensino pós secundário completo
NumeroIndiviEnsinoCompletoSuperior	INT	-	Número de indivíduos residentes com ensino superior completo

Tabela 16 - Tabela de factos **FactEnqGeograficoDemografico (c)**

FactEnqGeograficoDemografico			
Atributo	Tipo	Chave	Descrição
NumeroIndiviEnsinoCompletoPosSecF	INT	-	Número de indivíduos residentes a frequentar ensino pós secundário
NumeroIndiviEnsinoCompletoSupF	INT	-	Número de indivíduos residentes a frequentar ensino superior
NumeroIndiviNaoLerEscrever	INT	-	Número de indivíduos que não sabe ler nem escrever
NumeroIndiviEnsinoCompleto1BasicoF	INT	-	Número de indivíduos residentes a frequentar o 1º ciclo de escolaridade
NumeroIndiviEnsinoCompleto2BasicoF	INT	-	Número de indivíduos residentes a frequentar o 2º ciclo de escolaridade
NumeroIndiviEnsinoCompleto3BasicoF	INT	-	Número de indivíduos residentes a frequentar o 3º ciclo de escolaridade
NumeroIndiviEnsinoCompletoSecF	INT	-	Número de indivíduos residentes a frequentar ensino secundário
DimLocalizacao_idLocalizacao	INT	PK e FK	ID da DimLocalizacao
DimTempoAno_idTempoAno	INT	PK e FK	ID da DimTempo

A tabela de factos **FactEnqGeográficoDemográfico** contém todos os registos relativos aos censos do concelho de Guimarães. É possível verificar as duas chaves estrangeiras que relacionam esta tabela com as respetivas dimensões DimTempoAno e DimLocalização, que compõem a chave primária e permitem identificar uma determinada situação demográfica e geográfica. É importante referir que esta tabela é composta por 45 factos tendo sido carregados através de um ficheiro Excel que contém os dados relativos aos censos de 2001 e 2011.

A tabela de factos **FactSistemaEducativo**, representada na Figura 31, permite armazenar a informação referente ao sistema educativo das escolas do concelho de Guimarães. É possível analisar esta informação a partir de várias perspetivas, pois esta tabela de factos está relacionada com várias tabelas de dimensão (**DimAgrupamento, DimBridgeDisciplina, DimLocalizacao, DimEscola, DimBridgeCurso, DimBridgeProf, DimTempo, DimBridgeNaoDocente**).

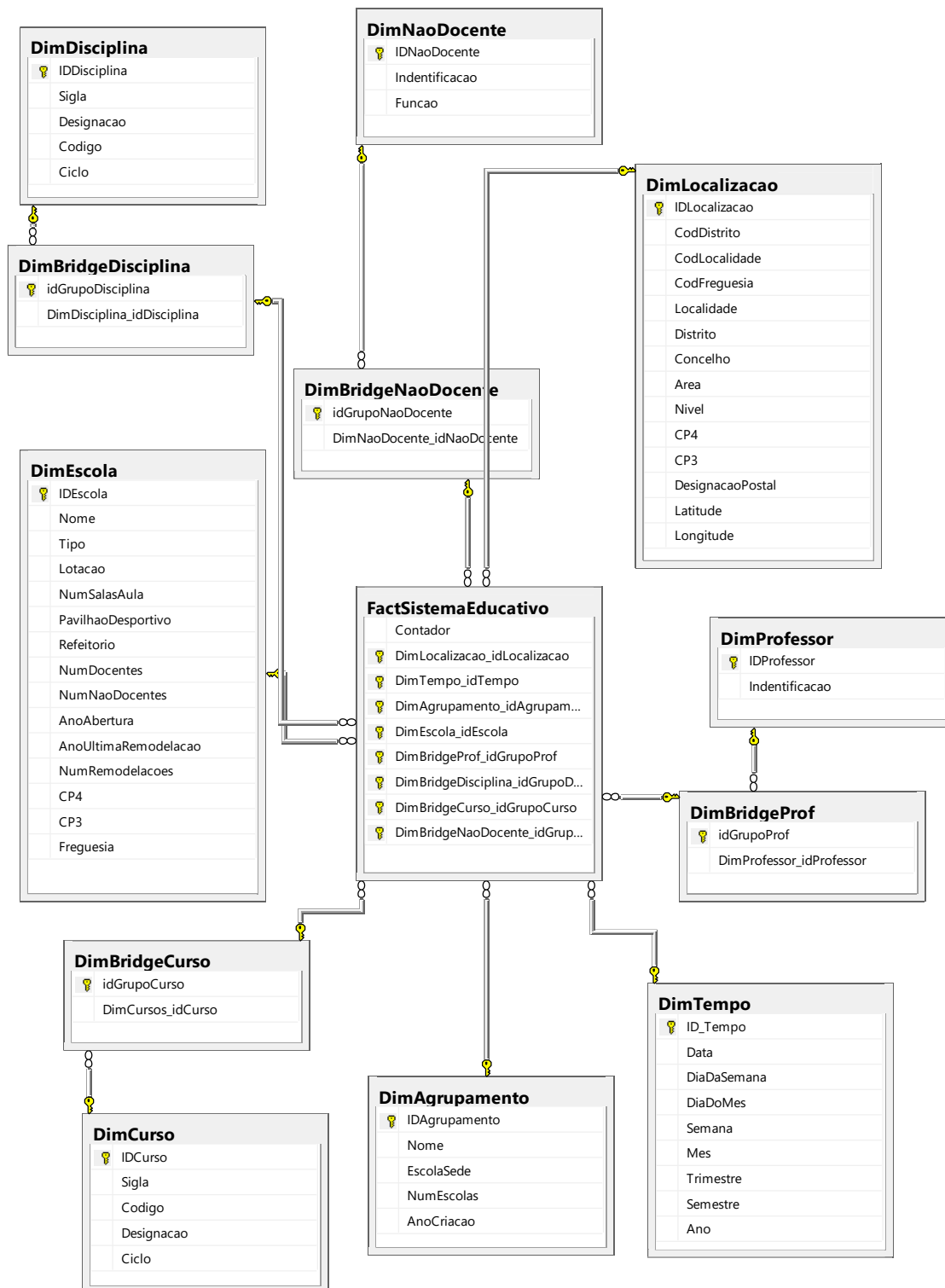


Figura 31 - Modelo de dados da tabela FactSistemaEducativo

Na Tabela 17 é possível visualizar com mais detalhe todos os atributos que integram a tabela de factos apresentada na Figura 31.

Tabela 17 - Tabela de factos FactSistemaEducativo

FactSistemaEducativo			
Atributo	Tipo	Chave	Descrição
Contador	INT	-	Contador de registos
DimLocalizacao_idLocalizacao	INT	PK e FK	ID da DimLocalizacao
DimTempo_idTempo	INT	PK e FK	ID da DimTempo
DimAgrupamento_idAgrupamento	INT	PK e FK	ID da DimAgrupamento
DimEscola_idEscola	INT	PK e FK	ID da DimEscola
DimBridgeProf_idGrupoProf	INT	PK e FK	ID da DimBrigeProf
DimBridgeDisciplina_idGrupoDisciplina	INT	PK e FK	ID da DimBridgeDisciplina
DimBridgeCurso_idGrupoCurso	INT	PK e FK	ID da DimBridgeCurso
DimBridgeNaoDocente_idGrupoNaoDocente	INT	PK e FK	ID da DimBridgeNaoDocente

Esta tabela de factos permite armazenar os registos referentes às escola, permitindo associá-la a uma série de outros atributos que fazem parte do ambiente escolar. De uma forma global, cada escola fica associada aos seus docentes, não docentes, disciplinas e cursos.

A tabela de factos **FactCaractClimatica**, representada na Figura 32, possibilita o armazenamento de informação relativa à caracterização do clima por localização e por dimensão temporal, de acordo com o relacionamento existente com as tabelas de dimensão (**DimLocalizacao e DimTempo**) que lhe conferem estas perspetivas.

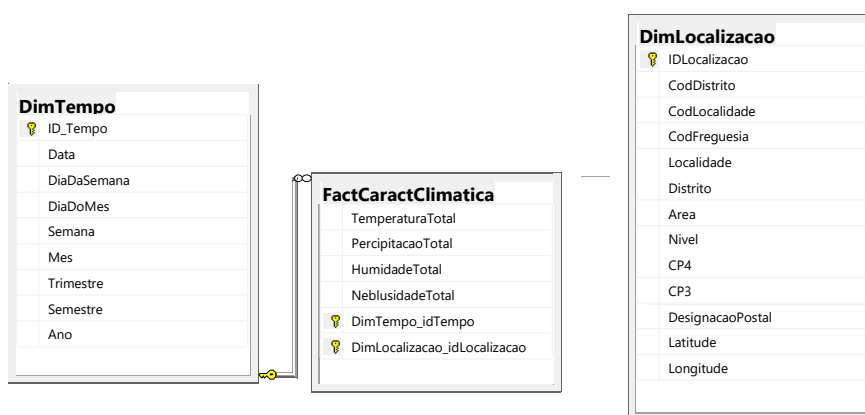


Figura 32 - Modelo de dados da tabela FactCaractClimatica

Na Tabela 18 é possível visualizar com mais detalhe todos os atributos que integram a tabela de factos apresentada na Figura 32.

Tabela 18 - Tabela de factos FactCaractClimatica

FactCaractClimatica			
Atributo	Tipo	Chave	Descrição
TemperaturaTotal	INT	-	Valor total da temperatura climatérica
PercipitacaoTotal	INT	-	Valor total da precipitação atmosférica
HumidadeTotal	INT	-	Valor total da humidade
NebulosidadeTotal	INT	-	Valor da nebulosidade
DimTempo_idTempo	INT	PK e FK	ID da DimTempo
DimLocalizacao_idLocalizacao	INT	PK e FK	ID da DimLocalizacao

Nesta tabela de factos destacam-se as chaves estrangeiras (FK) das dimensões Localização e Tempo, bem como a chave primária (PK) composta pelas duas chaves estrangeiras que permitem identificar em conjunto uma situação climatérica em específico.

A tabela de factos **FactAnaliseNotas**, representada na Figura 33, permite registar a informação relativa ao aproveitamento e às classificações obtidas por um determinado aluno. É possível analisar esta informação a partir de várias perspetivas, pois esta tabela de factos está relacionada com várias tabelas de dimensão (**DimNota, DimAluno, DimLocalizacao, DimDisciplina, DimNivelEnsino, DimPeriodo, DimTempoAno**).

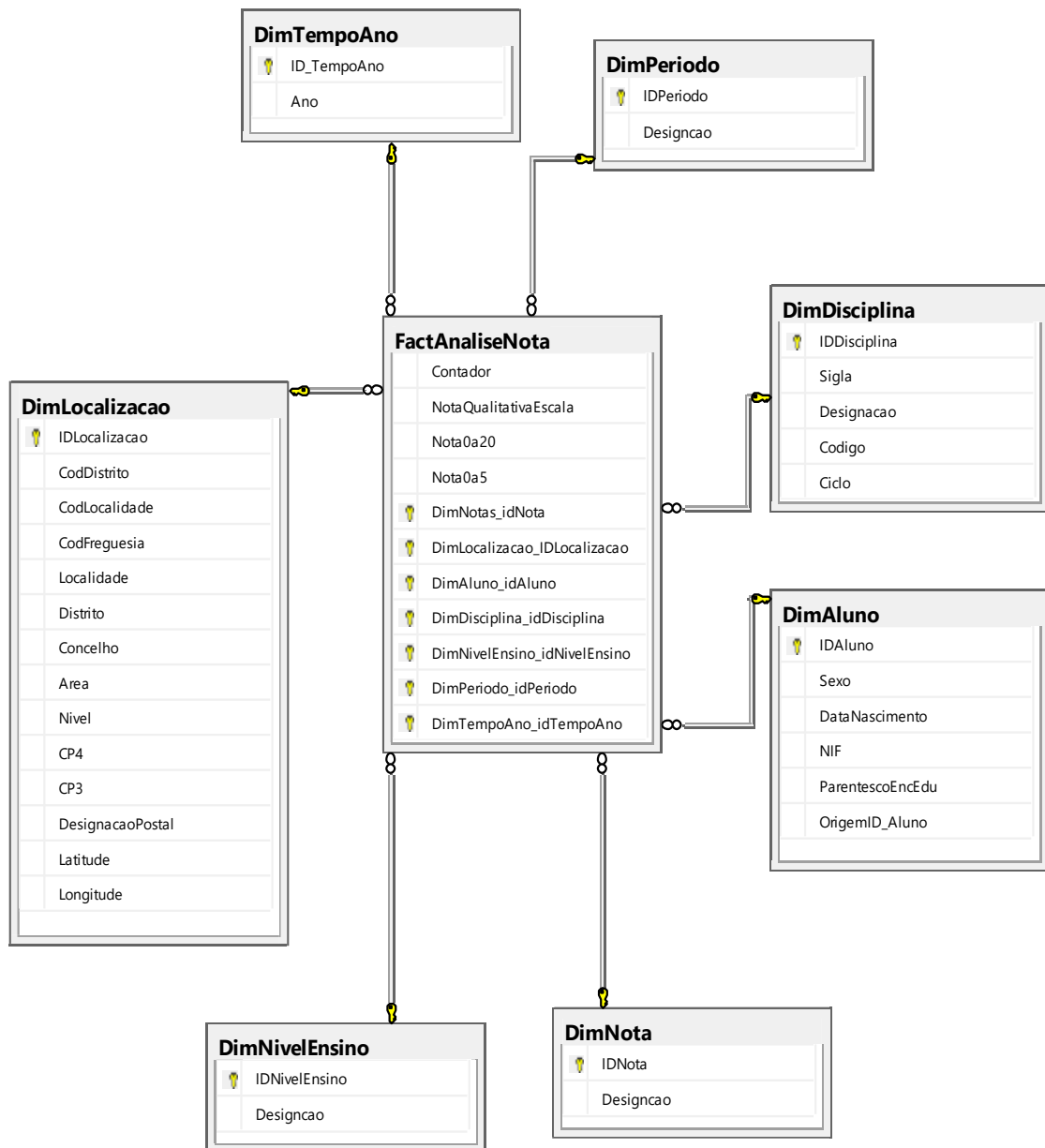


Figura 33 - Modelo de dados da tabela FactAnaliseNota

Na Tabela 19 é possível visualizar com mais detalhe todos os atributos que integram a tabela de factos apresentada na Figura 33.

Tabela 19 - Tabela de factos FactAnaliseNotas

FactAnaliseNotas			
Atributo	Tipo	Chave	Descrição
Contador	INT	-	Contador de registos
NotaQualitativaEscala	INT	-	Flag Avaliação qualitativa
Nota0a20	INT	-	Flag avaliação de 0 a 20 valores
Nota0a5	INT	-	Flag avaliação de 0 a 5 valores
DimNotas_idNota	INT	PK e FK	ID da DimNota
DimLocalizacao_idLocalizacao	INT	PK e FK	ID da DimLocalizacao
DimAluno_idAluno	INT	PK e FK	ID da DimAluno
DimDisciplina_idDisciplina	INT	PK e FK	ID da DimDisciplina
DimNivelEnsino_idNivelEnsino	INT	PK e FK	ID da DimNivelEnsino
DimPeriodo_idPeriodo	INT	PK e FK	ID da DimPeriodo
DimTempoAno_idTempoAno	INT	PK e FK	ID da DimTempoAno

Para além destas tabelas de factos, descritas anteriormente, o modelo de dados incorpora ainda 21 tabelas de dimensão. As tabelas de dimensão presentes no modelo apresentado na Figura 27 são representadas pelas seguintes designações:

- **DimTempo**
- **DimTempoAno**
- **DimLocalizacao**
- **DimCurso**
- **DimBridgeCurso**
- **DimDisciplina**
- **DimBridgeDisciplina**
- **DimAgrupamento**
- **DimEscola**
- **DimAluno**
- **DimProfessor**
- **DimBridgeProf**
- **DimNaoDocente**
- **DimBridgeNaoDocente**
- **DimNivelEnsino**
- **DimNotas**
- **DimPeriodo**
- **DimSituacao**
- **DimAnoEscolar**
- **DimEscalao**
- **DimTurma**

A tabela de dimensão **DimTempo** armazena informação relativa ao momento em que determinado facto ocorre. Na Tabela 20 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 20 - Tabela de dimensão DimTempo

DimTempo				
Atributo	Tipo	Chave	SCD	Descrição
ID_Tempo	INT	PK	-	Identificador único de cada registo
Data	Date	-	1	Data (yyyy-mm-dd)
DiaDaSemana	nvarchar(20)	-	1	Designação do Dia da Semana
DiaDoMes	INT	-	1	Dia do Mês
Semana	INT	-	1	Semana do Ano
Mes	INT	-	1	Mês do Ano
Trimestre	INT	-	1	Trimestre
Semestre	INT	-	1	Semestre
Ano	INT	-	1	Ano (yyyy)

A dimensão tempo armazena todas as datas possíveis para o período. Esta informação é gerada a partir de uma *script* que é executada através de um *stored procedure* ⁴ em sql. Esta *script* foi incluída no processo ETL para que sempre que seja executado, sejam carregados os registos até à data atual. Esta script pode ser consultada em anexo (Anexo B1).

A tabela de dimensão **DimTempoAno** armazena informação relativa ao calendário, em anos, em que determinado facto ocorreu. Na Tabela 21 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 21 - Tabela de dimensão DimTempoAno

DimTempoAno				
Atributo	Tipo	Chave	SCD	Descrição
ID_TempoAno	INT	PK	-	Identificador único de cada registo
Ano	INT	-	1	Ano (yyyy)

A dimensão DimTempoAno contém todos os anos desde 1950 até 2218.

A tabela de dimensão **DimLocalizacao** armazena informação relativa à localização em que determinado facto ocorreu. Na Tabela 22 é possível visualizar com mais detalhe todos os atributos desta dimensão.

⁴ *Stored Procedures* caracterizam-se como um conjunto de instruções SQL que podem ser executadas de várias maneiras. Como são executados diretamente no servidor podem reduzir os tempos de resposta e melhorar a performance das bases de dados.

Tabela 22 - Tabela de dimensão DimLocalizacao

DimLocalizacao				
Atributo	Tipo	Chave	SCD	Descrição
IDLocalizacao	INT	PK	-	Identificador único de cada registo
CodDistrito	INT	-	1	Código identificativo do Distrito
CodLocalidade	INT	-	1	Código identificativo da Localidade
CodFreguesia	INT	-	1	Código identificativo da Freguesia
Localidade	nvarchar(50)	-	1	Designação da Localidade
Distrito	nvarchar(50)	-	1	Designação do Distrito
Area	float	-	2	Área Geográfica da Localidade (km ²)
Nivel	nvarchar(20)	-	1	Nível da localidade (Vila, freguesia, etc)
CP4	INT	-	1	Código Postal de 4 dígitos
CP3	INT	-	1	Código Postal de 3 dígitos
DesignacaoPostal	nvarchar(255)	-	1	Designação da localidade referente ao Código de Postal
Latitude	nvarchar(20)	-	1	Coordenada Geográfica Latitude
Longitude	nvarchar(20)	-	1	Coordenada Geográfica Longitude

Esta tabela é uma dimensão espacial constituída unicamente por dados não geométricos, ou seja, dados apenas nominais. O facto de estes não serem dados geométricos não implica que não seja possível retirar informação que permita a georreferenciação (por exemplo, nomes de cidades ou localidades).

A tabela de dimensão **DimCurso** armazena informação relativa aos cursos lecionados no agrupamento de escolas Virgínia Moura do concelho de Guimarães. Apenas foram carregados para esta dimensão os dados fornecidos por este agrupamento, uma vez que os problemas de qualidade de dados, já referidos, levaram a que o protótipo incorporasse um conjunto mais restrito de dados. Na Tabela 23 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 23 - Tabela de dimensão DimCurso

DimCurso				
Atributo	Tipo	Chave	SCD	Descrição
IDCurso	INT	PK	-	Identificador único de cada registo
Sigla	nvarchar(100)	-	1	Abreviatura do curso
Codigo	nvarchar(100)	-	1	Código do curso
Designacao	nvarchar(255)	-	1	Designação do curso
Ciclo	nvarchar(50)	-	1	Ciclo de estudos em que o curso é lecionado

Segue-se a tabela de dimensão **DimBridgeCurso** que está relacionada com a dimensão Curso anteriormente descrita pela chave estrangeira idCurso. Esta dimensão é uma *Bridge Table Dimension*, criada para que seja possível associar a uma linha da tabela de factos, por exemplo uma escola, a um grupo de cursos. Nesta situação optou-se por criar uma entidade de grupo que se relaciona com a dimensão de múltiplo valor. Na Tabela 24 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 24 - Tabela de dimensão DimBridgeCurso

DimBridgeCurso				
Atributo	Tipo	Chave	SCD	Descrição
idGrupoCurso	INT	PK	-	Identificador único de cada grupo
DimCurso_idCurso	INT	FK e PK	1	ID da DimCurso

A tabela de dimensão **DimDisciplina** armazena informação relativa a todas as disciplinas lecionadas nos vários cursos do agrupamento de escola Virgínia Moura do concelho de Guimarães. Na Tabela 25 é possível visualizar detalhadamente todos os atributos desta dimensão.

Tabela 25 - Tabela de dimensão DimDisciplina

DimDisciplina				
Atributo	Tipo	Chave	SCD	Descrição
IDDisciplina	INT	PK	-	Identificador único de cada registo
Sigla	nvarchar(10)	-	1	Abreviatura da disciplina
Designacao	nvarchar(50)	-	1	Designação da disciplina
Codigo	nvarchar(50)	-	1	Código da disciplina
Ciclo	nvarchar(50)	-	1	Ciclo de estudo onde a disciplina é lecionada

Segue-se a tabela de dimensão **DimBridgeDisciplina** que está relacionada com a dimensão Disciplina anteriormente descrita. Esta dimensão é uma *Bridge Table Dimension*, criada para que seja possível associar a uma linha da tabela de factos, por exemplo uma escola, a um grupo de disciplinas. Nesta situação optou-se por criar uma entidade de grupo que se relaciona com a dimensão de múltiplo valor. Na Tabela 26 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 26 - Tabela de dimensão DimBridgeDisciplina

DimBridgeDisciplina				
Atributo	Tipo	Chave	SCD	Descrição
idGrupoDisciplina	INT	PK	-	Identificador único de cada grupo
DimDisciplina_idDisciplina	INT	FK e PK	-	ID da DimDisciplina

A tabela de dimensão **DimAgrupamento** armazena informação relativa aos agrupamentos de escolas do concelho de Guimarães (nome, número de escolas, ano de criação, etc.). Na Tabela 27 é possível visualizar com mais detalhe todos os atributos desta dimensão. Estes dados foram cedidos pela Câmara Municipal de Guimarães num ficheiro Excel.

Tabela 27 - Tabela de dimensão DimAgrupamento

DimAgrupamento				
Atributo	Tipo	Chave	SCD	Descrição
IDAgrupamento	INT	PK	-	Identificador único de cada registo
Nome	nvarchar(255)	-	1	Nome do Agrupamento
EscolaSede	nvarchar(255)	-	2	Nome da Escola Sede do Agrupamento
NumEscolas	INT	-	2	Número de Escolas do Agrupamento
AnoCriacao	INT	-	1	Ano de Criação do Agrupamento

A tabela de dimensão **DimEscola** armazena informação relativa a todas as escolas do concelho de Guimarães (nome, número de alunos, lotação, numero de docentes, etc.). Na Tabela 28 podem visualizar-se com mais detalhe todos os atributos desta dimensão. Estes dados foram cedidos pela Câmara Municipal de Guimarães num ficheiro Excel.

Tabela 28 - Tabela de dimensão DimEscola

DimEscola				
Atributo	Tipo	Chave	SCD	Descrição
IDEscola	INT	PK	-	Identificador único de cada registo
Nome	nvarchar(255)	-	1	Nome da Escola
Tipo	nvarchar(100)	-	2	Nível de Ensino lecionado
Lotacao	nvarchar(100)	-	2	Capacidade máxima da Escola
NumSalasAula	nvarchar(100)	-	2	Número de Salas de Aula
PavilhaoDesportivo	nvarchar(100)	-	2	Indicação da existência ou não de Pavilhão Desportivo
Refeitório	nvarchar(100)	-	2	Indicação da existência ou não de Refeitório
NumDocentes	nvarchar(100)	-	2	Número de Docentes
NumNaoDocentes	nvarchar(100)	-	2	Número de Não Docente
AnoAbertura	INT	-	1	Ano de Abertura da Escola (yyyy)
AnoUltimaRemodelacao	INT	-	1	Ano em que foi efetuada a última remodelação (yyyy)
NumRemodelacoes	INT	-	1	Número de remodelações efetuadas
CP4	INT	-	1	Código Postal de 4 dígitos
CP3	INT	-	1	Código Postal de 3 dígitos
Freguesia	nvarchar(255)	-	1	Designação da Freguesia a que a escola pertence

É importante referir que esta dimensão inclui informação geográfica redundante (CP3 e CP4) isto porque, a tabela de factos FactSistemaEducativo não está a ser carregada o que impossibilita o cruzamento desta informação com a tabela DimLocalizacao. Por este motivo optou-se por colocar estes atributos para que fosse possível realizar análises de cariz geoespacial. O SCD dos atributos “Tipo”, “Lotação”, “NumSalasAula”, “PavilhaoDesportivo”, “Refeitório”, “NumDocentes” e “NumNaoDocentes” são do tipo 2 o que significa que no caso de ocorrer uma atualização dos atributos, o registo histórico é guardado, sendo criado uma nova linha de registo atualizado.

A tabela de dimensão **DimAluno** armazena informação relativa aos alunos que frequentam todas as escolas do concelho de Guimarães (sexo, data de nascimento, etc.). Na Tabela 29 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 29 - Tabela de dimensão DimAluno

DimAluno				
Atributo	Tipo	Chave	SCD	Descrição
IDAluno	INT	PK	-	Identificador único de cada registo
Sexo	nvarchar(255)	-	1	Designação do Sexo do Aluno
DataNascimento	Date	-	1	Data de Nascimento (yyyy-mm-dd)
NIF	INT	-	1	Número de Identificação Fiscal do Aluno
ParentescoEncEdu	nvarchar(100)	-	1	Grau de Parentesco do Encarregado de Educação
OrigemID_Aluno	INT	-	1	Chave Original

A tabela de dimensão **DimProfessor** armazena informação que identifica os docentes que lecionam nas escolas do concelho de Guimarães. Na Tabela 30 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 30 - Tabela de dimensão DimProfessor

DimProfessor				
Atributo	Tipo	Chave	SCD	Descrição
IDProfessor	INT	PK	-	Identificador único de cada registo
Identificacao	varchar(50)	-	1	Identificação do Docente

Segue-se a tabela de dimensão **DimBridgeProf** que está relacionada com a dimensão professor anteriormente descrita. Esta dimensão é uma *Bridge Table Dimension*, criada para que seja possível associar a uma linha da tabela de factos, por exemplo uma escola, a um grupo de professores. Nesta situação optou-se por criar uma entidade de grupo que se relaciona com a dimensão de múltiplo valor. Na Tabela 31 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 31 - Tabela de dimensão DimBridgeProf

DimBridgeProf				
Atributo	Tipo	Chave	SCD	Descrição
idGrupoProf	INT	PK	-	Identificador único de cada grupo
DimProfessor_idProfessor	INT	FK e PK	-	ID da DimProfessor

A tabela de dimensão **DimNaoDocente** armazena informação que identifica o pessoal não docente que desempenha funções nas escolas do concelho de Guimarães. Na Tabela 32 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 32 - Tabela de dimensão DimNaoDocente

DimNaoDocente				
Atributo	Tipo	Chave	SCD	Descrição
IDNaoDocente	INT	PK	-	Identificador único de cada registo
Identificacao	varchar(50)	-	1	Identificação do Funcionário
Funcao	varchar(50)	-	2	Designação da Função desempenhada

Segue-se a tabela de dimensão **DimBridgeNaoDocente** que está relacionada com a dimensão não docente anteriormente descrita pela chave estrangeira. Esta dimensão é uma *Bridge Table Dimension*, criada para que seja possível associar a uma linha da tabela de factos, por exemplo uma escola, a um grupo de não docentes. Na Tabela 33 é possível visualizar todos os atributos desta dimensão.

Tabela 33 - Tabela de dimensão DimBridgeNaoDocente

DimBridgeNaoDocente				
Atributo	Tipo	Chave	SCD	Descrição
idGrupoNaoDocente	INT	PK	-	Identificador único de cada grupo
DimNaoDocente_idNaoDocente	INT	FK e PK	-	ID da DimNaoDocente

A tabela de dimensão **DimNivelEnsino** armazena informação acerca dos níveis de ensino legislados em Portugal. Na Tabela 34 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 34 - Tabela de dimensão DimNivelEnsino

DimNivelEnsino				
Atributo	Tipo	Chave	SCD	Descrição
IDNivelEnsino	INT	PK	-	Identificador único de cada registo
Designacao	varchar(50)	-	1	Designação do Nível de Ensino

A tabela de dimensão **DimNota** armazena informação acerca das notas de avaliação possíveis quer sejam elas quantitativas ou qualitativas. Na Tabela 35 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 35 - Tabela de dimensão DimNota

DimNota				
Atributo	Tipo	Chave	SCD	Descrição
IDNota	INT	PK	-	Identificador único de cada registo
Designacao	varchar(50)	-	1	Designação da Nota da escala de avaliação

A tabela de dimensão **DimPeriodo** armazena informação sobre os períodos letivos do 1º, 2º e 3º ciclos de escolaridade. Na Tabela 36 é possível visualizar em detalhe todos os atributos desta dimensão.

Tabela 36 - Tabela de dimensão DimPeriodo

DimPeriodo				
Atributo	Tipo	Chave	SCD	Descrição
IDPeriodo	INT	PK	-	Identificador único de cada registo
Designacao	varchar(50)	-	1	Designação do Período Letivo

A tabela de dimensão **DimSituacao** armazena informação relativa às situações possíveis da avaliação final do aluno (transitou, aprovado, etc.). Na Tabela 37 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 37 - Tabela de dimensão DimSituacao

DimSituacao				
Atributo	Tipo	Chave	SCD	Descrição
IDSituacao	INT	PK	-	Identificador único de cada registo
DesignacaoSituacao	varchar(50)	-	1	Situação do aluno quanto ao seu aproveitamento escolar
Sigla	varchar(3)	-	1	Abreviatura da Situação

A tabela de dimensão **DimAnoEscolar** armazena informação referente aos anos escolares existentes. Na Tabela 38 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 38 - Tabela de dimensão DimAnoEscolar

DimAnoEscolar				
Atributo	Tipo	Chave	SCD	Descrição
IDAnoEscolar	INT	PK	-	Identificador único de cada registo
AnoEscolar	varchar(50)	-	1	Designação do Ano de Escolaridade

A tabela de dimensão **DimEscalao** armazena informação sobre os diferentes níveis de escalão social que podem ser atribuídos aos alunos. Na Tabela 39 é possível visualizar em detalhe todos os atributos desta dimensão.

Tabela 39 - Tabela de dimensão DimEscalao

DimEscalao				
Atributo	Tipo	Chave	SCD	Descrição
IDEscalao	INT	PK	-	Identificador único de cada registo
Escalao	varchar(50)	-	1	Designação do Escalão Social

A tabela de dimensão **DimTurma** armazena informação sobre as turmas existentes nas escolas do concelho de Guimarães. Na Tabela 40 é possível visualizar com mais detalhe todos os atributos desta dimensão.

Tabela 40 - Tabela de dimensão DimTurma

DimTurma				
Atributo	Tipo	Chave	SCD	Descrição
IDTurma	INT	PK	-	Identificador único de cada registo
Turma	varchar(50)	-	1	Designação da Turma

Terminada a descrição de todas as tabelas de dimensão, excetuando as *bridge tables*, é de notar que não foram utilizadas chaves compostas. Foram utilizadas, por questões de desempenho e prevenção de duplicação de chaves, chaves substitutas (*Surrogate Key*), atribuídas de forma sequencial (iniciando em 1) conforme os registos foram sendo inseridos. Contudo, para algumas tabelas de dimensão, foram armazenadas as chaves originais (OrigemID_###) por questões de comparação durante as transformações no processo ETL.

É importante também referir que o modelo apresentado inclui tabelas que não serão carregadas por falta de dados. Porém, tomou-se a decisão de as manter e de apresentar no

trabalho associado a esta dissertação um modelo preparado para o cenário de disponibilidade de dados.

4.1.2. Processo ETL

Feita a análise exploratória dos dados operacionais e desenhado o modelo de dados do *Data Warehouse* apresentado nas secções anteriores, procedeu-se ao desenvolvimento do processo ETL. Nesta fase, e para que seja possível a criação do *Data warehouse* que suporte a tomada de decisão, é necessário proceder à extração dos dados dos sistemas OLTP, para que sejam submetidos a um processo de tratamento, e posterior carregamento para o *Data Warehouse*. A ferramenta utilizada para esta fase do projeto foi o *Talend Open Studio for Data Integration* que permite a criação *jobs* (packages de fluxo de dados) para desenvolver as tarefas de extração, transformação, integração e carregamento dos dados.

Para demonstrar, de uma forma mais simplificada, como o processo foi desenvolvido, é apresentada na Figura 34 a metodologia seguida para realizar o processo ETL necessário.

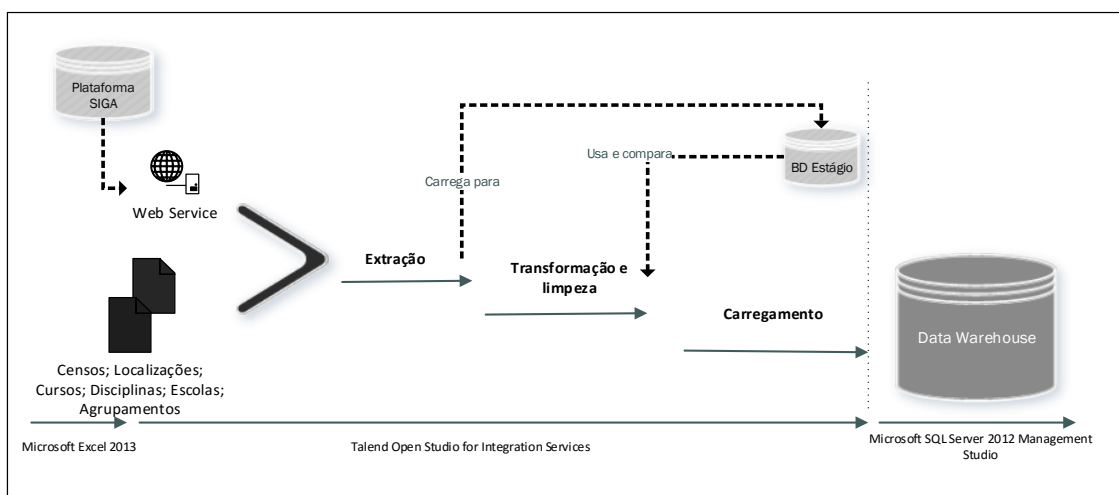


Figura 34 - Metodologia do processo ETL

Em primeiro lugar importa referir que tipo de fontes de dados foram utilizadas para alimentar o sistema. Como é possível verificar na Figura 34, existem dois tipos de fontes de dados distintas. A primeira (Plataforma SIGA) diz respeito a um sistema OLTP que disponibiliza os dados dos alunos via *web service*. Posto isto, optou-se pela criação de uma base de dados intermédia (BD Estágio) de suporte ao processo ETL para armazenar todos os dados provenientes deste sistema. A segunda fonte de dados diz respeito a vários ficheiros Excel que foram recolhidos, nomeadamente informações das escolas, agrupamentos de escolas, dados

dos censos, dados das localizações dos CTT, dados dos cursos e disciplinas lecionadas. Neste caso os dados não foram carregados para a base de dados de estágio. De uma forma geral, o processo ETL inicia-se com a extração dos dados para a base de dados intermédia, seguindo-se todas as operações de transformação e limpeza necessárias. Os dados que resultam destas transformações são depois carregados para o *Data Warehouse*.

Após explicada e apresentada a metodologia utilizada para o processo ETL, será, seguidamente, detalhado todo o processo ETL.

O desenvolvimento do processo ETL iniciou-se pela criação dos *scripts* de criação da base de dados de estágio e do *Data Warehouse*. Posteriormente, procedeu-se à criação do fluxo de dados com a invocação do *webservice* para a extração dos mesmos da plataforma SIGA da *Edubox*. O próximo passo foi desenvolver o processo ETL para as tabelas de dimensão e, posteriormente, desenvolver os processos ETL para as tabelas de factos. Na Figura 35 é possível verificar, de uma forma geral, todo o processo ETL.

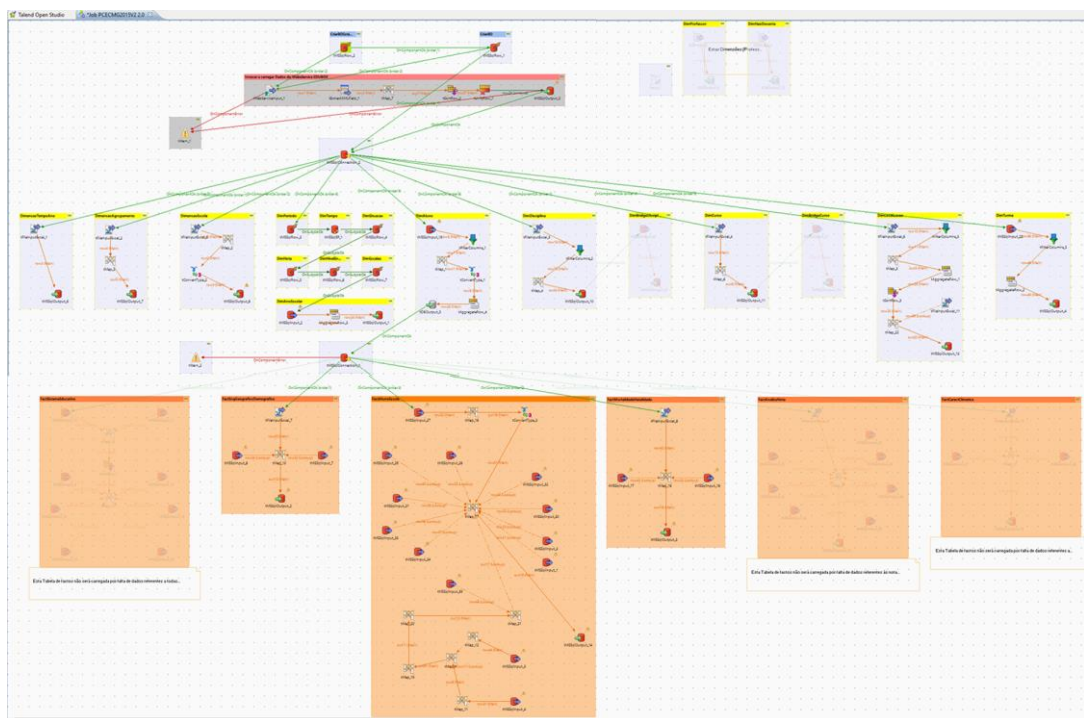


Figura 35 – Perspetiva Geral do Processo ETL implementado

O processo apresenta esta estrutura e tem um aspeto hierarquizado, pois importa referir que as tabelas de factos relacionam-se com várias tabelas de dimensão pela respetiva chave

estrangeira. Este é o motivo pela qual as tabelas de factos são carregadas apenas no fim do processo ETL.

O processo ETL desenvolvido é um único fluxo de dados que interliga todos os sub-fluxos existentes, desde a criação do *Data Warehouse* até ao carregamento das dimensões. Como é possível verificar na parte superior da Figura 35, o processo inicia-se com a criação de duas bases de dados como se pode ver mais detalhadamente na Figura 36.

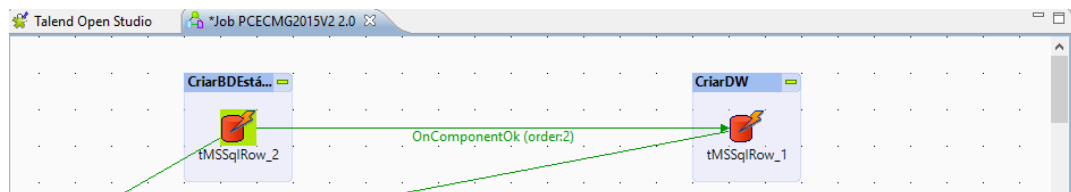


Figura 36 - Componentes de criação das bases de dados

O fluxo geral dos dados inicia-se com a criação da base de dados de estágio responsável por armazenar os dados provenientes da plataforma SIGA. Seguidamente é executada a segunda componente do processo para a criação do *Data Warehouse*. Estas duas componentes executam instruções *SQL* e as respetivas scripts podem ser consultadas nos anexos B2 e B3.

Após ter a bases de dados de estágio e o DW preparadas para receber os dados, segue-se o fluxo de invocação do *web service* e respetiva extração dos dados para a base de dados de estágio. Na Figura 37 é possível verificar todo o fluxo de dados desde a invocação até à inserção dos dados na tabela.

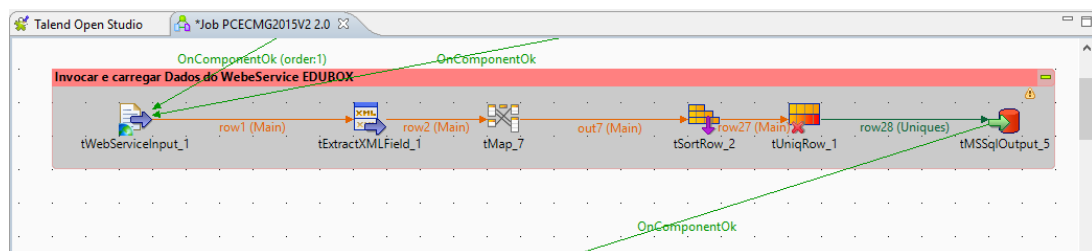


Figura 37 - Fluxo de Invocação e Extração dos Dados

O processo inicia-se com a chamada do *web service*, onde foram indicados como parâmetros o URL do WSDL⁵ bem como o método e um *token* de acesso. Após a receção do XML⁶, são

⁵ WSDL (*Web Service Description Language*) é um documento XML que descreve um *Web Service*. O conteúdo do WSDL descreve os métodos fornecidos pelo *Web Service* e quais as formas para lhes aceder.

⁶ XML (*Extensible Markup Language*) é um formato de texto simples para a criação de documentos com dados organizados hierarquicamente com o objetivo de facilitar a partilha de informação na Web.

definidos todos os campos que se pretendem obter, passando depois por um processo de tratamento dos dados que incide essencialmente sobre os valores a nulo, tratamento de datas e de caracteres especiais. Para finalizar, as linhas são ordenadas (por idUser e InitialDate), isto porque será necessária a ordenação para a implementar o facto FlagMatriculado da tabela de factos FactAlunoEscola. Para tal, ordenaram-se os registos para se analisar a data de entrada na turma do último registo de cada aluno. Caso existam linhas duplicadas, estas serão eliminadas para que se proceda à inserção na base de dados de estágio.

A fase que se segue diz respeito ao fluxo de carregamento das tabelas de dimensão. Na Figura 38 é possível observar todos os sub-fluxos de carregamento de cada uma das dimensões.

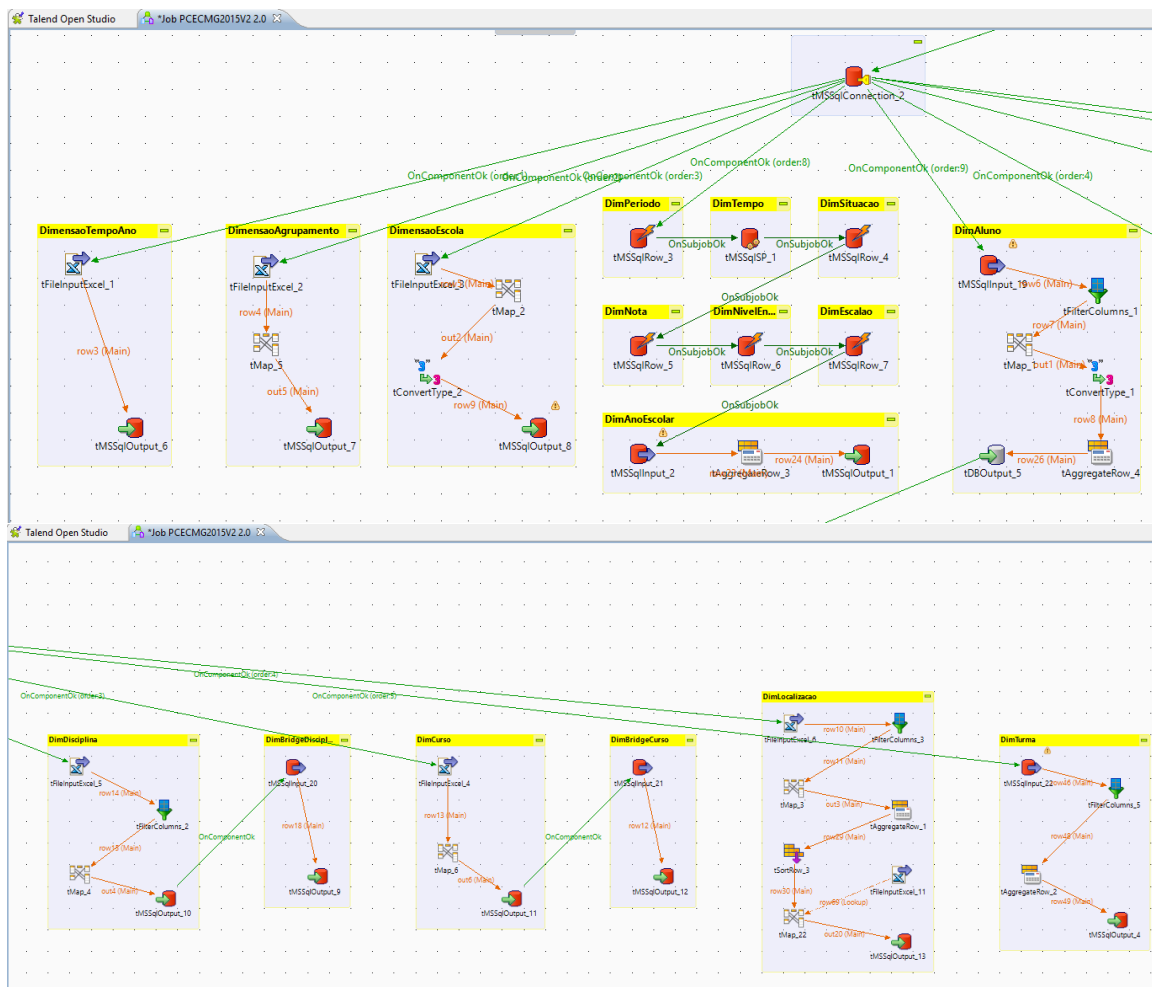


Figura 38 - Carregamento das Tabelas de Dimensão

As tabelas de dimensão TempoAno, Agrupamento, Escola, Disciplina, Curso e Localização foram carregadas através de ficheiros Excel disponibilizados pelo município de Guimarães e obtidas as localizações através das bases de dados dos CTT. Para além de algum tratamento

prévio realizado diretamente nos ficheiros por forma a homogeneizar e uniformizar determinadas colunas, foi necessário realizar algum tratamento essencialmente ao nível de registos nulos. A dimensão Localização foi carregada a partir dos dados obtidos de uma base de dados dos CTT com todas as localizações de Portugal. Foi necessário filtrar essas localizações e uniformizar a coluna localidade, criando uma nova coluna agrupada com as designações correntes das localidades existentes devido às uniões de freguesias. Dada a componente geoespacial do projeto, foi necessário obter as coordenadas geográficas (latitude e longitude) conseguidas através dos códigos postais e moradas existentes no Excel. Para isso, foi desenvolvido código em java⁷ que através da API⁸ da *Goggle*, converte os códigos postais e a localidade em coordenadas geográficas (latitude e longitude) que pode ser consultado em anexo (Anexo B4 – Script java Converter em Coordenadas Geográficas). Este código em java foi desenvolvido no âmbito de outra dissertação de mestrado que está a decorrer na mesma área de estudos e que já foi referida anteriormente. Dada a limitação imposta pela API que impossibilita a conversão de mais 2500 registos diários, optou-se por não incluir no fluxo a execução deste trecho de código pois iria inviabilizar o correto preenchimento da dimensão. A solução passou pela criação de um Excel com as coordenadas geográficas, procedendo-se depois à junção dos dois ficheiros e respetivo carregamento, como se pode verificar na Figura 39.

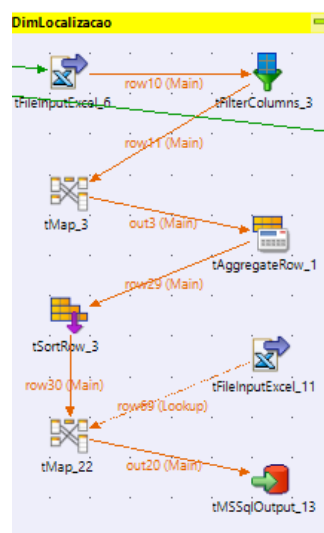


Figura 39 - Carregamento da dimensão Localização

⁷ Java é uma linguagem de programação e uma plataforma computacional.

⁸ API (*Application Programming Interface*) é um conjunto de padrões de programação disponibilizados por uma aplicação e que permitem a implementação das suas funcionalidades em aplicações externas.

As tabelas de dimensão Período, Situação, Nota, Nível de Ensino e Escalão foram carregadas através de scripts SQL (Figura 40), pois tratam-se de tabelas com menor complexidade e permitem, desta forma, reduzir a utilização de recursos de memória e tempo de processamento do ETL.

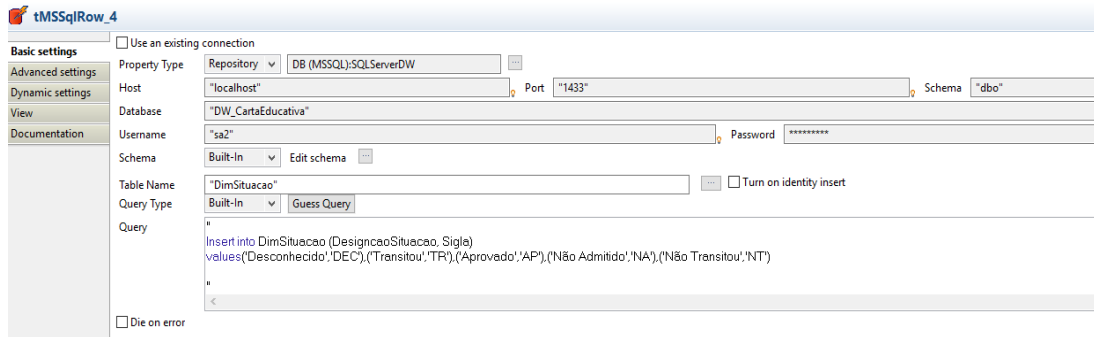


Figura 40 - Exemplo do carregamento de uma Tabela de Dimensão

Por último, as tabelas de dimensão Ano Escolar, Turma e Aluno foram carregadas com recurso aos dados armazenados na base de dados de estágio. A dimensão Aluno foi carregada com os atributos Data de Nascimento, Sexo, Grau de Parentesco do Encarregado de Educação, NIF do aluno e o idUser. As restantes colunas foram ignoradas por se tratar de informação de cariz confidencial ou não essencial. Foram realizados alguns tratamentos ao nível das datas e transformados os valores nulos. A cada transição de turma é criada uma nova linha de registo para o mesmo aluno, por este motivo foi necessário agrupar os alunos para se obter uma linha de registo para cada um.

É necessário referir que as tabelas de dimensão DimPofessor, DimNaoDocente, DimBridgeDisciplina e DimBridgeCurso, além de estarem presentes no modelo de dados e no processo ETL, encontram-se desativadas por falta de dados não sendo portanto carregadas.

Terminado todo processo de carregamento das dimensões, segue-se por último no fluxo de dados o carregamento das tabelas de factos. É importante referir que as tabelas de factos FactSistemaEducativo, FactAnaliseNota e FactCaracterizaçãoClimatica além de estarem presentes no modelo de dados do *Data Warehouse* e no processo ETL, encontram-se desativadas não sendo carregadas por falta de dados. Posto isto, seguidamente serão

apresentados apenas os fluxos de dados das tabelas de factos FactEnqGeograficoDemografico, FactAlunoEscola e FactMortalidadeNatalidade.

O processo de ETL referente às tabelas de factos é mais complexo comparado com o das tabelas de dimensão, isto porque estas necessitam de armazenar as chaves estrangeiras das várias dimensões, bem como os factos numéricos que necessitam de ser calculados através de um tratamento especial. A Figura 41 mostra o fluxo dos dados da tabela de factos FactEnqGeograficoDemografico.

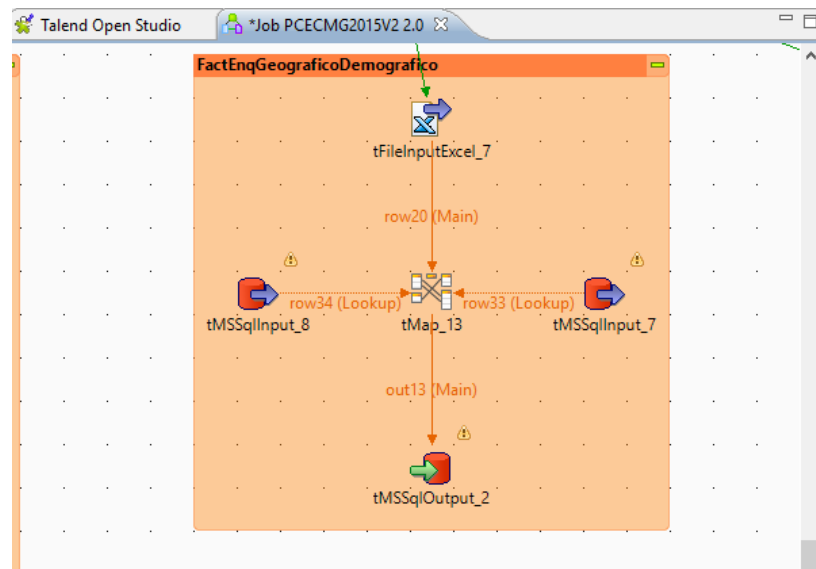


Figura 41 - Fluxo ETL da tabela de factos FactEnqGeograficoDemografico

Para o carregamento desta tabela de factos, foram extraídos os dados de um ficheiro Excel com a informação dos censos realizados nos anos de 2001 e 2011.

Nesta tabela existem chaves estrangeiras para as dimensões DimLocalizacao e DimTempoAno, sendo necessário carregar todos os factos bem como as respetivas chaves estrangeiras (FK) destas tabelas de dimensão. Assim, após a extração dos dados, é feito o *mapping* para preencher os factos e proceder à comparação, para cada registo, dos atributos (ano e localidade) que irão referenciar as respetivas dimensões antes de serem carregados.

Segue-se, na Figura 42, o fluxo dos dados, desde a extração até ao carregamento da tabela de factos FactAlunoEscola.

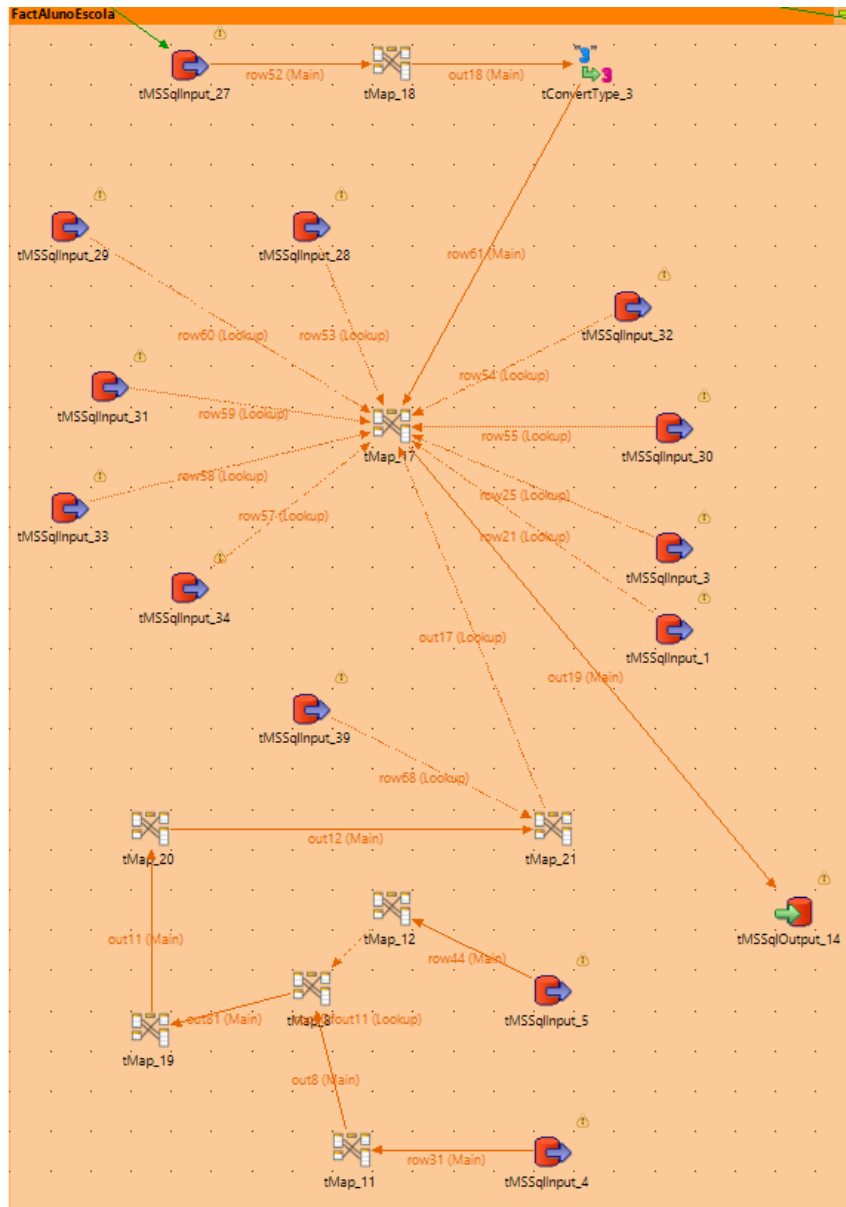


Figura 42 - Fluxo ETL da tabela de factos FactAlunoEscola

Os dados são extraídos da base de dados de estágio que armazena toda a informação dos alunos, obtida a partir do *web service* disponibilizado. Após a realização de algumas transformações, são feitas todas as comparações com as tabelas de dimensão e carregados os factos. Os factos são valores numéricos e normalmente são obtidos a partir do cálculo de vários atributos, para determinar por exemplo, se o aluno está matriculado no ano corrente ou se recebe apoio social, tal como é possível analisar na Figura 43.

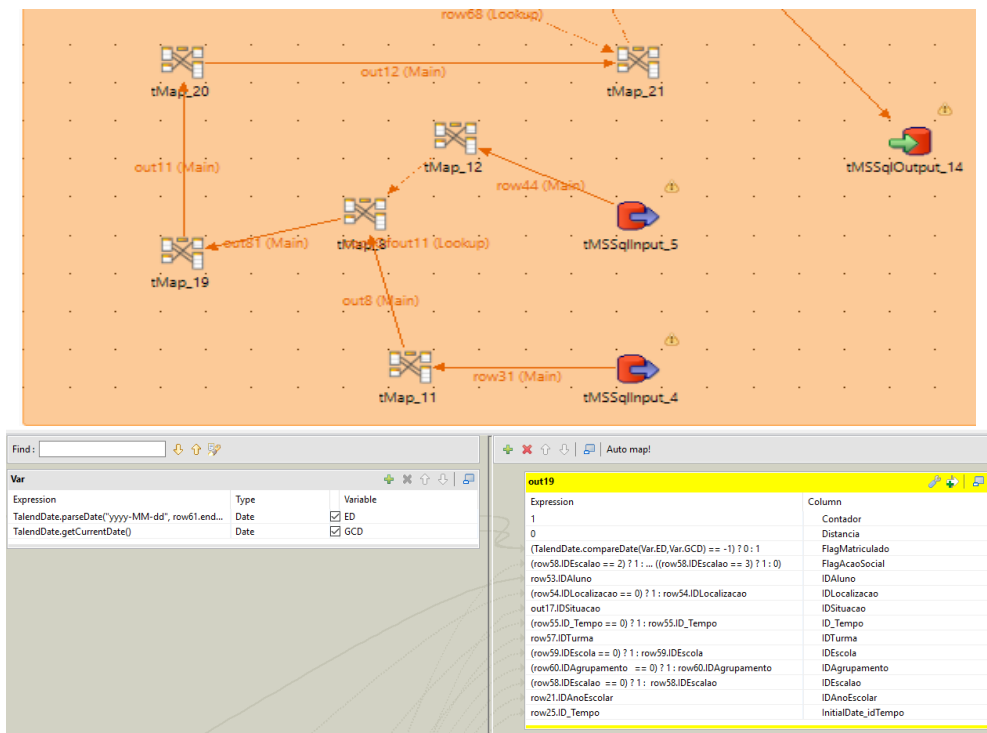


Figura 43 - Parte do Fluxo que transforma os factos da FactAlunoEscola

Devido ao facto de não existir um atributo que indique qual a situação do aluno em relação ao ano anterior (Aprovado ou Não Aprovado), foi necessário proceder à análise das datas de entrada nas turmas (initialDate) e verificar se no ano seguinte o aluno mudou de turma e de ano escolar. Para isso, foi necessário ordenar o fluxo pelo identificador único do aluno (idUser). Foram criados inicialmente dois fluxos, com duas chamadas à mesma fonte de dados, para que fosse feito um *join* com a criação de uma coluna com o valor do ano seguinte de cada uma das linhas. Seguidamente, é feita a comparação de cada linha com a linha seguinte e feita a verificação do idUser e AnoEscolar.

Nesta tabela existem chaves estrangeiras para as dimensões DimLocalizacao, DimTempo, DimAluno, DimAgrupamento, DimEscola, DimEscalao, DimTurma, DimSituação e DimAnoEscolar sendo necessário carregar todas as chaves estrangeiras destas tabelas de dimensão.

Por último, na Figura 44 é apresentado o fluxo de dados para o carregamento da tabela de factos FactMortalidadeNatalidade, com dados referentes à evolução da natalidade durante os últimos anos.

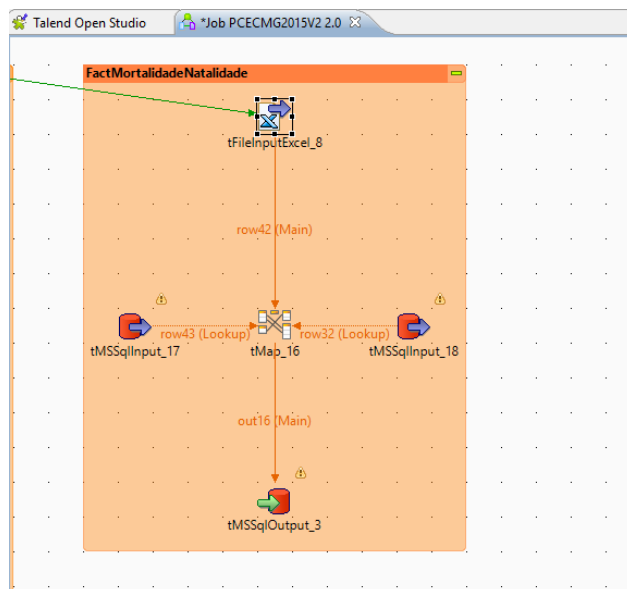


Figura 44 - Fluxo ETL para transformação dos factos da FactMortalidadeNatalidade

Para o carregamento desta tabela de factos foram extraídos os dados de um ficheiro Excel com a informação do número de nados-mortos e vivos nos últimos quatro anos. Nesta tabela existem chaves estrangeiras para as dimensões DimLocalizacao e DimTempo, portanto é necessário carregar todos os factos bem como as respetivas chaves estrangeiras (FK) destas tabelas de dimensão. Assim, após a extração dos dados, é feito o *mapping* para preencher os factos e a respetiva comparação, para cada registo, dos atributos (ano e código da localidade) que irão referenciar as respetivas dimensões antes de serem carregados.

É importante referir que os mecanismos de refrescamento do sistema não foram implementados, motivado pelas dificuldades na obtenção e análise dos dados, que se traduziu num maior esforço nessa fase. Isso levou a um atraso geral do projeto inviabilizando a implementação dos mesmos. Posto isto, o sistema implementado, corresponde apenas ao primeiro carregamento dos dados para o *Data Warehouse*. Também pelos motivos enumerados anteriormente, e pela dificuldade em carregar todo modelo, foi impossível a implementação de todos os atributos do tipo 2 do SCD.

Após a execução de todo o processo ETL, todos os dados (14968 registos) que foram extraídos e que passaram pelo fluxo de dados foram transformados e devidamente carregados para o *Data Warehouse*, ficando agora disponíveis para serem acedidos pelos sistemas OLAP implementados no âmbito do projeto.

4.1.3. Processamento Analítico de Dados

Feita a descrição do modelo de dados do *Data Warehouse* e feito o carregamento do mesmo através do processo de ETL, também descrito anteriormente, nesta subsecção, pretende-se apresentar as análises e resultados obtidos através dos sistemas de processamento OLAP.

Desta forma, estão reunidas as condições necessárias para que sejam implementados os cubos OLAP. Para isso, nesta fase do projeto foi utilizada a componente *Analysis Services do Microsoft SQL Server*.

Numa fase inicial da implementação, foi criada a ligação (*OLE DB SQL Server Native Client 11.0*) às fontes de dados dos cubos, nomeadamente ao *Data Warehouse* que está alojado no *SQL Management Studio*. A partir desse *data source*, foram criadas as vistas (*data source views*) que funcionam como abstrações da fonte de dados. Este passo tem como objetivo dar um maior controlo e manipulação, de forma independente, das fontes de dados podendo ser criadas várias conforme as necessidades do projeto. Para esta implementação foi apenas criada uma vista com as tabelas de factos que foram carregadas e respetivas tabelas de dimensão.

Analizadas as necessidades do projeto verificou-se que seria necessário nesta fase, criar 3 cubos de acordo com o número de tabelas de factos existentes. Deu-se início à construção dos cubos OLAP com a seleção da tabela de factos e respetivas tabelas de dimensão para os seguintes cubos OLAP:

- FactNatalidadeMortalidade – incorpora a Estrela FactNatalidadeMortalidade para que seja possível a análise dos números da mortalidade e natalidade sobre as perspetivas fornecidas;
- FactEnqGeograficoDemografico – incorpora a Estrela FactEnqGeograficoDemografico para que seja possível a análise sobre as características demográficas e geográficas da população sobre as diversas perspetivas fornecidas.
- FactAlunoEscola – incorpora a Estrela FactAlunoEscola para que seja possível a análise do aluno com a respetiva escola sobre as diversas perspetivas fornecidas.

As tabelas de dimensão que incluem hierarquias são:

- DimTempo
- DimLocalizacao
- DimAgrupamento
- DimAluno
- DimEscola

A Figura 45 e Figura 46 mostram os atributos e as hierarquias definidas para as tabelas de dimensão DimTempo e DimLocalizacao respetivamente.

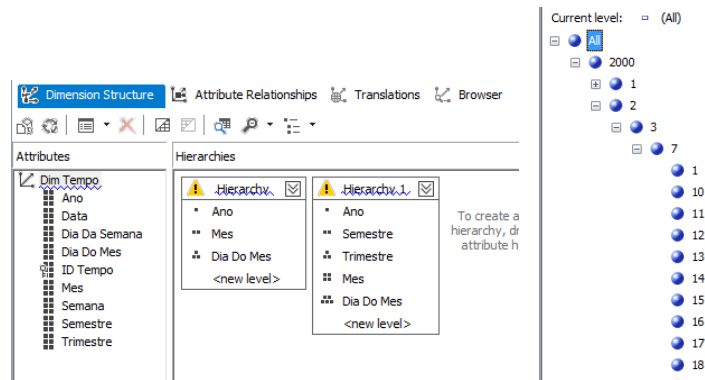


Figura 45 - Hierarquia da tabela de dimensão DimTempo

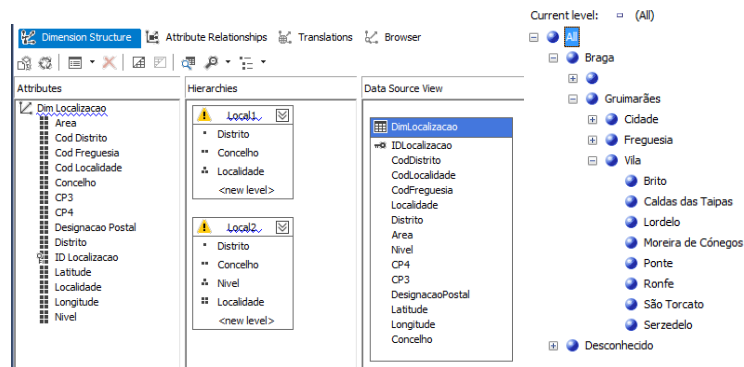


Figura 46 - Hierarquia da tabela de dimensão DimLocalizacao

Nas figuras apresentadas anteriormente, é também ilustrada a pesquisa ao longo das hierarquias de ambas as dimensões. Exemplificando, na utilização da hierarquia localização {Distrito – Concelho – Localidade} é possível analisar os alunos por Distrito e, através da operação *Drill-Down*, detalhar a análise por concelho até chegar especificamente à localidade da residência dos mesmos.

O sistema de processamento analítico de cubos OLAP desenvolvido assenta numa arquitetura MOLAP, pois este caracteriza-se pela utilização de bases de dados multidimensionais que lhe permite obter várias vistas multidimensionais sobre os dados. Em questões de desempenho e indexação de dados é extremamente eficaz e indicado para o projeto.

O processamento deste sistema OLAP pode ocorrer sempre que assim for necessário aquando da conclusão do processo ETL, a qual se prevê uma periodicidade anual.

A implementação deste sistema de processamento analítico tornou possível a utilização de ferramentas analíticas dos dados, através das quais é possível criar gráficos, tabelas, mapas. Na subsecção seguinte são apresentados exemplos da utilização deste tipo de ferramentas analíticas na exploração dos dados a partir do sistema OLAP desenvolvido.

4.1.4. Aplicações de *Front-End*

Realizadas todas as etapas do sistema de *Business Intelligence* proposto, resta por fim apresentar a implementação das aplicações de *front-end*. Esta subsecção tem como objetivo mostrar como é que os responsáveis pela gestão da rede escolar vão poder aceder à informação sobre a rede, fazendo assim referência às aplicações de *front-end* utilizadas para a componente de visualização de dados do projeto.

O objetivo para esta fase prende-se com a concretização de mecanismos de análise que permitam o acesso à informação de forma simples e intuitiva. Para concretizar este objetivo foram implementadas aplicações que fossem capazes de permitir o acesso local e via web dos dados para que os utilizadores possam usar e criar os relatórios, bem como consultas desejadas através de operações *drill-down* e *roll-up* de qualquer um dos cubos OLAP implementados. Para tal, foram utilizadas ferramentas como o Microsoft Excel 2013 com recurso à componente *PowerPivot* e utilizada a aplicação Microsoft *PowerBI*. Através destas ferramentas foi possível cumprir com os requisitos de simplicidade com grande versatilidade de análises que permite a criação de *Dashboards* e *Reports* relevantes para a gestão da rede escolar do concelho de Guimarães.

É importante referir que a aplicação *PowerBI* que a *Microsoft* disponibiliza não é *open source*, tendo sido obtida, pelo seu baixo custo, uma licença para que fosse possível desenvolver todas as análises pretendidas. Optou-se por esta ferramenta pois cumpre com todos os requisitos pretendidos que serão demonstrados de seguida.

Para a elaboração das diversas Dashboards e Relatórios foram tomados em consideração os seguintes aspetos:

- Facilidade e simplicidade na utilização e manuseamento das ferramentas;
- Facilidade na partilha das análises com recurso a visualização das mesmas através de interfaces móveis;
- Interface simples e agradável com vários recursos analíticos, como vários tipos de gráficos, tabelas e mapas;
- Possibilidade de criação e definição de filtros nas análises realizadas.

A implementação das análises efetuadas foi pensada tendo em conta a estrutura da atual carta educativa vigente no município de Guimarães. Como tal, foi efetuada uma divisão das análises de acordo com as três áreas distintas que foram alvo de análise (Caracterização Demográfica, Enquadramento Geográfico e Sistema Educativo). Para criar os diversos relatórios e análises foram utilizadas para cada uma das áreas os seguintes cubos OLAP:

- Caracterização Demográfica – Cubo FactEnqGeograficoDemografico que permite analisar os indicadores demográficos do concelho;
- Enquadramento Geográfico – Cubos FactEnqGeograficoDemografico e FactMortalidadeNatalidade que permitem a análise dos fatores geográficos e respetiva caracterização da região;
- Sistema Educativo - Cubos FactEnqGeograficoDemografico e FactSistemaEducativo que permitem a análise aos indicadores relevantes no estudo da rede escolar do concelho de Guimarães.

Para evitar multiplataformas e tornar este processo mais simples, pensou-se numa solução integrada através do Microsoft Excel 2013. Para isso, foi estabelecida a ligação aos cubos OLAP e criada uma interface simples e intuitiva que permita ao utilizar navegar pelas três áreas de análise (Figura 47) e visualizar todos os *reports* e *dashboards* disponibilizados de forma dinâmica.



Figura 47 - Interface do Sistema de *Business Intelligence* Implementado

Após a implementação das análises, na aba *PowerPivot* da aplicação, procedeu-se à gestão dos dados criando várias consultas que serão necessárias para a utilização na aplicação *PowerBI*. É importante realçar que o ficheiro Excel desenvolvido foi guardado na plataforma *cloud* da *Microsoft OneDrive* para que fosse possível aceder aos dados, por parte do *PowerBI*, criando desta forma uma integração entre as plataformas. Optou-se por esta estratégia para que fosse possível a conexão, através do *PowerBI*, aos cubos OLAP implementados. As conexões aos dados desta ferramenta são um pouco limitadas, sendo a ligação ao Excel, via *cloud*, a solução ideal para se obter uma integração entre plataformas.

Para permitir o acesso *ad-hoc* aos dados e realizar as operações necessárias de análise, utilizou-se a ferramenta *Microsoft PowerBI*. Esta disponibiliza duas formas de desenvolvimento, uma plataforma *Web* e uma aplicação *Desktop* que são praticamente idênticas. Esta última destaca-se, pois permite fazer a gestão das fontes de dados, permitindo inclusive criar novas colunas, campos calculados bem como gerir as relações entre as tabelas (Figura 48).

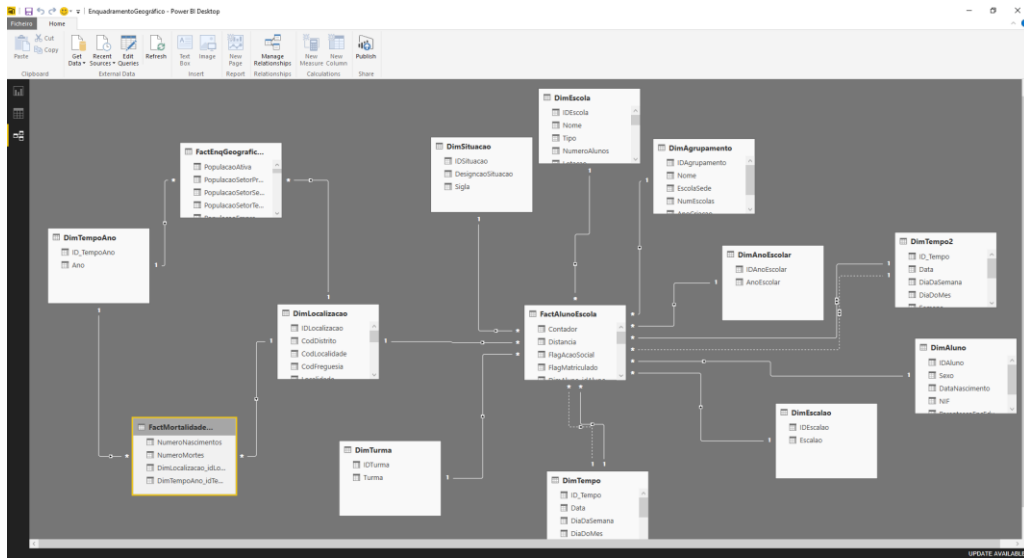


Figura 48 - Aplicação Desktop da ferramenta *PowerBI*

Através da aplicação *desktop* da ferramenta *PowerBI* foram criadas duas conexões aos dados, uma através do *cloud Microsoft OneDrive* ao ficheiro Excel lá armazenado e outra ao *Data Warehouse* através de uma conexão ao *SQL Server*. Com base nestas conexões foram criadas três *dashboards* (Figura 49) para a análise dos dados disponíveis. Para além destas, foram também criados uma série de relatórios que permitem uma análise integrada e mais detalhada sobre determinado foco de análise.

Na Figura 49 é ilustrada a *Dashboard* referente às análises da área do Sistema Educativo do concelho de Guimarães.

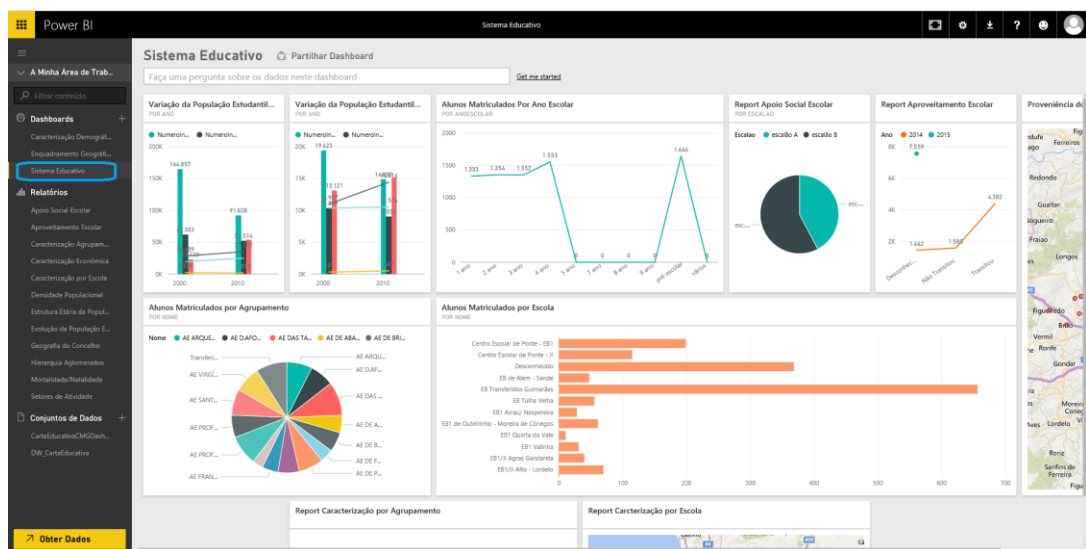


Figura 49 – Vista da *Dashboard* “Sistema Educativo” da plataforma *PowerBI*

Através desta *dashboard* é possível realizar diversas análises que permitem verificar a evolução da população estudantil do concelho. Através destas é possível analisar as variações de ano para ano dos alunos matriculados, quer ciclo de estudos, ano escolar, agrupamento e por escola. Para além destas, a ferramenta disponibiliza uma funcionalidade bastante interessante que permite ao utilizador questionar, na parte superior da página, o sistema acerca de alguma análise que este pretenda (por exemplo Matriculados *by* local).

O *PowerBi* permite também criar uma *tag* nas análises da *dashboard* que possibilita, ao clicar na respetiva análise, direcionar para o respetivo relatório obtendo, desta forma, um conjunto de análises mais detalhadas sobre determinado assunto. Tomando como exemplo a Figura 49, é de notar que existem duas análises (*Report* Apoio Social Escolar e *Report* Aproveitamento Escolar) que aquando da sua seleção direcionam para os respetivos relatórios de apoio social escolar e aproveitamento escolar.

Na Figura 50 é ilustrado o relatório referente às análises, com maior nível de detalhe, do aproveitamento escolar do concelho de Guimarães.

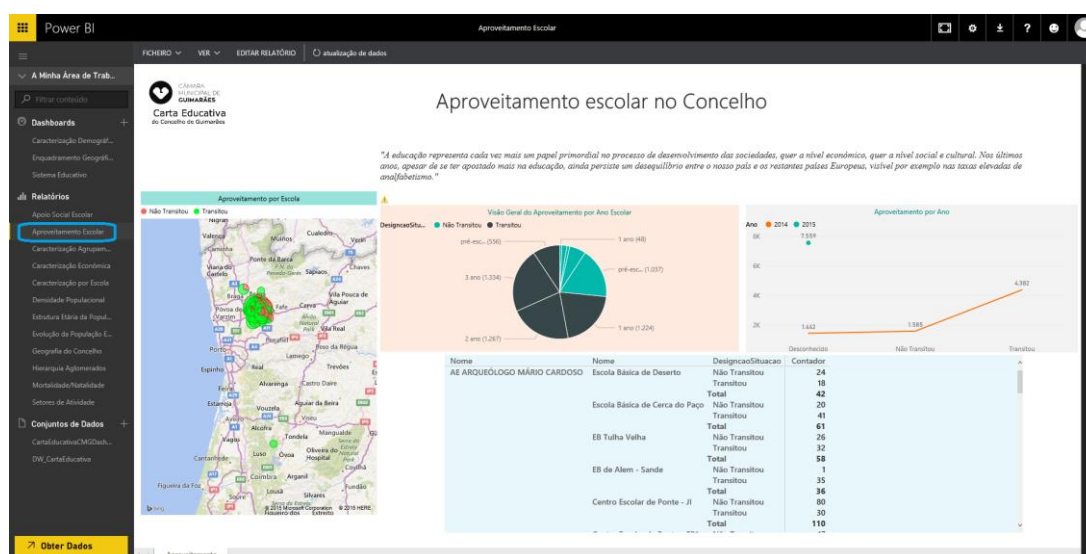


Figura 50 - Relatório do Aproveitamento Escolar do Concelho

O relatório da Figura 50 apresenta um conjunto de análises que permite formular conclusões acerca do aproveitamento escolar dos alunos quer por ano civil, ano escolar ou por escola com o respetivo agrupamento de escolas. A acompanhar todos estes relatórios pode verificar-se, que na sua grande maioria, a componente geográfica com a inclusão de mapas. Neste relatório pode verificar-se a inclusão de um mapa onde se pode analisar o número de alunos que

transitaram e não transitaram por cada escola do concelho. É importante referir que todas estas análises são dinâmicas, ou seja, clicando por exemplo em cima de um dos círculos do mapa é selecionada a respetiva escola nas restantes análises do relatório.

Para além deste, foram criados uma série de outros relatórios com diversas análises a outras áreas. Estes podem ser consultados em anexo (Anexo D – Relatórios nas Aplicações *Front-End*) bem como as outras duas *dashboards* (Anexo C – *Dashboards* nas Aplicações *Front-End*).

Apesar de toda esta integração e facilidade de acesso à informação via web, optou-se por tirar total partido de todas as potencialidades que a ferramenta *PowerBI* oferece. Para isso, esta ferramenta disponibiliza uma aplicação para dispositivos móveis que permite facilitar ainda mais o acesso à informação a partir de qualquer lugar. A aplicação *mobile* está disponível gratuitamente tanto para dispositivos *android* e *ios* nas respetivas lojas. Esta possibilita ver as análises que integram as *dashboards*, permitindo a sua seleção e respetiva navegação pelos diferentes tipos de gráficos e mapas (Figura 51).

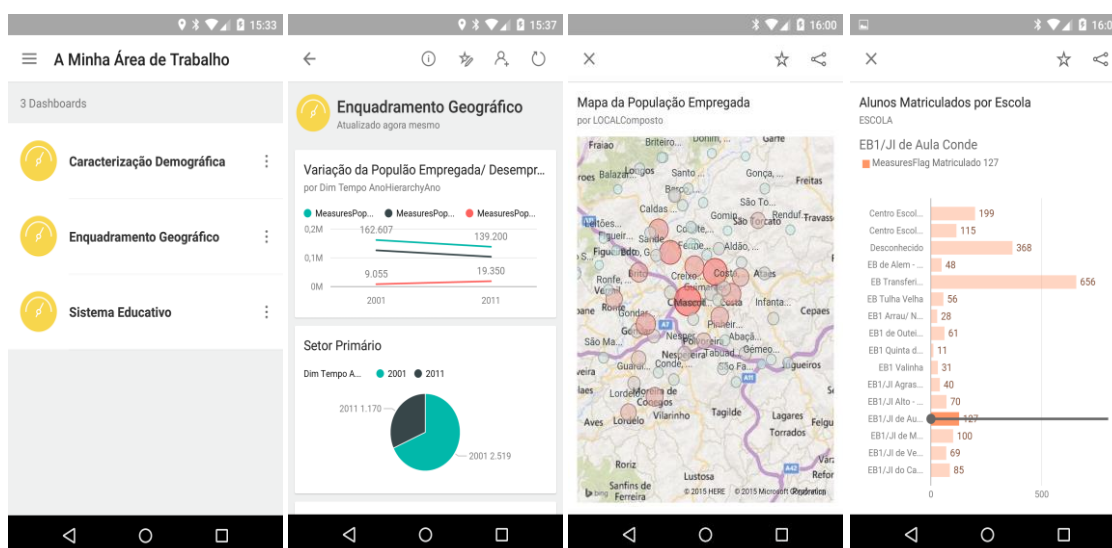


Figura 51 - Vistas das Dashboards na Aplicação Mobile

Na Figura 51 apresentam-se as várias vistas da aplicação móvel *PowerBI*. A aplicação apresenta inicialmente a área de trabalho onde se pode selecionar uma das três *dashboards* criadas. Após a seleção da *dashboard* Enquadramento Geográfico, que visa mostrar análises referentes à composição geográfica do concelho, é possível visualizar todas as análises criadas na plataforma *web* da ferramenta. A partir daí é possível selecionar uma delas e navegar no gráfico ou no mapa que é disponibilizado. Como exemplo, pode ver-se na Figura 51 um mapa

com os valores da população empregada por freguesia, sendo possível verificar quais as zonas mais povoadas e que mais se destacam nesta análise.

Apesar de se ter notado uma grande versatilidade desta ferramenta, quer no tipo de gráficos e análises disponíveis, como no próprio acesso e partilha de conteúdos num ambiente web, optou-se por manter o ficheiro Excel descrito no início desta secção. Isto porque, para além de servir de suporte à ferramenta *PowerBI*, permite realizar, como alternativa, operações mais complexas e diversificadas.

Para que melhor se entenda quais os aspetos da carta educativa que o protótipo desenvolvido consegue dar resposta, foi criada uma matriz, representada nas Tabela 41 e Tabela 42, que demonstra através de três níveis (Realizado, Parcialmente Realizado, Não Realizado) qual o estado dos indicadores alvo de análise.

Tabela 41 – Matriz de relação do estado dos indicadores (a)

Enquadramento Geográfico		
Indicador	Estado	Observações
Enquadramento Geográfico		-
Caracterização Física (Clima)		Falta de dados
Caracterização Económica		Impossível análise por ramo de atividade
Setores de Atividade		Implementadas análises mais detalhadas
Rede Viária		Falta de dados
Hierarquia dos Aglomerados		Falta de informação para definir a área urbana
Caracterização Demográfica		
População Residente		Implementadas análises mais detalhadas
Densidade Populacional		-
Estrutura Etária da População		-
Natalidade e Mortalidade		Implementadas análises mais detalhadas
Envelhecimento da população		-

Legenda:  - Realizado;  - Parcialmente Realizado;  - Não Realizado

Tabela 42 – Matriz de relação do estado dos indicadores (b)

Sistema Educativo		
Indicador	Estado	Observações
Distribuição da População		-
Abandono escolar, saída antecipada e precoce		Falta de dados
Sucesso Escolar		-
Classificação dos Exames Nacionais		Falta de dados
Proveniência dos Alunos Para a escola		-
Caracterização das IPSS's		Falta de dados
Distância média percorrida pelos alunos		Não implementado
Caracterização dos Agrupamentos de Escola		-
Caracterização das Escolas		-
Escolas com Excesso de Lotação		Faltam os dados referentes aos alunos dos outros ciclos de estudos
Evolução do número de alunos		-
Alunos por ciclo de estudo/Agrupamento/Escolas		-
Alunos com educação especial		Falta de dados
Alunos deslocados		Falta de dados
Distribuição pela oferta formativa		Falta de dados
Alunos subsidiados		Falta de dados
Corpo Docente		Falta de dados
Infraestruturas de ensino existentes		Apenas informação relativa aos pavilhões desportivos e refeitórios

Legenda:  - Realizado;  - Parcialmente Realizado;  - Não Realizado

Para se perceber como funcionam as aplicações *front-end* implementadas é possível visualizar a navegação das mesmas através do [vídeo](#) disponibilizado ([Vídeo de Demonstração](#)).

Para finalizar, é importante referir que mesmo estando a lidar com dados de características não tão sensíveis e a maioria ser até de acesso público, a segurança da plataforma de análise é assegurada através de contas de utilizador. Posto isto, apenas os responsáveis pela gestão da rede escolar do município de Guimarães terão acesso ao sistema implementado.

5. Conclusão

Neste capítulo são apresentadas as conclusões relativas ao trabalho desenvolvido para a elaboração desta dissertação de mestrado, desde a recolha bibliográfica até todo o desenvolvimento técnico da solução final, as dificuldades sentidas bem como as limitações do projeto. Por último, e tendo em vista a continuidade do trabalho, são também identificadas algumas propostas de desenvolvimento de trabalho futuro.

5.1. Síntese do Trabalho Realizado

Como objetivo principal, esta dissertação de mestrado teve como finalidade propor e implementar um protótipo de um sistema de *Business Intelligence*, que incluía como repositório um *Data Warehouse*, com a capacidade de suportar a tomada de decisão na gestão da rede escolar do Conselho de Guimarães por parte da Câmara Municipal.

Para que tal fosse possível, e fazendo um alinhamento com os objetivos propostos, numa primeira fase foi necessário realizar o trabalho de pesquisa e revisão da literatura dos conceitos de *Business Intelligence*, como forma de adquirir os conhecimentos necessários para o desenrolar do projeto. Para além da definição e dos conceitos de *Business Intelligence*, foram também descritas as componentes que fazem parte de um sistema deste tipo. Isso levou a que se aprofundassem também os conceitos associados aos sistemas de *Data Warehousing*, desde a sua definição, metodologia de desenvolvimento, arquiteturas e componentes dos sistemas deste tipo. Pelas necessidades particulares deste projeto, foi necessário realizar também uma revisão aos mecanismos de dados espaciais e aos conceitos associados que permitem o seu armazenamento e respetiva análise.

Como forma de garantir que a proposta aqui apresentada é inovadora, foi também incluída nesta revisão, uma pesquisa aos trabalhos já realizados na área do ensino. Isso levou à conclusão de que existem diversos trabalhos desenvolvidos nesta área, nomeadamente no estrangeiro, sendo que nenhum vai de encontro aos objetivos definidos nesta dissertação de mestrado.

Da revisão da literatura realizada, foi possível concretizar um dos objetivos definidos para este projeto, que é a elaboração de uma proposta de arquitetura do sistema que seja capaz de suportar o sistema de *Business Intelligence* desejado. Para tal, foi elaborado o enquadramento tecnológico que permitiu definir quais as ferramentas que satisfazem as necessidades de cada uma das etapas de construção de um sistema BI. Foi realizada uma descrição para cada uma

dessas ferramentas, bem como as suas funcionalidades e em que etapa do projeto se inserem. Nesta seleção foram consideradas também as necessidades da organização.

Os dados, para este tipo de sistema, são fundamentais e imprescindíveis. Como tal, e cumprindo com dois outros objetivos, foram identificadas e caracterizadas as fontes de dados necessárias para alimentar o sistema. Numa fase inicial, foram identificadas as fontes de dados provenientes dos *softwares* de gestão dos estabelecimentos de ensino. Contudo, após a devida análise e caracterização dos dados, esta hipótese teve de ser descartada por falta de qualidade dos mesmos. Para salvaguardar a execução e a viabilidade da solução apresentada, foi identificada uma outra fonte de dados que visa a gestão dos alunos do pré-escolar e do 1º ciclo de estudos. Estes dados são provenientes de um sistema OLTP do município de Guimarães. Foram apresentados alguns resultados da análise exploratória dos dados de ambas as fontes de maneira a justificar as decisões tomadas.

Depois de cumpridos os objetivos anteriores, foi possível iniciar o desenvolvimento propriamente dito do sistema. Numa fase inicial foi concebido o modelo de dados do *Data Warehouse* que servirá de base para o projeto. Este modelo baseia-se num esquema em constelação, constituído por seis estrelas (tabelas de factos) e 21 tabelas de dimensão. As tabelas de factos (FactSistemaEducativo, FactCarctClimatica e FactAnaliseNota) não foram carregadas por falta de dados. As restantes (FactAlunoEscola, FactEnqGeograficoDemografico e FactMortalidadeNatalidade) foram devidamente carregadas e encontram-se ligadas a diversas tabelas de dimensão que permitem análises de várias perspetivas.

Para por em prática o modelo de dados e concretizar o carregamento dos dados no *Data Warehouse*, desenvolveu-se um processo ETL com a capacidade de extrair os dados via *web service* do sistema OLTP e, por conseguinte, realizar diversas transformações para que, por fim, sejam carregados de forma automatizada no *Data Warehouse*.

Após a modelação dos dados e feito o respetivo carregamento no *Data Warehouse*, foi implementado um sistema de processamento analítico de cubos OLAP através da ferramenta *Microsoft SQL Analysis Services*. Foram implementados três cubos OLAP de acordo com as três tabelas de factos carregadas, sendo feito também o processamento das tabelas de dimensão, bem como definidas as suas hierarquias.

Por fim, e de maneira a cumprir com o último dos objetivos definidos, foram implementados mecanismos de acesso e análise da informação através de aplicações *front-end*. A partir de

ferramentas Microsoft, foram elaborados um conjunto de *dashboards* e relatórios que auxiliam os responsáveis da gestão da rede escolar na tomada de decisão. O sistema possibilita aos gestores o acesso *online* das análises, de forma dinâmica, através da ferramenta *Microsoft PowerBI*. É possível também criar novas análises recorrendo a gráficos, tabelas ou mapas. De maneira a facilitar ainda mais o acesso à informação, é possível visualizar as análises criadas através da aplicação móvel do *PowerBI*.

Para além de todo o trabalho aqui desenvolvido e descrito, importa também fazer referência ao trabalho realizado e que não está documentado. A par do trabalho desenvolvido, foram realizadas uma série de reuniões com a Câmara Municipal de Guimarães, com o agrupamento de escolas Virgínia Moura e com a empresa J.P.M & Abreu que fornece o *software* de gestão escolar para todos os agrupamentos de escolas do Município de Guimarães à exceção de um. Foi possível obter a partir dessas reuniões um *backup* dos dados de um agrupamento de escolas e uma versão do *software*, instalada numa máquina do Departamento de Sistemas de Informação da Universidade do Minho. Com a disponibilização deste *software* e do *backup*, foi possível ter uma perspetiva mais funcional do programa bem como realizar algumas análises à qualidade dos dados que foram disponibilizados.

5.2. Dificuldades e Limitações

Os sistemas de *Business Intelligence* apresentam-se como facilitadores no processo de tomada de decisão. Estes caracterizam-se como um processo de recolha, tratamento e análise dos dados organizacionais dos quais estes sistemas necessitam para “sobreviver”.

Apesar de vivermos numa era digital, ainda é possível encontrar sistemas OLTP com falhas graves ao nível estrutural com bases de dados bastante deficitárias e com níveis de homogeneidade bastantes reduzidos.

Esta foi sem dúvida a maior das dificuldades sentidas ao longo de toda o trabalho desenvolvido. As primeiras dificuldades surgiram logo no acesso às fontes de dados quando foram detetadas diversas debilidades no *software* de gestão de alunos dos estabelecimentos de ensino. O facto das bases de dados do *software* estarem descentralizadas, tornou desde logo impossível reunir toda a informação disponível. Como solução optou-se por adquirir um *backup* de apenas uma destas bases de dados referente ao agrupamento de escolas Virgínia de Moura do concelho de Guimarães.

Porém, para que a extração dos dados fosse possível foi necessário ter uma versão experimental do *software* que teve de ser acordada com a empresa responsável. Todo este processo de recolha e obtenção de todas as condições para obter os dados, acarretou atrasos significativos no desenrolar da dissertação.

Aliado a toda esta situação, ainda foi necessário lidar com a falta de qualidade dos dados que tornou a utilização dos mesmos impossível. Foi portanto necessário recorrer a outra fonte de dados para que o desenrolar do projeto não ficasse comprometido.

Pode-se concluir que as dificuldades sentidas influenciaram o desenrolar de todo o projeto. Contudo, tais dificuldades não inviabilizaram a concretização da prova de conceito necessária.

5.3. Trabalho Futuro

O trabalho desenvolvido e aqui apresentado, apesar de funcional e de todos os objetivos traçados inicialmente terem sido cumpridos com sucesso, apresenta aspetos que requerem uma evolução futura. Este protótipo do sistema requer várias melhorias e implementações futuras, por forma a ultrapassar algumas das limitações atualmente existentes. Destacam-se, portanto, um conjunto de implementações futuras que visam dar continuidade ao trabalho aqui iniciado:

- Numa primeira fase, seria pertinente rever todo o modelo de dados e incluir novas medidas (factos) que permitam análises complementares. Também seria importante implementar os mecanismos de frescamento dos dados, bem como criar *KPI*'s com metas precisas e implementar os atributos do tipo 2 da SCD (*Slowly Changing Dimension*).
- Como o sistema atual só incorpora os dados relativos aos alunos do pré-escolar e do 1º ciclo de estudos só é possível obter análises destes mesmos ciclos. Posto isto, seria também necessário alargar o sistema aos restantes ciclos de estudos, obtendo para tal todas as fontes de dados necessárias para o carregamento integral do *Data Warehouse*.
- Estender o modelo do *Data Warehouse* através da criação de novas tabelas de factos e de dimensão por forma a alargar as potencialidades destes sistemas obtendo novas perspetivas sobre o estado da rede escolar do concelho. Seria assim necessário proceder à reconstrução do respetivo processo ETL bem como dos cubos OLAP. Por

exemplo, seria importante perceber o que leva uma determinada região a ter um elevado número de alunos a reprovar, tentando para isso obter respostas através do cruzamento com um conjunto de dados sobre a criminalidade.

- Explorar técnicas de *Data Mining* para todas as áreas alvo de análise para que seja possível retirar mais informação e obter novas perspectivas que tragam valor à solução apresentada.

Referências Bibliográficas

- 1keydata. (2015). Data Warehouse Architecture. Retrieved February 17, 2015, from <http://www.1keydata.com/datawarehousing/data-warehouse-architecture.html>
- Abreu, J. P. M. &. (2004). J.P.M. & Abreu, Lda. Retrieved September 17, 2015, from <http://www.jpmafreu.com/cgi-bin/jpmcgi.jpma/pagina?id=empresa>
- Bédard, Y., Merrett, T., & Han, J. (2001). Fundamentals of spatial data warehousing for geographic knowledge discovery. In H. J. Miller & J. Han (Eds.), *Geographic data mining and knowledge discovery* (Second Edi, pp. 53–73). Taylor & Francis Group.
- Carrilho, M., & Patrício, L. (2010). A situação demográfica recente em Portugal. *Revista de Estudos Demográficos*, 48(5), 101–146.
- Castro, R. A. (2011). School Performance Evaluation in Portugal: A Data Warehouse Implementation to Automate Information Analysis. In *DSIE'11 - 6th Doctoral Symposium on Informatics Engineering* (pp. 1–12). Porto.
- Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 28(1), 65–74. <http://doi.org/10.1145/248603.248616>
- CMGuimarães. (2006). Carta Educativa do Concelho de Guimarães. Guimarães: Câmara Municipal de Guimarães.
- Dinis, R. (2011). *Padrões de Incidência Geográfica das Doenças Pulmonares Obstrutivas Crônicas e da Asma em Portugal*. Dissertação de Mestrado, Universidade do Minho.
- Educação, M. (2007). Educação e Formação em Portugal. Ministério da Educação.
- Educação, V. (2015). Elaboração de Cartas Educativas e Projeto Educativo Municipal. Retrieved February 12, 2015, from http://www.viaeducacao.pt/content/default.asp?idcat=cartas_Educativas&idCatM=cartas_Educativas&idContent=17011F91-D99E-485E-8D23-CD32C4E36C9E
- Elmasri, R., & Navathe, S. B. (2000). *Fundamentals of database systems* (Third Edit). Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA ©1999.
- Ferreira, I., Ferreira, S. G. G., Silva, C., & Carvalho, J. Á. (2012). Dilemas iniciais na investigação em TSI: design science e design research, uma clarificação de conceitos.
- Gartner. (2013a). Big Data Analytics. Retrieved February 17, 2015, from <http://www.gartner.com/it-glossary/big-data/>
- Gartner. (2013b). Business Intelligence - BI - Gartner IT Glossary. Retrieved December 17, 2014, from <http://www.gartner.com/it-glossary/business-intelligence-bi/>

- Gartner. (2014). Magic Quadrant for Business Intelligence and Analytics Platforms. Retrieved February 21, 2015, from <http://www.gartner.com/technology/reprints.do?id=1-1QLGACN&ct=140210&st=sb>
- GRT. (2015). Data Warehouse Architecture - A Study in Evolution. Retrieved February 17, 2015, from <http://www.grtcorp.com/content/data-warehouse-architecture-study-evolution>
- Hevner, A., & Chatterjee, S. (2010). *Design Research in Information Systems* (Vol. 22). Boston, MA: Springer US. <http://doi.org/10.1007/978-1-4419-5653-8>
- INE. (2013). *Censos 2011 - Preparação, Metodologia e Conceitos*. (I. N. de Estatística, Ed.) (First Edit). Lisboa: INE.
- Inmon, W. H. (2002). *Building the Data Warehouse Third Edition*. (R. Elliott, E. Herman, & J. Atkins, Eds.) *Wiley Publishing, Inc.* (Third Edit). Canada: Robert Ipsen.
- Jensen, C. S., & Pedersen, T. B. (2010). Multidimensional Databases. *Synthesis Lectures on Data Management*. Aalborg, Denmark.
- Kazi, L., Kazi, Z., & Radulovic, B. (2012). Data warehouse based evaluation of students' achievements in information systems education. In *MIPRO 2012 - 35th International Convention on Information and Communication Technology, Electronics and Microelectronics - Proceedings* (pp. 1377–1382).
- Kimball, R. (1997). A Dimensional Modeling Manifesto | Kimball Group. Retrieved February 9, 2015, from <http://www.kimballgroup.com/1997/08/a-dimensional-modeling-manifesto/>
- Kimball, R. (2013). Additive, Semi-Additive, and Non-Additive Facts | Kimball Group. Retrieved February 11, 2015, from <http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/additive-semi-additive-non-additive-fact/>
- Kurniawan, Y., & Halim, E. (2013). Use data warehouse and data mining to predict student academic performance in schools: A case study (perspective application and benefits). In *Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)* (pp. 98–103). IEEE. <http://doi.org/10.1109/TALE.2013.6654408>
- Loshin, D. (2012). *Business intelligence: the Savvy Manager's Guide*. (M. Kaufmann, Ed.) (2nd Editio). ELSEVIER.
- Miami, U. of. (2015). Data Marts | University of Miami. Retrieved February 3, 2015, from http://www.miami.edu/index.php/data_warehouse/data_marts/
- Mickael Santos, C. (2014). *Conceção e projeto de Data Warehouse: Olhar o Futuro do Ensino Básico e Secundário em Portugal*. Dissertação de Mestrado, Universidade de Aveiro.
- Microsoft. (2015). SQL Server Business Intelligence Edition | Microsoft. Retrieved February 15, 2015, from <http://www.microsoft.com/pt-pt/server-cloud/products/sql-server-editions/sql-server-business-intelligence.aspx>

- Moody, D. L., & Kortink, M. a R. (2003). From ER Models to Dimensional Models: Bridging the Gap between OLTP and OLAP Design, Part I. *Business Intelligence Journal*, 7–24.
- Morgado, A. (2013). *Data Warehouses Espaciais – Projeto e implementação*. Dissertação de Mestrado, Universidade do Minho.
- Negash, S. (2004). Business intelligence. *The Communications of the Association for Information*, 13(15), 177–195.
- Nobre, T., Trigo, A., & Sanches, P. (2014). SBIAES – Business intelligence system for analysis of access to higher education: The case of the Polytechnic Institute of Coimbra. In *2014 9th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1–6). IEEE. <http://doi.org/10.1109/CISTI.2014.6876866>
- Oliveira, B., Coragem, C., & Martins, É. (2000). *Manual para a Elaboração da Carta Educativa*. (M. da Educação, Ed.) (1st Editio). Lisboa: Ministério da Educação.
- Peppers, K., Tuunanen, T., Rothenberger, M. a., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77. <http://doi.org/10.2753/MIS0742-1222240302>
- Pennsylvania, U. of. (2015). University of Pennsylvania Data Warehouse. Retrieved February 3, 2015, from <http://www.upenn.edu/computing/da/dw/>
- Público, M. (2003). DL n.º 7/2003, de 15 de Janeiro. Retrieved December 5, 2014, from http://www.pgdlisboa.pt/leis/lei_mostra_articulado.php?nid=1620&tabela=leis
- Rainardi, V. (2008). *Building a data warehouse: with examples in SQL Server*. (J. Pepper, Ed.). Apress.
- Ralph Kimball, Reeves, L., Ross, M., & Thornthwaite, W. (1998). *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses*. (R. M. Elliott, Ed.) (1st Editio). Canada: Robert Ipsen.
- Ralph Kimball, & Ross, M. (2003). *The data warehouse toolkit: The Complete Guide to Dimensional Modeling*. (R. Elliott, E. Herman, J. Atkins, & B. Snapp, Eds.) (2nd Editio, Vol. 32). Canada: Robert Ipsen.
- Rivest, S., Bédard, Y., Proulx, M. J., & Nadeau, M. (2003). Solap: a New Type of User Interface To Support Spatio-Temporal Multidimensional Data Exploration and Analysis. *Knowledge Creation Diffusion Utilization*, 2–3.
- Sá, J. (2010). *Metodologia de sistemas de data warehouse*. Tese de Doutoramento, Universidade do Minho.
- Santos, M. Y., & Ramos, I. (2006). *Business Intelligence : tecnologias da informação na gestão de conhecimento* (1st Editio). FCA - Editora de Informática.
- Tableau. (2015). Tableau Desktop. Retrieved October 12, 2015, from <http://www.tableau.com/products/desktop>

- Talend. (2015a). Talend Data Integration & ETL Tools. Retrieved October 12, 2015, from <https://www.talend.com/products/data-integration>
- Talend. (2015b). Talend Data Quality. Retrieved October 12, 2015, from <https://www.talend.com/products/data-quality>
- Taniar, D. (2009). *Progressive Methods in Data Warehousing and Business Intelligence : Concepts and Competitive Analytics*. (X. Hu & W. Rahayu, Eds.)*Atlantic*. Monash University, Australia: Information Science Reference.
- Turban, E., Sharda, R., Aronson, J., & King, D. (2008). *Business intelligence: A managerial approach* (2nd Editio). Prentice Hall.
- Turban, E., Sharda, R., Delen, D., & Efraim, T. (2007). *Decision support and business intelligence systems* (8th Editio). USA: Prentice-Hall, Inc.
- Turkmen, G. (2007). *Developing a Data Warehouse for a University Decision Support System*. Atılım University, Tese de Mestrado.
- Wang, J. (2008). *Encyclopedia of Data Warehousing and Mining*. IGI Global. <http://doi.org/10.4018/978-1-60566-010-3>
- Watson, H. J., & Wixom, B. H. (2007). The Current State of Business Intelligence. *Computer*, 40(9), 96–99. <http://doi.org/10.1109/MC.2007.331>
- Zimányi, E., Gajda, E. M., & Approfondies, D. d'Etudes. (2003). *Concepts and Methodological Framework for Spatio-Temporal Data Warehouse Design*. Dissertação de Mestrado, Universite Libre de Bruxelles.

Anexos

Anexo A – Atributos das Fontes de Dados

A1. Fonte de Dados Software JPM & Abreu (AlunosGeraTotal.csv)

Tabela 43 - Atributos da fonte de dados JPM & Abreu (a)

Atributo	Descrição
N_Processo	Número do processo do aluno
IdMatricula	Identificador do número de matrícula do aluno
Nome	Nome do aluno
Nome Abreviado	Nome Abreviado do aluno
País_Origem	País de origem do aluno
Nacionalidade	Nacionalidade do aluno
sexo	Género sexual do aluno
Data_Nascimento	Data de nascimento do aluno
Lingua_Estrangeira_I	Língua estrangeira escolhida pelo aluno
Vacina	Data da última vacina
Morada1	Morada da residência do aluno
Morada2	Freguesia onde está sediada a residência
0000	Código postal de 4 dígitos
000	Código postal de 3 dígitos
CodigoPostal	Código postal
Telefone	Contacto telefónico do aluno
Email	Contacto de email do Aluno
N_BI	Número do bilhete de identidade ou cartão de cidadão
Arquivo_BI	Arquivo do bilhete de identidade ou cartão de cidadão
Data_Emissao_BI	Data de emissão do documento de identificação
Validade_BI	Data de validade do documento de identificação
ViveComPais	Indicação se o aluno vive com os pais
NecEnsiEspecial	Indicação se o aluno necessita de ensino especial
Subsidiado	Tipo de subsídio que o aluno usufrui
Escalaoaf	Grau do escalão que o aluno usufrui
Einiciativasase	Campo obsoleto
Ascendencia_Cultural	Referência à ascendência cultural do aluno
Deficiencia	Indicação se o aluno é portador de algum tipo de deficiência
TransporteUtilizado	Transporte utilizada na deslocação para a escola
Informações_Adicionais	Informação adicional relevante
Nome_Do_Pai	Nome do pai do aluno

Tabela 44 - Atributos da fonte de dados JPM & Abreu (b)

Atributo	Descrição
Nome_da_Mãe	Nome da mãe do aluno
ProfissaoPai	Profissão do pai
ProfissaoMae	Profissão da mãe
Telefone_de_Casa	Contacto telefónico da residência do aluno
Telefone_Emprego	Contacto telefónico do emprego do encarregado de educação
Situação_Sócio_Económica_Pai	Código referente à situação socioeconómica do pai
Situação_Sócio_Económica_Mãe	Código referente à situação socioeconómica da mãe
HabilitacoesPai	Habilitações literárias do pai
HabilitacoesMae	Habilitações literárias da mãe
Nome_Encarregado_Educação	Nome do encarregado de educação do aluno
Morada_1_Enc_Educação	Morada da residência do encarregado de educação
Morada_2_Enc_Educação	Freguesia onde está sediada a residência do encarregado de educação
CP_0000_Enc_Educação	Código postal de 4 dígitos
CP_000_Enc_Educação	Código postal de 3 dígitos
Código_Postal_Enc_Educação	Código postal
Profissão_Enc_Educação	Profissão do encarregado de educação
Telefone_Enc_Educação	Contacto telefónico do encarregado de educação do aluno
Email_Enc_Educação	Email do encarregado de educação do aluno
ParentescoED	Grau de parentesco do encarregado de educação
TipoCartao	Tipo de cartão utilizado na escola pelo aluno
Assistência_Médica	Número de assistência médica do sistema de saúde do aluno
Instituição	Instituição de saúde que o aluno está associado
Habilitações_Literárias_Enc_Educação	Habilitações do Encarregado de Educação
Situação_Sócio_Economica_Enc_Educação	Código referente à situação socioeconómica do encarregado de educação
Residência	Código referente ao local de residência do aluno
Passaporte	Código identificativo do passaporte do aluno
Cédula	Código identificativo da cédula de nascimento do aluno
Deslocado	Indicação se o aluno se encontra deslocado da sua região geográfica
NºElementosAgrFami	Número de elementos do agregado familiar

Tabela 45 - Atributos da fonte de dados JPM & Abreu (c)

Atributo	Descrição
Naturalidade_Freguesia_D_App	Código da freguesia do aluno
Computador	Indicação se o aluno possui computador
Internet	Indicação se o aluno possui acesso à internet
Naturalidade_Pai	Sigla do país do Pai
Naturalidade_Mãe	Sigla do país da Mãe
Naturalidade_Enc_Educação	Sigla do país do Encarregado de Educação
Nacionalidade_Pai	Sigla da nacionalidade do Pai
Nacionalidade_Mãe	Sigla da nacionalidade do mãe
Nacionalidade_Enc_Educação	Sigla da nacionalidade do encarregado de educação
Residência_Distrito	Código do Distrito do aluno
Residência_Concelho	Código do Concelho do aluno
Naturalidade_Concelho	Concelho do aluno
Naturalidade_Freguesia	Freguesia de residência do aluno
Curso	Curso que o aluno frequenta
N_Inscrição	Número de inscrição do aluno num curso
Texto_Instituição	Nome da instituição de saúde que o aluno está associado
NISS	Número de identificação da Segurança Social do aluno
NIF	Número de identificação fiscal do aluno
NCGA	Número de identificação da caixa geral de aposentações
Abandono	Indicação se o aluno abandonou a escola
AbandonoData	Data em que o aluno abandonou a escola
Despacho_50	Plano de recuperação aplicado ao aluno ao abrigo do Despacho Normativo n° 50/2005
PercursoCurricularAlternativo	Indicação se o aluno teve um percurso curricular alternativo
Texto_freguesia	Freguesia do aluno
Texto_Profissão_Pai	Profissão do pai
Texto_Profissão_Mãe	Profissão da mãe
Texto_habilitações_Pai	Habilitações literárias do pai
Texto_habilitações_Mãe	Habilitações literárias da mãe
Texto_Profissão_Enc_Educação	Profissão do encarregado de educação
Texto_habilitações_Enc_Educação	Habilitações literárias do encarregado de educação
Texto_Parentesco	Grau de parentesco do encarregado de educação
Texto_Naturalidade	País de onde o aluno é natural

Tabela 46 - Atributos da fonte de dados JPM & Abreu (d)

Atributo	Descrição
Texto_Distrito	Distrito de onde o aluno é natural
Texto_Concelho	Concelho de onde o aluno é natural
Computador_1	Indica se o aluno tem computador
Internet_1	Indica se o aluno tem internet
Texto_Nacionalidade_pai	Nacionalidade do pai
Texto_Nacionalidade_mae	Nacionalidade da mãe
Texto_Naturalidade_pai	Naturalidade do pai
Texto_Naturalidade_mae	Naturalidade da mãe

A2. Fonte de Dados Plataforma SIGA - Edubox

Tabela 47 - Atributos da fonte de dados Edubox (a)

Atributo	Descrição
BIAluno	Número do documento de identificação do aluno
BIEncEdu	Número do documento de identificação do encarregado de educação
BIMae	Número do documento de identificação da mãe
BIPai	Número do documento de identificação do pai
CodPostalAluno	Código postal (4 dígitos) do aluno
CodPostalEncEdu	Código postal (4 dígitos) do encarregado de educação
CodPostalMae	Código postal (4 dígitos) da mãe
CodPostalPai	Código postal (4 dígitos) do pai
DataNascimento	Data de nascimento do aluno
EmailAluno	Endereço de email do aluno
EmailEncEdu	Endereço de email do encarregado de educação
EmailMae	Endereço de email da mãe
EmailPai	Endereço de email do pai
EscalaoAluno	Escalão de apoio social escolar do aluno
ExtensaoCodPostalAluno	Código postal (3 dígitos) do aluno
ExtensaoCodPostalEncEdu	Código postal (3 dígitos) do encarregado de educação
ExtensaoCodigoPostalMae	Código postal (3 dígitos) da mãe
ExtensaoCodigoPostalPai	Código postal (3 dígitos) do pai
GrauParentescoEncEdu	Grau de parentesco do encarregado de educação em relação ao aluno
LocalidadeAluno	Localidade do aluno
LocalidadeMae	Localidade da mãe
LocalidadePai	Localidade do pai
Localidade	Localidade do encarregado de educação
MoradaEncEdu	Morada do encarregado de educação

Tabela 48 - Atributos da fonte de dados Edubox (b)

Atributo	Descrição
MoradaMae	Morada da mãe
MoradaPai	Morada do pai
NifAluno	Número de identificação fiscal do aluno
NifEncEdu	Número de identificação fiscal do encarregado de educação
NifMae	Número de identificação fiscal da mãe
NifPai	Número de identificação fiscal do pai
NomeAluno	Nome do aluno
NomeAnoEscolaridade	Ano de escolaridade que o aluno frequenta
NomeEncEdu	Nome do encarregado de educação
NomeEscola	Nome da escola que o aluno frequenta
NomeMae	Nome da mãe
NomePai	Nome do pai
NomeTurma	Indicação da turma que aluno frequenta
TelefoneAluno	Número de telefone do aluno
TelefoneEncEdu	Número de telefone do encarregado de educação
TelefoneMae	Número de telefone da mãe
TelefonePai	Número de telefone do pai
moradaAluno	Morada do aluno
nomeAgrupamento	Nome do agrupamento de escolas que o aluno frequenta
sexo	Indicação do sexo do aluno
InitialDate	Data de entrada do aluno numa turma
EndDate	Data de saída do aluno da turma
IdUser	Identificador único de um determinado aluno

Anexo B – Scripts

B1. Criação da Tabela DimTempo

```
USE [DW_CartaEducativa]
GO
/***** Object: StoredProcedure [dbo].[dim_tempo_sp]    Script Date: 08/09/2015 18:09:31 *****/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO

ALTER PROCEDURE [dbo].[dim_tempo_sp]
AS
BEGIN
    -- SET NOCOUNT ON added to prevent extra result sets from
    -- interfering with SELECT statements.
    SET NOCOUNT ON;

    -- Insert statements for procedure here

    INSERT INTO DW_CartaEducativa.dbo.DimTempo ([Data],[DiaDaSemana],[DiaDoMes],[Semana],[Mes],[Trimestre],[Semestre],[Ano])
    VALUES ('2000/01/01', 'Desconhecido', 0, 0, 0, 0, 0, 0)

    DECLARE @OrderDate DATETIME
    DECLARE @OrderFim DATETIME
    DECLARE @Semestre int
    SET @OrderDate = '2000/01/01'
    SET @OrderFim = GETDATE()

    WHILE @Orderdate <= @Orderfim
    BEGIN
        if DATEPART(mm,@OrderDate) < 7
            SET @Semestre =1
        else
            SET @Semestre =2

        INSERT INTO DW_CartaEducativa.dbo.DimTempo ([Data],[DiaDaSemana],[DiaDoMes],[Semana],[Mes],[Trimestre],[Semestre],[Ano])
        VALUES (@OrderDate, DATENAME(dw,@OrderDate), DATEPART(dd,@OrderDate), DATEPART(wk,@OrderDate), DATEPART(mm,@OrderDate), DATEPART(qq,@OrderDate),
        @Semestre,DATEPART(yy,@OrderDate))
        SET @OrderDate = @OrderDate + 1
    END
END
```

Figura 52 - Script de Criação da tabela DimTempo

B2. Script de Criação Base de Dados de Estágio

```
"
USE DadosCartaEducativa;

IF Object_ID('Aluno') is NOT null DROP TABLE Aluno

CREATE TABLE Aluno (
    IDAluno int IDENTITY (1,1) ,
    BIALuno nvarchar(255) ,
    BIEncEdu nvarchar(255) ,
    BIMae nvarchar(255) ,
    BIPai nvarchar(255) ,
    CodPostalAluno nvarchar(255) ,
    CodPostalEncEdu nvarchar(255) ,
    CodPostalMae nvarchar(255) ,
    CodPostalPai nvarchar(255) ,
    DataNascimento nvarchar(255) ,
    EmailAluno nvarchar(255) ,
    EmailEncEdu nvarchar(255) ,
    EmailMae nvarchar(255) ,
    EmailPai nvarchar(255) ,
    EscalaoAluno nvarchar(255) ,
    ExtensaoCodPostalAluno nvarchar(255) ,
    ExtensaoCodPostalEncEdu nvarchar(255) ,
    ExtensaoCodigoPostalMae nvarchar(255) ,
    ExtensaoCodigoPostalPai nvarchar(255) ,
    GrauParentescoEncEdu nvarchar(255) ,
    Localidade nvarchar(255) ,
    LocalidadeAluno nvarchar(255) ,
    LocalidadeMae nvarchar(255) ,
    LocalidadePai nvarchar(255) ,
    MoradaEncEdu nvarchar(255) ,
    MoradaMae nvarchar(255) ,
    MoradaPai nvarchar(255) ,
    NifAluno nvarchar(255) ,
    NifEncEdu nvarchar(255) ,
    NifMae nvarchar(255) ,
    NifPai nvarchar(255) ,
    NomeAluno nvarchar(255) ,
    NomeAnoEscolaridade nvarchar(255) ,
    NomeEncEdu nvarchar(255) ,
    NomeEscola nvarchar(255) ,
    NomeMae nvarchar(255) ,
    NomePai nvarchar(255) ,
    NomeTurma nvarchar(255) ,
    TelefoneAluno nvarchar(255) ,
    TelefoneEncEdu nvarchar(255) ,
    TelefoneMae nvarchar(255) ,
    TelefonePai nvarchar(255) ,
    moradaAluno nvarchar(255) ,
    nomeAgrupamento nvarchar(255) ,
    sexo nvarchar(255) ,
    initialDate nvarchar(255) ,
    endDate nvarchar(255) ,
    idUser int,
    PRIMARY KEY (IDAluno));"
```

B3. Script de Criação do *Data Warehouse*

```
"
USE DW_CartaEducativa;

IF Object_ID('FactEnqGeograficoDemografico') is NOT null DROP TABLE
FactEnqGeograficoDemografico
IF Object_ID('FactSistemaEducativo') is NOT null DROP TABLE
factSistemaEducativo
IF Object_ID('FactCaractClimatica') is NOT null DROP TABLE
FactCaractClimatica
IF Object_ID('FactAnaliseNota') is NOT null DROP TABLE
FactAnaliseNota
IF Object_ID('FactMortalidadeNatalidade') is NOT null DROP TABLE
FactMortalidadeNatalidade
IF Object_ID('FactAlunoEscola') is NOT null DROP TABLE
FactAlunoEscola

IF Object_ID('DimBridgeProf') is NOT null DROP TABLE DimBridgeProf
IF Object_ID('DimBridgeDisciplina') is NOT null DROP TABLE
DimBridgeDisciplina
IF Object_ID('DimBridgeCurso') is NOT null DROP TABLE DimBridgeCurso
IF Object_ID('DimBridgeNaoDocente') is NOT null DROP TABLE
DimBridgeNaoDocente

IF Object_ID('DimTempo') is NOT null DROP TABLE DimTempo
IF Object_ID('DimLocalizacao') is NOT null DROP TABLE DimLocalizacao
IF Object_ID('DimAluno') is NOT null DROP TABLE DimAluno
IF Object_ID('DimDisciplina') is NOT null DROP TABLE DimDisciplina
IF Object_ID('DimCurso') is NOT null DROP TABLE DimCurso
IF Object_ID('DimEscola') is NOT null DROP TABLE DimEscola
IF Object_ID('DimAgrupamento') is NOT null DROP TABLE DimAgrupamento
IF Object_ID('DimTempoAno') is NOT null DROP TABLE DimTempoAno
IF Object_ID('DimSituacao') is NOT null DROP TABLE DimSituacao
IF Object_ID('DimPeriodo') is NOT null DROP TABLE DimPeriodo
IF Object_ID('DimNota') is NOT null DROP TABLE DimNota
IF Object_ID('DimNivelEnsino') is NOT null DROP TABLE DimNivelEnsino
IF Object_ID('DimProfessor') is NOT null DROP TABLE DimProfessor
IF Object_ID('DimNaoDocente') is NOT null DROP TABLE DimNaoDocente
IF Object_ID('DimEscalao') is NOT null DROP TABLE DimEscalao
IF Object_ID('DimAnoEscolar') is NOT null DROP TABLE DimAnoEscolar
IF Object_ID('DimTurma') is NOT null DROP TABLE DimTurma

CREATE TABLE DimTempo (
    ID_Tempo      int IDENTITY (1,1),
    Data          date ,
    DiaDaSemana  nvarchar(20) ,
    DiaDoMes     int ,
    Semana       int ,
    Mes          int ,
    Trimestre    int ,
    Semestre     int ,
    Ano          int ,
    PRIMARY KEY (ID_Tempo));

CREATE TABLE DimTempoAno (
    ID_TempoAno  int IDENTITY (1,1),
    Ano          int ,
    PRIMARY KEY (ID_TempoAno));

CREATE TABLE DimLocalizacao (
    IDLocalizacao int IDENTITY (1,1) ,
    CodDistrito   int,
    CodLocalidade int,
    CodFreguesia  int,
    Localidade    nvarchar(50) ,
    Distrito      nvarchar(50),
    Concelho      nvarchar(50),
    Area          nvarchar(50),
```

```

Nivel                                nvarchar(50),
CP4                                  int,
CP3                                  nvarchar(3),
DesignacaoPostal                     nvarchar(255),
Latitude                             nvarchar(20),
Longitude                             nvarchar(20),
PRIMARY KEY (IDLocalizacao));

CREATE TABLE DimAluno (
  IDAluno          int IDENTITY (1,1) ,
  Sexo             nvarchar(255),
  DataNascimento  Date,
  NIF              int,
  ParentescoEncEdu nvarchar(100),
  OrigemID_Aluno  int ,
  PRIMARY KEY (IDAluno));

CREATE TABLE DimDisciplina (
  IDDisciplina    int IDENTITY (1,1) ,
  Sigla           nvarchar(10),
  Designacao      nvarchar(50),
  Codigo          nvarchar(50),
  Ciclo           nvarchar(50),
  PRIMARY KEY (IDDisciplina));

CREATE TABLE DimCurso (
  IDCurso         int IDENTITY (1,1) ,
  Sigla           nvarchar(100),
  Codigo          nvarchar(100),
  Designacao      nvarchar(255),
  Ciclo           nvarchar(50),
  PRIMARY KEY (IDCurso));

CREATE TABLE DimEscola (
  IDEscola        int IDENTITY (1,1) ,
  Nome             nvarchar(255),
  Tipo            nvarchar(100),
  Lotacao         nvarchar(100),
  NumSalasAula    nvarchar(100),
  PavilhaoDesportivo nvarchar(100),
  Refeitório      nvarchar(100),
  NumDocentes     nvarchar(100),
  NumNaoDocentes nvarchar(100),
  AnoAbertura     int,
  AnoUltimaRemodelacao int,
  NumRemodelacoes int,
  CP4             nvarchar(4),
  CP3             nvarchar(3),
  Freguesia       nvarchar(255),
  PRIMARY KEY (IDEscola));

CREATE TABLE DimAgrupamento (
  IDAgrupamento  int IDENTITY (1,1) ,
  Nome            nvarchar(255),
  EscolaSede      nvarchar(255),
  NumEscolas      int,
  AnoCriacao      int,
  PRIMARY KEY (IDAgrupamento));

CREATE TABLE DimSituacao (
  IDSituacao      int IDENTITY (1,1),
  DesigncaoSituacao varchar(50),
  Sigla           varchar(3),
  PRIMARY KEY (IDSituacao));

CREATE TABLE DimPeriodo (
  IDPeriodo       int IDENTITY (1,1),
  Designcao       varchar(50) ,
  PRIMARY KEY (IDPeriodo));

CREATE TABLE DimNota (
  IDNota          int IDENTITY (1,1),

```

```

Designcao          varchar(50) ,
PRIMARY KEY (IDNota));

CREATE TABLE DimNivelEnsino (
  IDNivelEnsino    int IDENTITY (1,1),
  Designcao        varchar(50) ,
PRIMARY KEY (IDNivelEnsino));

CREATE TABLE DimEscalao (
  IDEscalao       int IDENTITY (1,1),
  Escalao         varchar(50) ,
PRIMARY KEY (IDEscalao));

CREATE TABLE DimAnoEscolar (
  IDAnoEscolar    int IDENTITY (1,1),
  AnoEscolar      varchar(50) ,
PRIMARY KEY (IDAnoEscolar));

CREATE TABLE DimTurma (
  IDTurma         int IDENTITY (1,1),
  Turma           varchar(50) ,
PRIMARY KEY (IDTurma));

CREATE TABLE DimProfessor (
  IDProfessor     int IDENTITY (1,1),
  Identificacao   varchar(50) ,
PRIMARY KEY (IDProfessor));

CREATE TABLE DimNaoDocente (
  IDNaoDocente    int IDENTITY (1,1),
  Identificacao   varchar(50) ,
  Funcao          varchar(50),
PRIMARY KEY (IDNaoDocente));

CREATE TABLE DimBridgeProf (
  idGrupoProf    int IDENTITY (1,1),
  DimProfessor_idProfessor INT NOT NULL,
PRIMARY KEY (idGrupoProf),
INDEX fk_DimGrupoBridge_DimProfessor1_idx
(DimProfessor_idProfessor ASC),

CONSTRAINT fk_DimGrupoBridge_DimProfessor1
FOREIGN KEY (DimProfessor_idProfessor)
REFERENCES DimProfessor (IDProfessor)
ON DELETE NO ACTION
ON UPDATE NO ACTION);

CREATE TABLE DimBridgeDisciplina (
  idGrupoDisciplina int IDENTITY (1,1),
  DimDisciplina_idDisciplina INT NOT NULL,
PRIMARY KEY (idGrupoDisciplina),
INDEX fk_DimBridgeDisciplina_DimDisciplinal_idx
(DimDisciplina_idDisciplina ASC),

CONSTRAINT fk_DimBridgeDisciplina_DimDisciplinal
FOREIGN KEY (DimDisciplina_idDisciplina)
REFERENCES DimDisciplina (IDDisciplina)
ON DELETE NO ACTION
ON UPDATE NO ACTION);

CREATE TABLE DimBridgeCurso (
  idGrupoCurso   int IDENTITY (1,1),
  DimCursos_idCurso INT NOT NULL,
PRIMARY KEY (idGrupoCurso),
INDEX fk_DimBridgeCurso_DimCursos1_idx (DimCursos_idCurso ASC),

CONSTRAINT fk_DimBridgeCurso_DimCursos1

```

```

FOREIGN KEY (DimCursos_idCurso)
REFERENCES DimCurso (IDCurso)
ON DELETE NO ACTION
ON UPDATE NO ACTION);

CREATE TABLE DimBridgeNaoDocente (
    idGrupoNaoDocente int IDENTITY (1,1),
    DimNaoDocente_idNaoDocente INT NOT NULL,
    PRIMARY KEY (idGrupoNaoDocente),
    INDEX fk_DimBridgeNaoDocente_DimNaoDocente1_idx
(DimNaoDocente_idNaoDocente ASC),

    CONSTRAINT fk_DimBridgeNaoDocente_DimNaoDocente1
    FOREIGN KEY (DimNaoDocente_idNaoDocente)
    REFERENCES DimNaoDocente (IDNaoDocente)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION);

CREATE TABLE FactSistemaEducativo (
    Contador int NOT NULL,
    DimLocalizacao_idLocalizacao INT NOT NULL,
    DimTempo_idTempo INT NOT NULL,
    DimAgrupamento_idAgrupamento INT NOT NULL,
    DimEscola_idEscola INT NOT NULL,
    DimBridgeProf_idGrupoProf INT NOT NULL,
    DimBridgeDisciplina_idGrupoDisciplina INT NOT NULL,
    DimBridgeCurso_idGrupoCurso INT NOT NULL,
    DimBridgeNaoDocente_idGrupoNaoDocente INT NOT NULL,
    PRIMARY KEY (DimLocalizacao_idLocalizacao, DimTempo_idTempo,
DimAgrupamento_idAgrupamento, DimEscola_idEscola,
DimBridgeProf_idGrupoProf, DimBridgeDisciplina_idGrupoDisciplina,
DimBridgeCurso_idGrupoCurso, DimBridgeNaoDocente_idGrupoNaoDocente),
    INDEX fk_FactSistemaEducativo_DimTempo1_idx (DimTempo_idTempo
ASC),
    INDEX fk_FactSistemaEducativo_DimAgrupamento1_idx
(DimAgrupamento_idAgrupamento ASC),
    INDEX fk_FactSistemaEducativo_DimEscola1_idx (DimEscola_idEscola
ASC),
    INDEX fk_FactSistemaEducativo_DimBridgeProf1_idx
(DimBridgeProf_idGrupoProf ASC),
    INDEX fk_FactSistemaEducativo_DimBridgeDisciplinal_idx
(DimBridgeDisciplina_idGrupoDisciplina ASC),
    INDEX fk_FactSistemaEducativo_DimBridgeCurso1_idx
(DimBridgeCurso_idGrupoCurso ASC),
    INDEX fk_FactSistemaEducativo_DimBridgeNaoDocente1_idx
(DimBridgeNaoDocente_idGrupoNaoDocente ASC),
    CONSTRAINT fk_FactSistemaEducativo_DimLocalizacao1
    FOREIGN KEY (DimLocalizacao_idLocalizacao)
    REFERENCES DimLocalizacao (idLocalizacao)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
    CONSTRAINT fk_FactSistemaEducativo_DimTempo1
    FOREIGN KEY (DimTempo_idTempo)
    REFERENCES DimTempo (ID_Tempo)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
    CONSTRAINT fk_FactSistemaEducativo_DimAgrupamento1
    FOREIGN KEY (DimAgrupamento_idAgrupamento)
    REFERENCES DimAgrupamento (IdAgrupamento)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
    CONSTRAINT fk_FactSistemaEducativo_DimEscola1
    FOREIGN KEY (DimEscola_idEscola)
    REFERENCES DimEscola (idEscola)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
    CONSTRAINT fk_FactSistemaEducativo_DimBridgeProf1
    FOREIGN KEY (DimBridgeProf_idGrupoProf)
    REFERENCES DimBridgeProf (idGrupoProf)

```

```

        ON DELETE NO ACTION
        ON UPDATE NO ACTION,
    CONSTRAINT fk_FactSistemaEducativo_DimBridgeDisciplinal
        FOREIGN KEY (DimBridgeDisciplina_idGrupoDisciplina)
        REFERENCES DimBridgeDisciplina (idGrupoDisciplina)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION,
    CONSTRAINT fk_FactSistemaEducativo_DimBridgeCurso1
        FOREIGN KEY (DimBridgeCurso_idGrupoCurso)
        REFERENCES DimBridgeCurso (idGrupoCurso)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION,
    CONSTRAINT fk_FactSistemaEducativo_DimBridgeNaoDocente1
        FOREIGN KEY (DimBridgeNaoDocente_idGrupoNaoDocente)
        REFERENCES DimBridgeNaoDocente (idGrupoNaoDocente)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION);

CREATE TABLE FactAlunoEscola (
    Contador INT NOT NULL,
    FlagAcaoSocial INT NOT NULL,
    FlagMatriculado INT NOT NULL,
    DimAluno_idAluno INT NOT NULL,
    DimEscola_idEscola INT NOT NULL,
    DimAgrupamento_idAgrupamento INT NOT NULL,
    DimTempo_idTempo INT NOT NULL,
    DimLocalizacao_idLocalizacao INT NOT NULL,
    DimSituacao_idSituacao INT NOT NULL,
    DimEscalao_IDEscalao INT NOT NULL,
    DimAnoEscolar_IDAnoEscolar INT NOT NULL,
    DimTurma_IDTurma INT NOT NULL,
    InitialDate_idTempo INT NOT NULL,
    PRIMARY KEY (DimAluno_idAluno, DimEscola_idEscola,
    DimAgrupamento_idAgrupamento, DimTempo_idTempo,
    DimLocalizacao_idLocalizacao, DimSituacao_idSituacao,
    DimEscalao_IDEscalao, DimAnoEscolar_IDAnoEscolar, DimTurma_IDTurma,
    InitialDate_idTempo),
    INDEX fk_FactAlunoEscola_DimEscola1_idx (DimEscola_idEscola ASC),
    INDEX fk_FactAlunoEscola_DimAgrupamento1_idx
    (DimAgrupamento_idAgrupamento ASC),
    INDEX fk_FactAlunoEscola_DimTempo1_idx (DimTempo_idTempo ASC),
    INDEX fk_FactAlunoEscola_DimLocalizacao1_idx
    (DimLocalizacao_idLocalizacao ASC),
    INDEX fk_FactAlunoEscola_DimSituacao1_idx (DimSituacao_idSituacao
    ASC),
    INDEX fk_FactAlunoEscola_DimEscalao1_idx (DimEscalao_IDEscalao
    ASC),
    INDEX fk_FactAlunoEscola_DimAnoEscolar1_idx
    (DimAnoEscolar_IDAnoEscolar ASC),
    INDEX fk_FactAlunoEscola_DimTurma1_idx (DimTurma_IDTurma ASC),
    INDEX fk_FactAlunoEscola_DimTempo2_idx (InitialDate_idTempo ASC),
    CONSTRAINT fk_FactAlunoEscola_DimAluno1
        FOREIGN KEY (DimAluno_IDAluno)
        REFERENCES DimAluno (IDAluno)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION,
    CONSTRAINT fk_FactAlunoEscola_DimEscola1
        FOREIGN KEY (DimEscola_IDEscola)
        REFERENCES DimEscola (IDEscola)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION,
    CONSTRAINT fk_FactAlunoEscola_DimAgrupamento1
        FOREIGN KEY (DimAgrupamento_IDAgrupamento)
        REFERENCES DimAgrupamento (IDAgrupamento)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION,
    CONSTRAINT fk_FactAlunoEscola_DimTempo1
        FOREIGN KEY (DimTempo_IDTempo)
        REFERENCES DimTempo (ID_Tempo)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION,

```

```

CONSTRAINT fk_FactAlunoEscola_DimLocalizacao1
  FOREIGN KEY (DimLocalizacao_idLocalizacao)
  REFERENCES DimLocalizacao (IDLocalizacao)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
CONSTRAINT fk_FactAlunoEscola_DimSituacao1
  FOREIGN KEY (DimSituacao_IDSituacao)
  REFERENCES DimSituacao (IDSituacao)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
CONSTRAINT fk_FactAlunoEscola_DimEscalao1
  FOREIGN KEY (DimEscalao_IDEscalao)
  REFERENCES DimEscalao (IDEscalao)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
CONSTRAINT fk_FactAlunoEscola_DimAnoEscolar1
  FOREIGN KEY (DimAnoEscolar_IDAnoEscolar)
  REFERENCES DimAnoEscolar (IDAnoEscolar)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
CONSTRAINT fk_FactAlunoEscola_DimTurma1
  FOREIGN KEY (DimTurma_IDTurma)
  REFERENCES DimTurma (IDTurma)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
CONSTRAINT fk_FactAlunoEscola_DimTempo2
  FOREIGN KEY (InitialDate_idTempo)
  REFERENCES DimTempo (ID_Tempo)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION);

```

```

CREATE TABLE FactEnqGeograficoDemografico (
  PopulacaoAtiva INT NOT NULL,
  PopulacaoSetorPrimario INT NOT NULL,
  PopulacaoSetorSecundario INT NOT NULL,
  PopulacaoSetorTerciario INT NOT NULL,
  PopulacaoEmpregada INT NOT NULL,
  PopulacaoReformada INT NOT NULL,
  PopulacaoSemAtividade INT NOT NULL,
  PopulacaoDesempregada INT NOT NULL,
  NumeroHabitantes INT NOT NULL,
  NumeroHabiFeminino INT NOT NULL,
  NumeroHabiMasculino INT NOT NULL,
  NumeroIndiviResidentes0a4 INT NOT NULL,
  NumeroIndiviResidentes0a4M INT NOT NULL,
  NumeroIndiviResidentes0a4F INT NOT NULL,
  NumeroIndiviResidentes5a9 INT NOT NULL,
  NumeroIndiviResidentes5a9M INT NOT NULL,
  NumeroIndiviResidentes5a9F INT NOT NULL,
  NumeroIndiviResidentes10a13 INT NOT NULL,
  NumeroIndiviResidentes10a13M INT NOT NULL,
  NumeroIndiviResidentes10a13F INT NOT NULL,
  NumeroIndiviResidentes14a19 INT NOT NULL,
  NumeroIndiviResidentes14a19M INT NOT NULL,
  NumeroIndiviResidentes14a19F INT NOT NULL,
  NumeroIndiviResidentes20a25 INT NOT NULL,
  NumeroIndiviResidentes20a25M INT NOT NULL,
  NumeroIndiviResidentes20a25F INT NOT NULL,
  NumeroIndiviResidentes25a64 INT NOT NULL,
  NumeroIndiviResidentes25a64M INT NOT NULL,
  NumeroIndiviResidentes25a64F INT NOT NULL,
  NumeroIndiviResidentesMais65 INT NOT NULL,
  NumeroIndiviResidentesMais65M INT NOT NULL,
  NumeroIndiviResidentesMais65F INT NOT NULL,
  NumeroIndiviEnsinoCompleto1Basico INT NOT NULL,
  NumeroIndiviEnsinoCompleto2Basico INT NOT NULL,
  NumeroIndiviEnsinoCompleto3Basico INT NOT NULL,
  NumeroIndiviEnsinoCompletoSec INT NOT NULL,
  NumeroIndiviEnsinoCompletoPosSec INT NOT NULL,
  NumeroIndiviEnsinoCompletoSuperior INT NOT NULL,

```

```

NumeroIndiviEnsinoCompleto1BasicoF INT NOT NULL,
NumeroIndiviEnsinoCompleto2BasicoF INT NOT NULL,
NumeroIndiviEnsinoCompleto3BasicoF INT NOT NULL,
NumeroIndiviEnsinoCompletoSecF INT NOT NULL,
NumeroIndiviEnsinoCompletoPosSecF INT NOT NULL,
NumeroIndiviEnsinoCompletoSupF INT NOT NULL,
NumeroIndiviNaoLerEscrever INT NOT NULL,
DimLocalizacao_idLocalizacao INT NOT NULL,
DimTempoAno_idTempoAno INT NOT NULL,
PRIMARY KEY (DimLocalizacao_idLocalizacao,
DimTempoAno_idTempoAno),
INDEX fk_FactEnqGeograficoDemografico_DimTempoAno1_idx
(DimTempoAno_idTempoAno ASC),
CONSTRAINT fk_FactEnqGeografico_DimLocalizacao1
FOREIGN KEY (DimLocalizacao_idLocalizacao)
REFERENCES DimLocalizacao (IDLocalizacao)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT fk_FactEnqGeograficoDemografico_DimTempoAno1
FOREIGN KEY (DimTempoAno_idTempoAno)
REFERENCES DimTempoAno (ID_TempoAno)
ON DELETE NO ACTION
ON UPDATE NO ACTION);

CREATE TABLE FactCaractClimatica (
TemperaturaTotal INT NOT NULL,
PercipitacaoTotal INT NOT NULL,
HumidadeTotal INT NOT NULL,
NeblusidadeTotal INT NOT NULL,
DimTempo_idTempo INT NOT NULL,
DimLocalizacao_idLocalizacao INT NOT NULL,
PRIMARY KEY (DimTempo_idTempo, DimLocalizacao_idLocalizacao),
INDEX fk_FactCaractClimatica_DimLocalizacao1_idx
(DimLocalizacao_idLocalizacao ASC),
CONSTRAINT fk_FactCaractClimatica_DimTempo1
FOREIGN KEY (DimTempo_idTempo)
REFERENCES DimTempo (ID_Tempo)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT fk_FactCaractClimatica_DimLocalizacao1
FOREIGN KEY (DimLocalizacao_idLocalizacao)
REFERENCES DimLocalizacao (IDLocalizacao)
ON DELETE NO ACTION
ON UPDATE NO ACTION);

CREATE TABLE FactMortalidadeNatalidade (
NumeroNascimentos INT NOT NULL,
NumeroMortes INT NOT NULL,
DimLocalizacao_idLocalizacao INT NOT NULL,
DimTempoAno_idTempoAno INT NOT NULL,
PRIMARY KEY (DimLocalizacao_idLocalizacao,
DimTempoAno_idTempoAno),
INDEX fk_FactMortalidadeNatalidade_DimLocalizacao1_idx
(DimLocalizacao_idLocalizacao ASC),
INDEX fk_FactMortalidadeNatalidade_DimTempoAno1_idx
(DimTempoAno_idTempoAno ASC),
CONSTRAINT fk_FactMortalidadeNatalidade_DimLocalizacao1
FOREIGN KEY (DimLocalizacao_idLocalizacao)
REFERENCES DimLocalizacao (IDLocalizacao)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT fk_FactMortalidadeNatalidade_DimTempoAno1
FOREIGN KEY (DimTempoAno_idTempoAno)
REFERENCES DimTempoAno (ID_TempoAno)
ON DELETE NO ACTION
ON UPDATE NO ACTION);

CREATE TABLE FactAnaliseNota (
Contador INT NOT NULL,
NotaQualitativaEscala INT NOT NULL,
Nota0a20 DECIMAL NOT NULL,

```



```

Nota0a5 INT NOT NULL,
DimNotas_idNota INT NOT NULL,
DimLocalizacao_IDLocalizacao INT NOT NULL,
DimAluno_idAluno INT NOT NULL,
DimDisciplina_idDisciplina INT NOT NULL,
DimNivelEnsino_idNivelEnsino INT NOT NULL,
DimPeriodo_idPeriodo INT NOT NULL,
DimTempoAno_idTempoAno INT NOT NULL,
PRIMARY KEY (DimNotas_idNota, DimLocalizacao_IDLocalizacao,
DimAluno_idAluno, DimDisciplina_idDisciplina,
DimNivelEnsino_idNivelEnsino, DimPeriodo_idPeriodo,
DimTempoAno_idTempoAno),
INDEX fk_FactAnaliseNotas_DimLocalizacao1_idx
(DimLocalizacao_IDLocalizacao ASC),
INDEX fk_FactAnaliseNotas_DimAluno1_idx (DimAluno_idAluno ASC),
INDEX fk_FactAnaliseNotas_DimDisciplinal1_idx
(DimDisciplina_idDisciplina ASC),
INDEX fk_FactAnaliseNotas_DimNivelEnsino1_idx
(DimNivelEnsino_idNivelEnsino ASC),
INDEX fk_FactAnaliseNotas_DimPeriodo1_idx (DimPeriodo_idPeriodo
ASC),
INDEX fk_FactAnaliseNotas_DimTempoAno1_idx (DimTempoAno_idTempoAno
ASC),
CONSTRAINT fk_FactAnaliseNotas_DimNotas1
FOREIGN KEY (DimNotas_idNota)
REFERENCES DimNota (idNota)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT fk_FactAnaliseNotas_DimLocalizacao1
FOREIGN KEY (DimLocalizacao_IDLocalizacao)
REFERENCES DimLocalizacao (IDLocalizacao)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT fk_FactAnaliseNotas_DimAluno1
FOREIGN KEY (DimAluno_idAluno)
REFERENCES DimAluno (IdAluno)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT fk_FactAnaliseNotas_DimDisciplinal1
FOREIGN KEY (DimDisciplina_idDisciplina)
REFERENCES DimDisciplina (idDisciplina)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT fk_FactAnaliseNotas_DimNivelEnsino1
FOREIGN KEY (DimNivelEnsino_idNivelEnsino)
REFERENCES DimNivelEnsino (idNivelEnsino)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT fk_FactAnaliseNotas_DimPeriodo1
FOREIGN KEY (DimPeriodo_idPeriodo)
REFERENCES DimPeriodo (idPeriodo)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT fk_FactAnaliseNotas_DimTempoAno1
FOREIGN KEY (DimTempoAno_idTempoAno)
REFERENCES DimTempoAno (Id_TempoAno)
ON DELETE NO ACTION
ON UPDATE NO ACTION);

```

..

B4. Script Java Converter em Coordenadas Geográficas

```
package coordenadas;

import java.io.File;
import java.io.FileOutputStream;
import java.io.IOException;
import java.net.MalformedURLException;
import java.net.URL;
import java.nio.channels.Channels;
import java.nio.channels.ReadableByteChannel;
import java.sql.*;
import javax.xml.parsers.DocumentBuilder;
import javax.xml.parsers.DocumentBuilderFactory;
import javax.xml.parsers.ParserConfigurationException;
import javax.xml.transform.TransformerException;
import org.w3c.dom.DOMException;
import org.w3c.dom.Element;
import org.w3c.dom.Node;
import org.w3c.dom.NodeList;
import org.xml.sax.SAXException;

/**
 *
 * @author TiagoCC
 */
public class Coordenadas {

    public static void main(String[] args) throws ClassNotFoundException,
        SQLException, IOException, SAXException, ParserConfigurationException,
        TransformerException {

        //Estabelecer a conexão à BD SQL SERVER
        Class.forName("com.microsoft.sqlserver.jdbc.SQLServerDriver");
        String connectionURL =
"jdbc:sqlserver://localhost:1433;databaseName=Dim_Local;user=TiagoCordenada
s;password=13591220";
        Connection con = DriverManager.getConnection(connectionURL);

        //Ir buscar os dados
        Statement st = con.createStatement();
        ResultSet rs = st.executeQuery("SELECT TOP 2200 FORMAT (CP3,
'000'),CP4,Localidade from DimLocalizacao where latitude = -1");

        while (rs.next()) {
            int CP4 = rs.getInt(2);
            String CP3 = rs.getString(1);
            String Localidade = rs.getString(3);
            System.out.println("Código Postal: " + CP4 + "-" + CP3 + "-" +
Localidade.replace(" ", "%"));
            //conectar ao google maps
            try {

                // Minha URL Gmap = new
                URL("https://maps.googleapis.com/maps/api/geocode/xml?address=" + CP4 + "+"
+ CP3 + "+" + Localidade.replace(" ", "&") + "&key=AIzaSyC3vyj4d-
YqIYrGL5Fb27iF_csBVze97SQ");
                // Sara URL Gmap = new
                URL("https://maps.googleapis.com/maps/api/geocode/xml?address="+CP4+"+"+CP3
+"+"+Localidade.replace("
", "&")+ "&key=AIzaSyCTz50nX6_lAobJdGwVt007hThxZVrasI4");
                URL Gmap = new
                URL("https://maps.googleapis.com/maps/api/geocode/xml?address="+CP4+"+"+CP3
+"+"+Localidade.replace("
", "&")+ "&key=AIzaSyCCiOmrPKZ4_AS9swy6dEhCDzPbRN2R9Kk");
                ReadableByteChannel rbc =
                Channels.newChannel(Gmap.openStream());
                new File("C://GoogleMapsInfo").mkdir();
                FileOutputStream fos = new
                FileOutputStream("C://GoogleMapsInfo/GoogleMapsInfo.xml");
                fos.getChannel().transferFrom(rbc, 0, Long.MAX_VALUE);
            }
        }
    }
}
```

```

        try {
            //Gravar um ficheiro XML com a informação que vem do
            google maps (ficheiro temporário)
            File fXmlFile = new
            File("C://GoogleMapsInfo/GoogleMapsInfo.xml");
            DocumentBuilderFactory dbFactory =
            DocumentBuilderFactory.newInstance();
            DocumentBuilder dBuilder =
            dbFactory.newDocumentBuilder();
            org.w3c.dom.Document doc = dBuilder.parse(fXmlFile);
            doc.getDocumentElement().normalize();
            NodeList nList = doc.getElementsByTagName("result");
            System.out.println("-----");

            for (int temp = 0; temp < nList.getLength(); temp++) {
                Node nNode = nList.item(temp);
                if (nNode.getNodeType() == Node.ELEMENT_NODE) {
                    Element eElement = (Element) nNode;
                    System.out.println("Endereço = " +
                    eElement.getElementsByTagName("formatted_address").item(0).getTextContent()
                    );
                    System.out.println("Latitude : " +
                    eElement.getElementsByTagName("lat").item(0).getTextContent());
                    System.out.println("Longitude : " +
                    eElement.getElementsByTagName("lng").item(0).getTextContent());
                    String latitude1 =
                    eElement.getElementsByTagName("lat").item(0).getTextContent();
                    String longitude1 =
                    eElement.getElementsByTagName("lng").item(0).getTextContent();
                    Float F1 = Float.valueOf(latitude1);
                    Float F2 = Float.valueOf(longitude1);
                    Statement st1 = con.createStatement();

                    if (F1 >= 40 && F1 <= 42 && F2 <= -7.7 && F2 >=
                    -9) {
                        st1.executeQuery("update DimLocalizacao Set
                        Latitude = " + F1 + ", Longitude = " + F2 + " where CP4 =" + CP4 + "and
                        CP3= " + CP3);
                    } else {
                        System.out.println("NÃO ESTÁ BEM");
                    }
                }
            }
        } catch (ParserConfigurationException | SAXException |
        IOException | DOMException | NumberFormatException | SQLException e) {
        }

        } catch (MalformedURLException e) {
        // new URL() failed
        // ...
        } catch (IOException e) {
        // openConnection() failed
        // ...
        }
    }
}

```

Anexo C – Dashboards das Aplicações Front-End

C1. Vista Geral da Dashboard Caracterização Demográfica

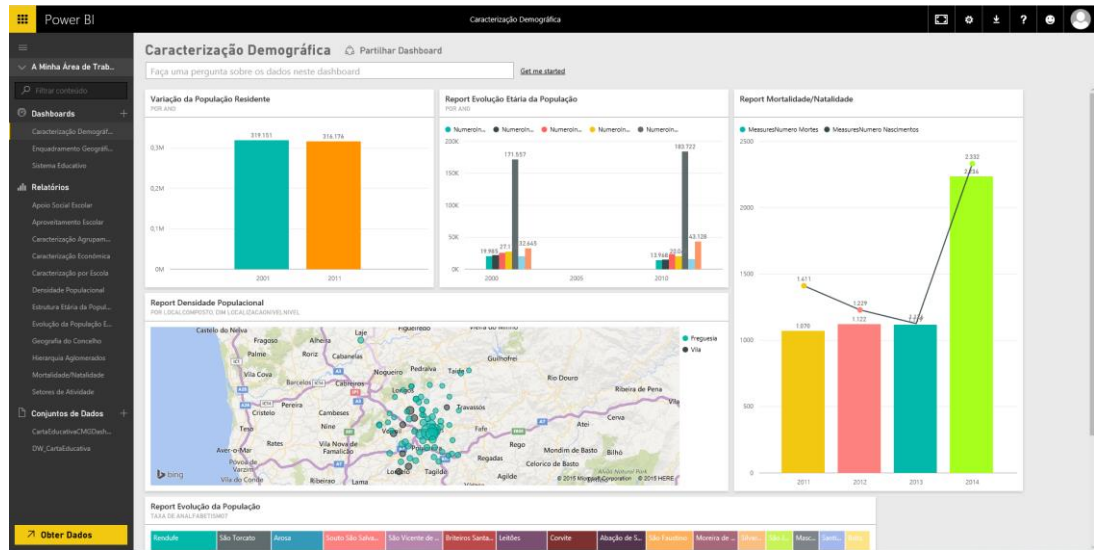


Figura 53 - Dashboard Caracterização Demográfica (PowerBI)

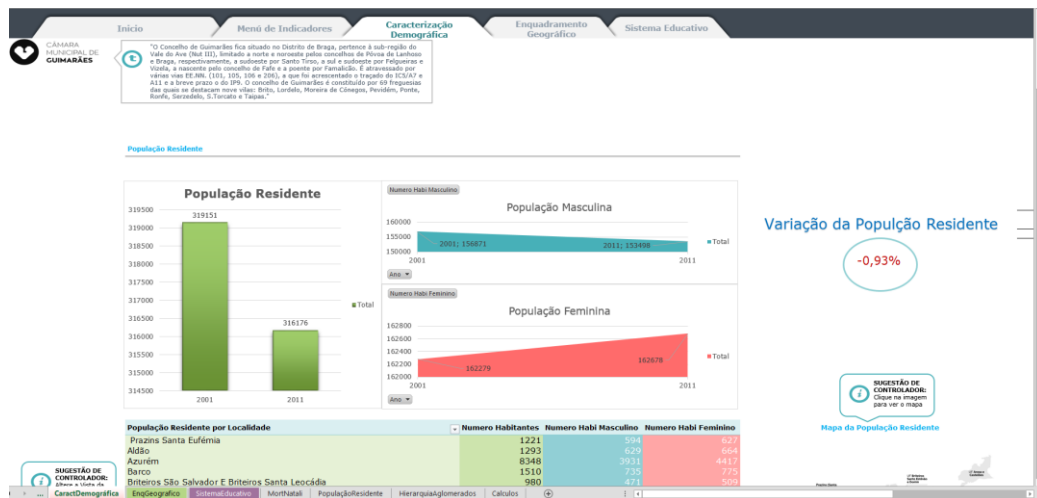


Figura 54 - Dashboard Caracterização Demográfica 1 (Excel 2013)

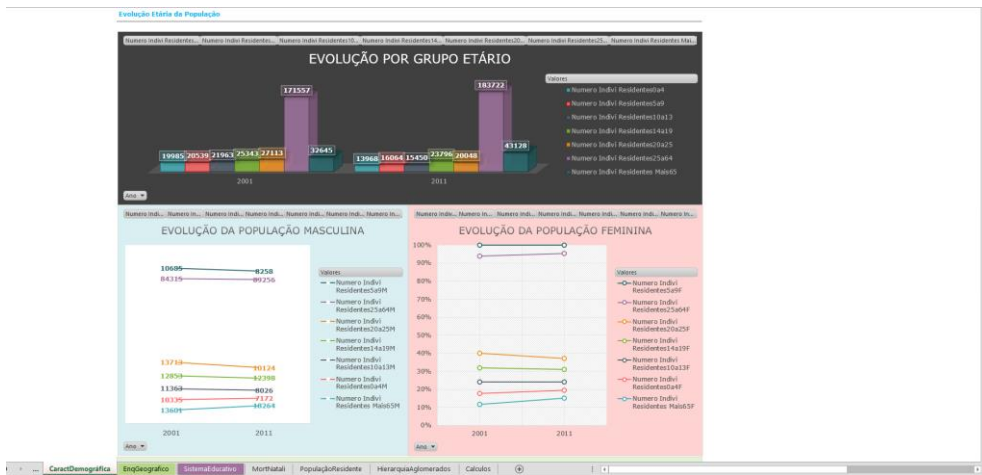


Figura 55 - Dashboard Caracterização Demográfica 2 (Excel 2013)

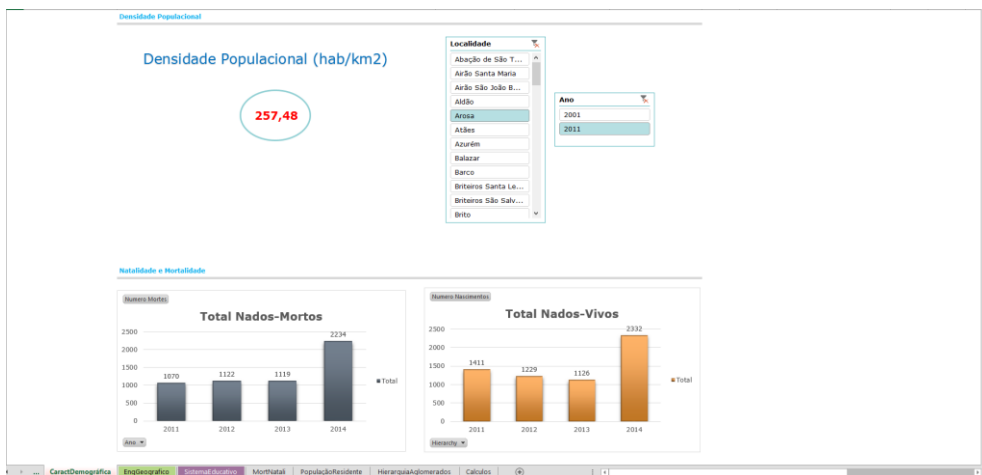


Figura 56 - Dashboard Caracterização Demográfica 3 (Excel 2013)

C2. Vista Geral da Dashboard Enquadramento Geográfico

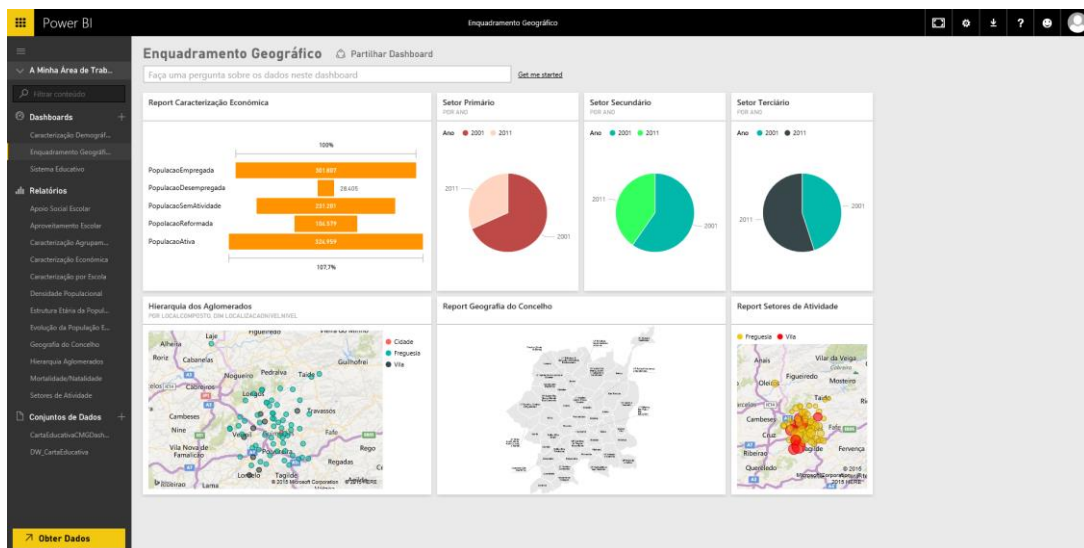


Figura 57 - Dashboard Enquadramento Geográfico

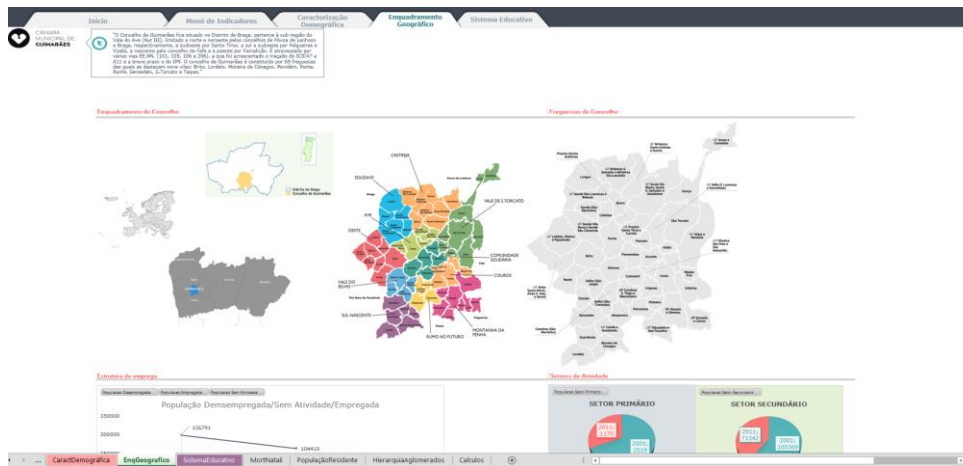


Figura 58 - Dashboard Enquadramento Geográfico 1 (Excel 2013)

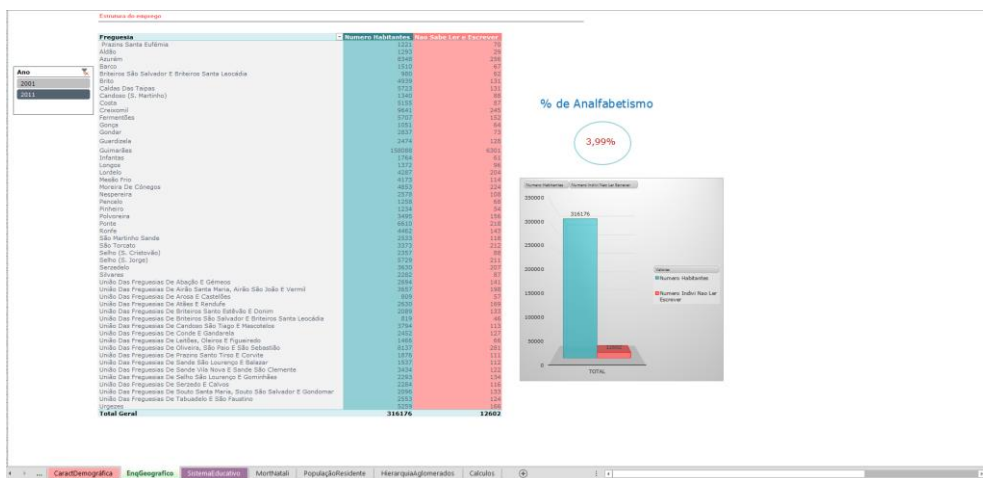


Figura 59 - Dashboard Enquadramento Geográfico 2 (Excel 2013)

C3. Vista Geral da Dashboard Sistema Educativo

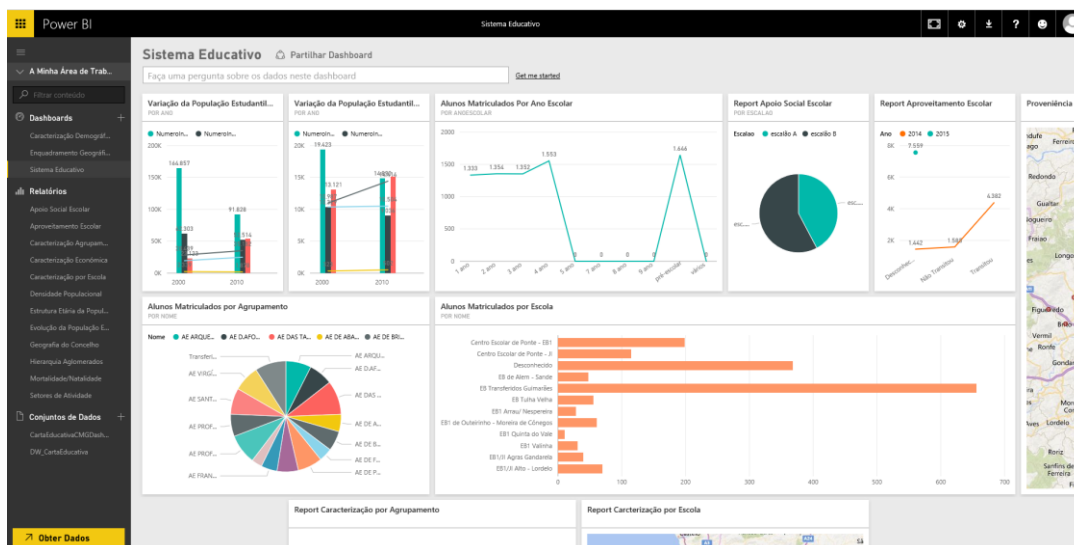


Figura 60 - Dashboard Sistema Educativo



Figura 61 - Dashboard Sistema Educativo 1 (Excel 2013)

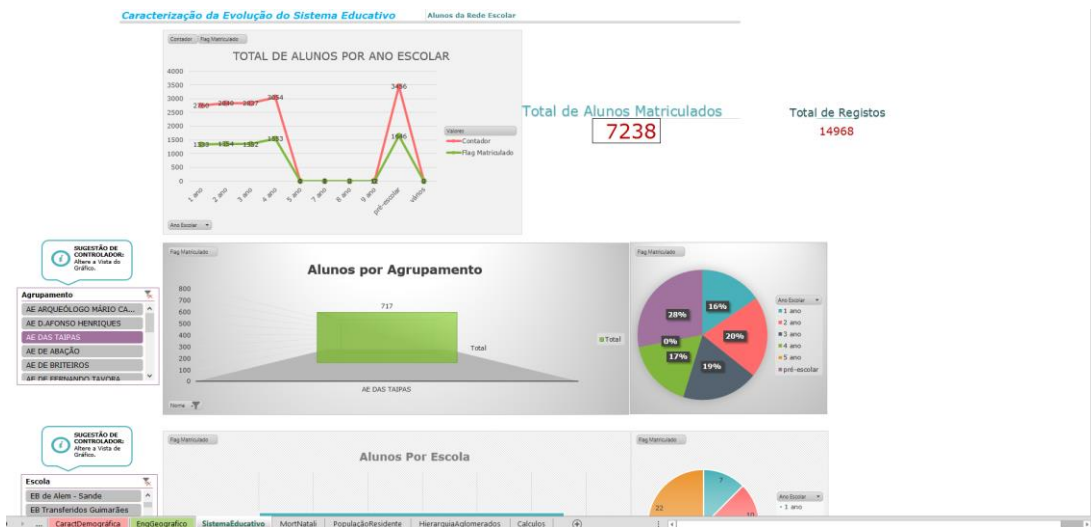


Figura 62 - Dashboard Sistema Educativo 2 (Excel 2013)

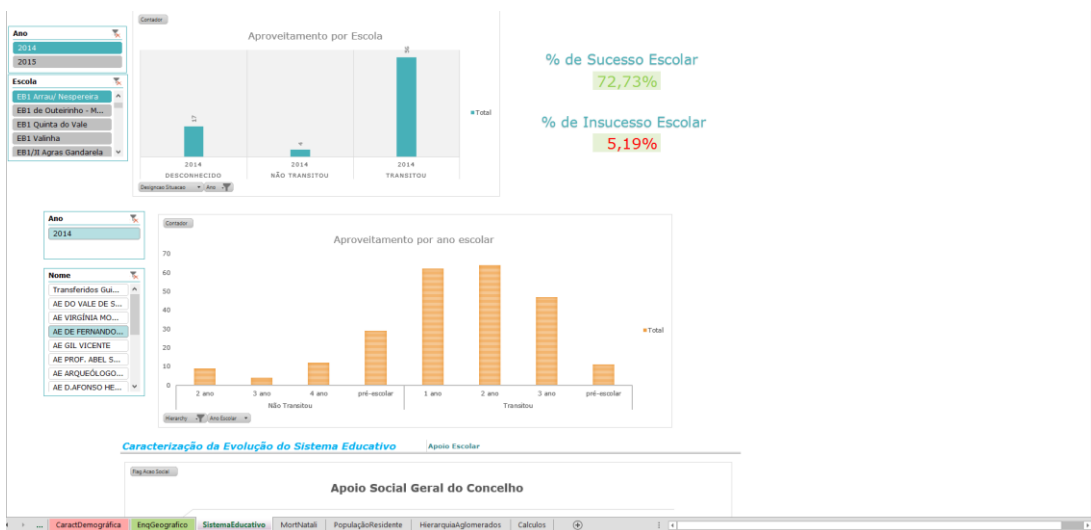


Figura 63 - Dashboard Sistema Educativo 3 (Excel 2013)

Anexo D – Relatórios nas Aplicações *Front-End*

D1. Vista Geral do Relatório Apoio Social Escolar



Figura 64 - Relatório Apoio Social Escolar

D2. Vista Geral do Relatório Aproveitamento Escolar

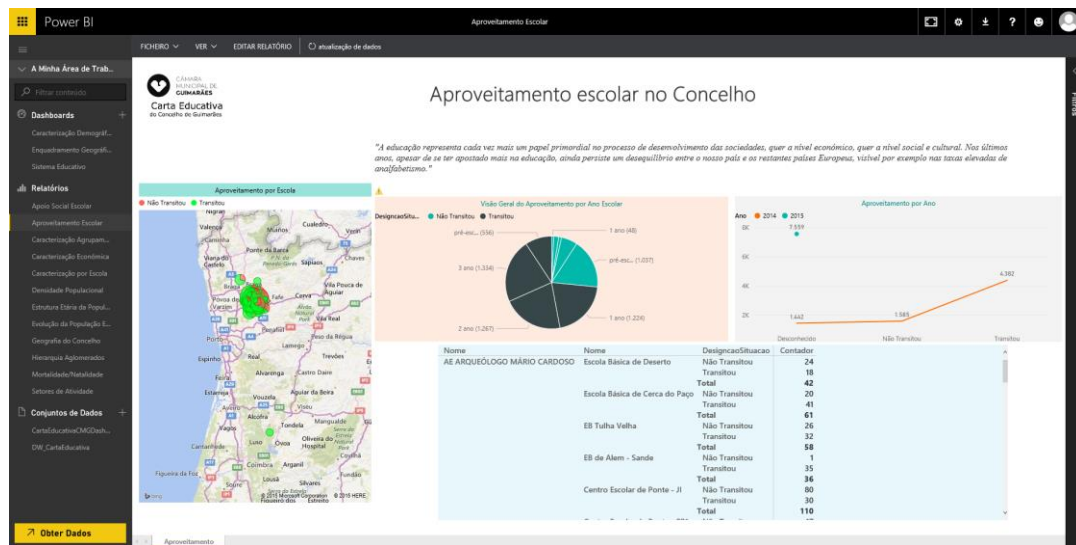


Figura 65 - Relatório Aproveitamento Escolar

D3. Vista Geral do Relatório Caracterização por Agrupamento

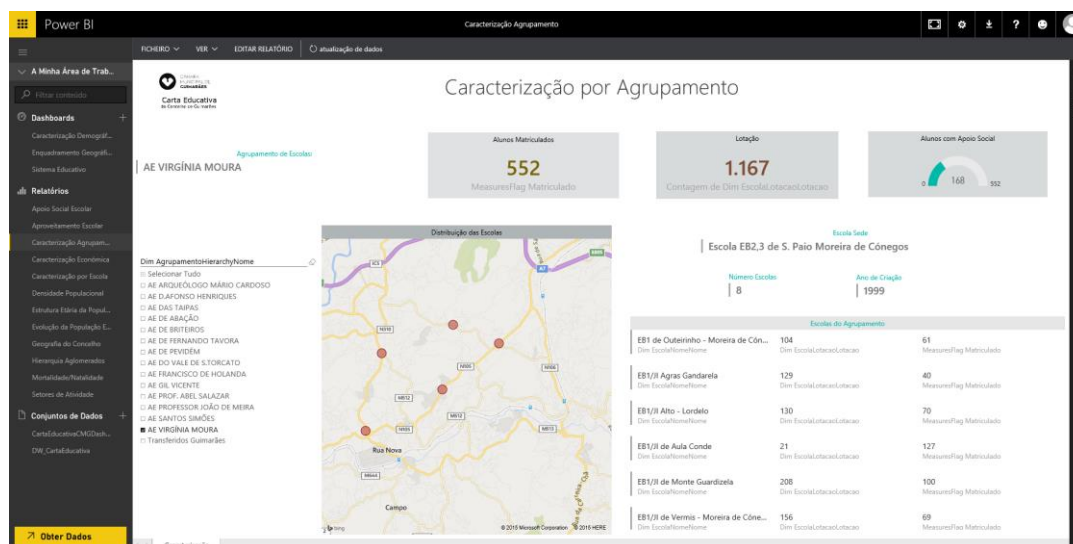


Figura 66 - Relatório Caracterização por Agrupamento

D4. Vista Geral do Relatório Caracterização Económica

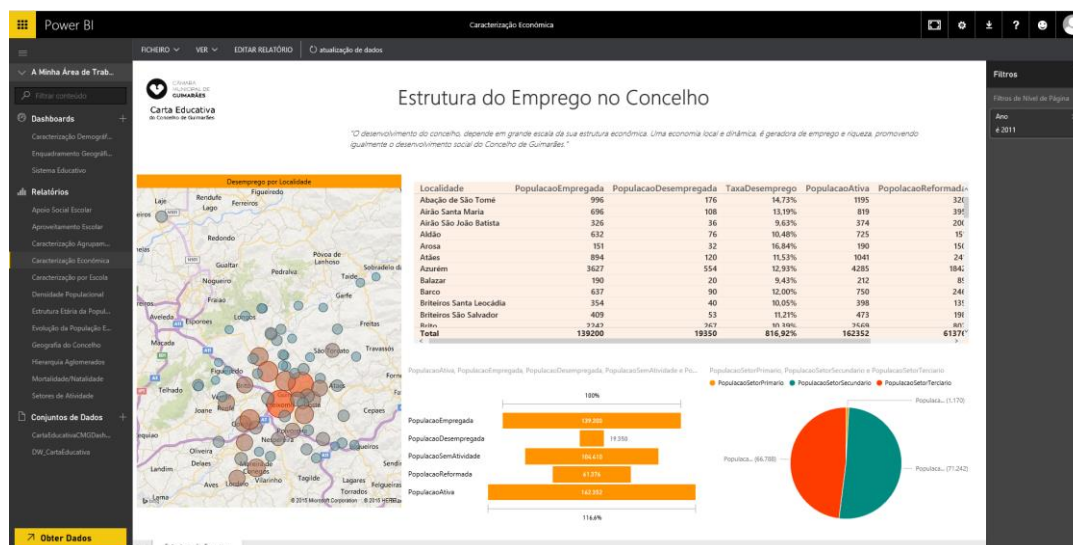


Figura 67 - Relatório Caracterização Económica

D5. Vista Geral do Relatório Caracterização por Escola

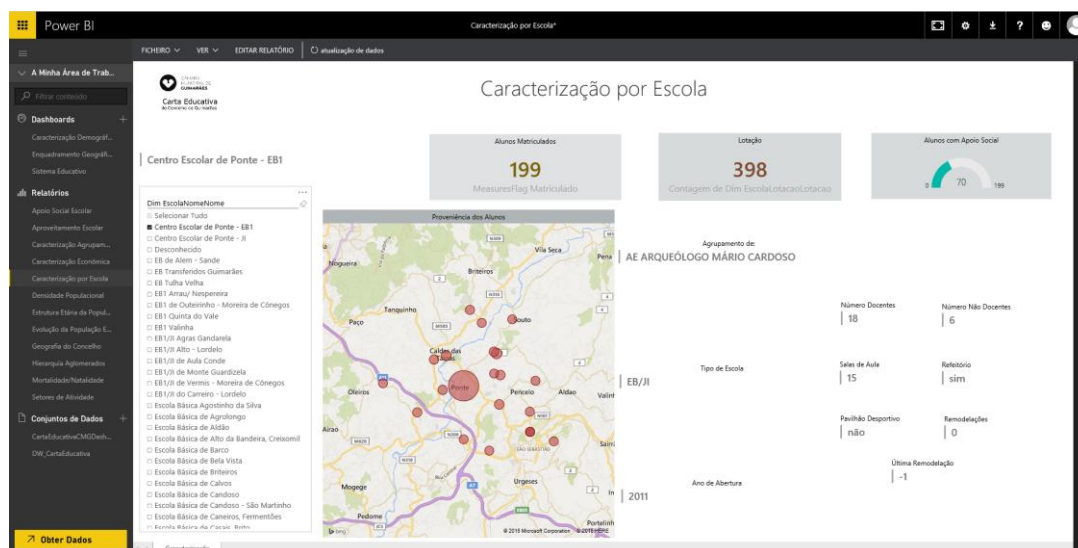


Figura 68 - Caracterização por Escola

D6. Vista Geral do Relatório Densidade Populacional

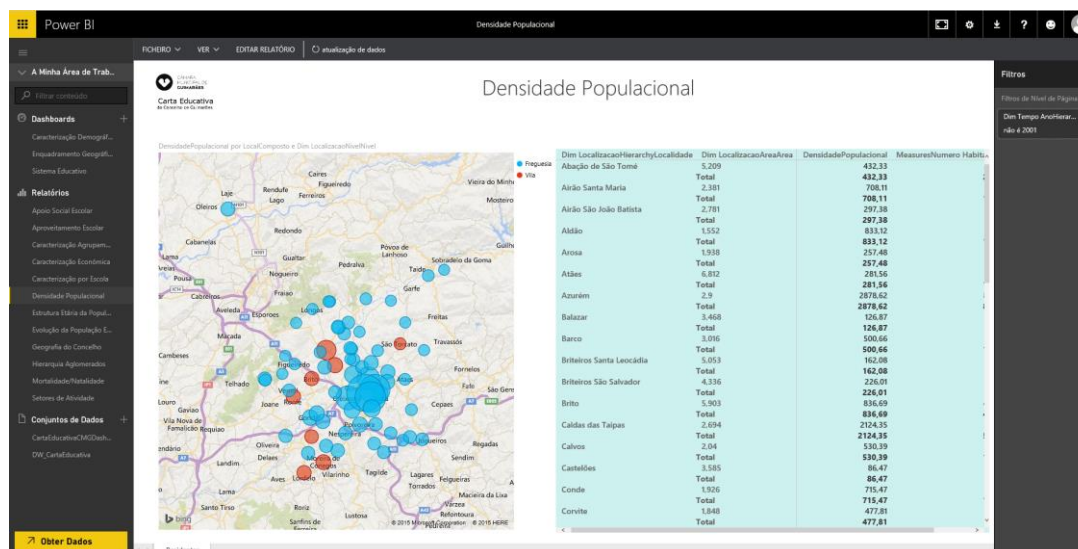


Figura 69 - Relatório Densidade Populacional

D7. Vista Geral do Relatório Evolução Etária da População

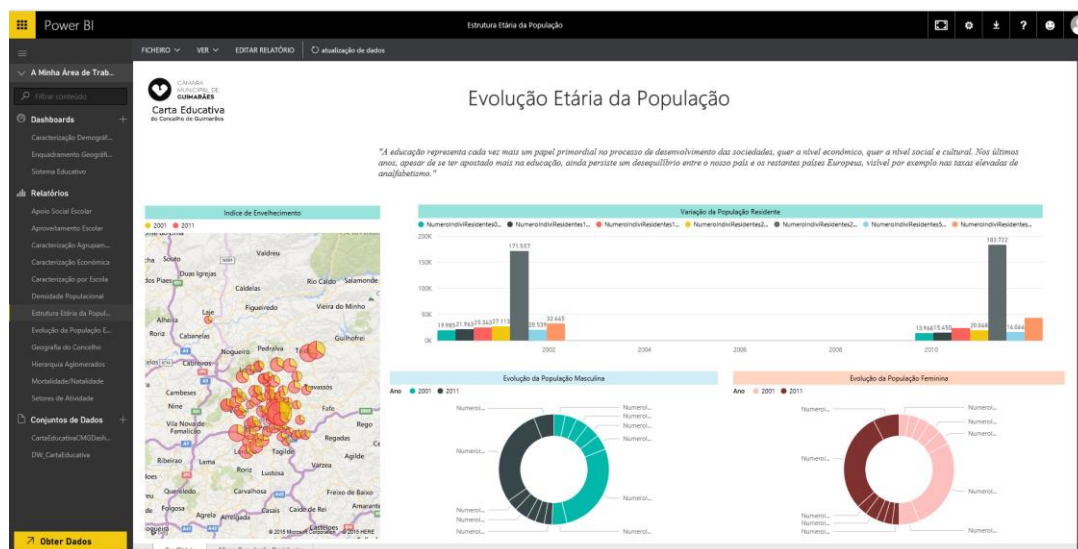


Figura 70 - Relatório Evolução Etária da População 1

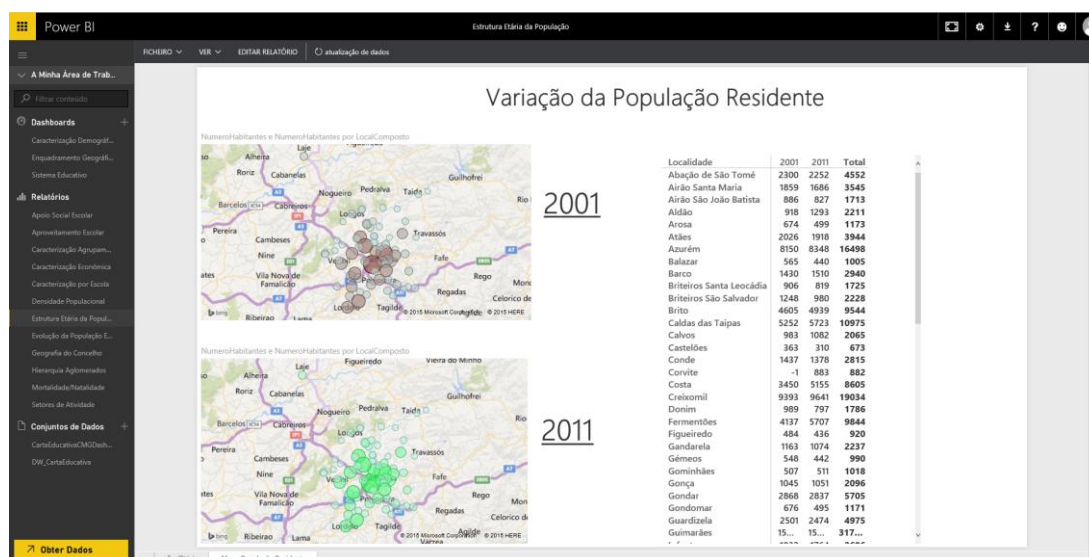


Figura 71 - Relatório Evolução Etária da População 2

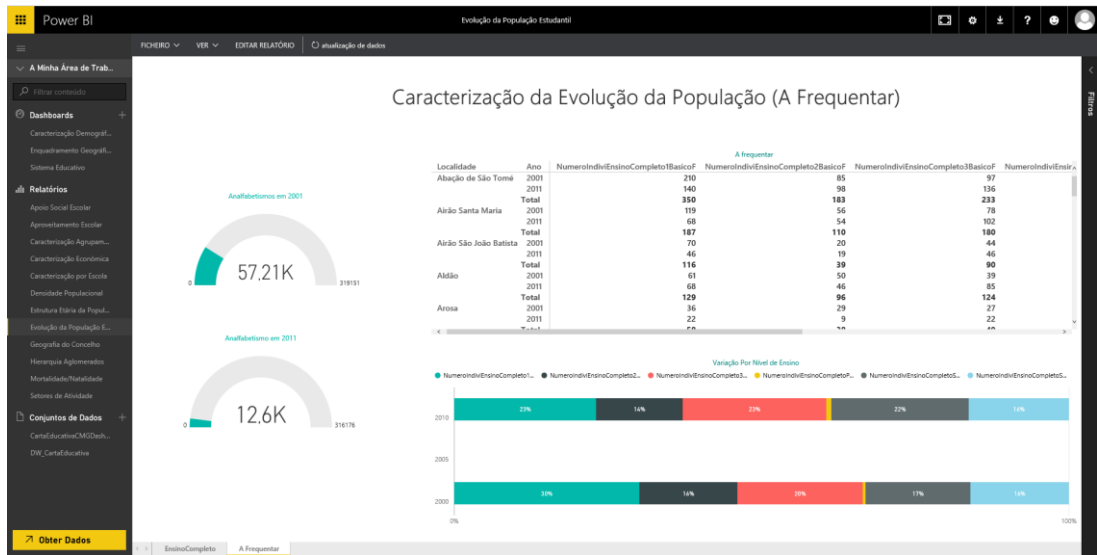


Figura 74 - Relatório Evolução da População Estudantil 2

D9. Vista Geral do Relatório Geografia do Concelho

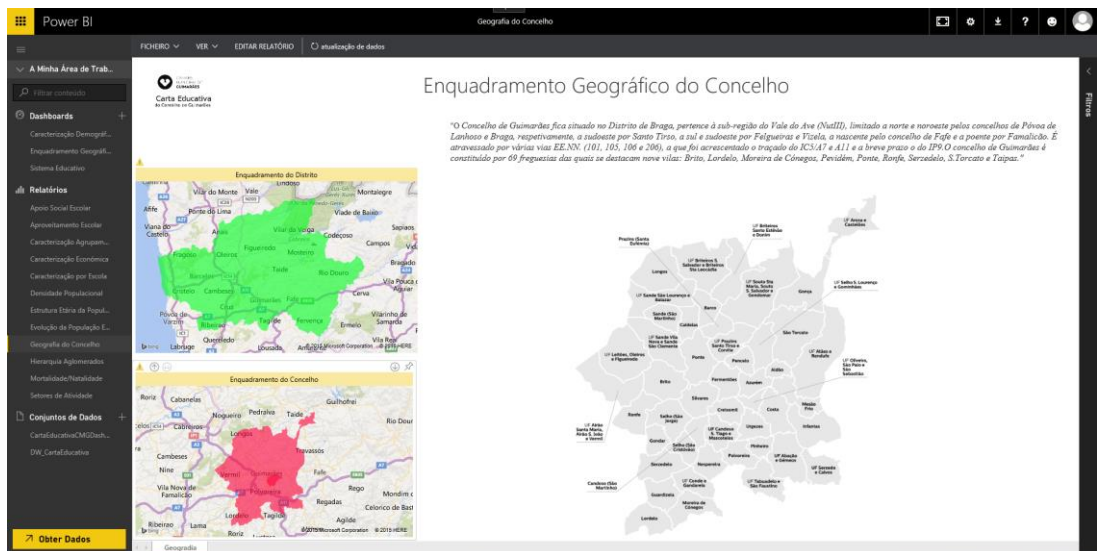


Figura 75 - Relatório Geografia do Concelho

D11. Vista Geral do Relatório Hierarquia dos Aglomerados

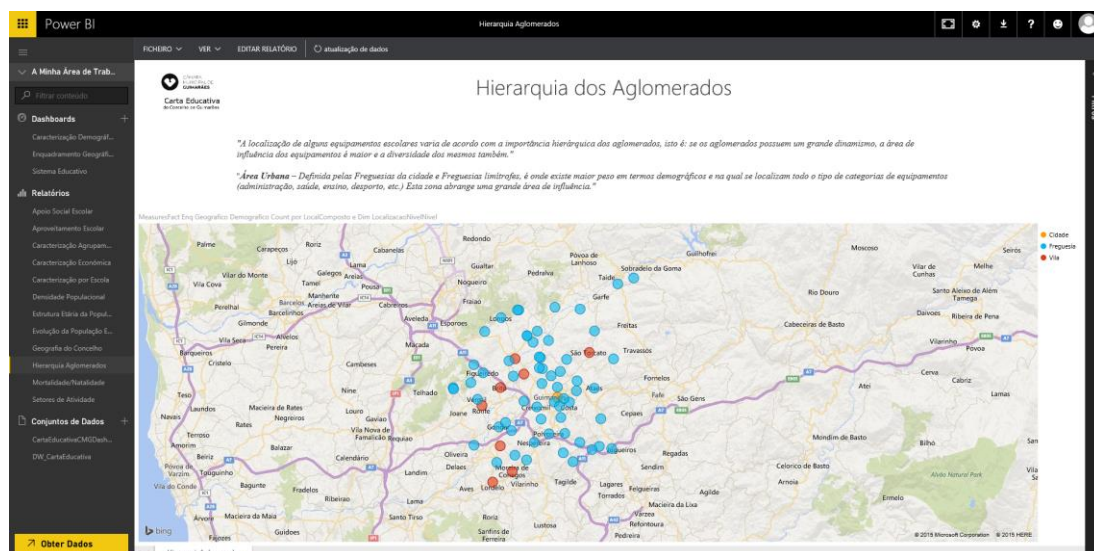


Figura 76 - Relatório Hierarquia dos Aglomerados

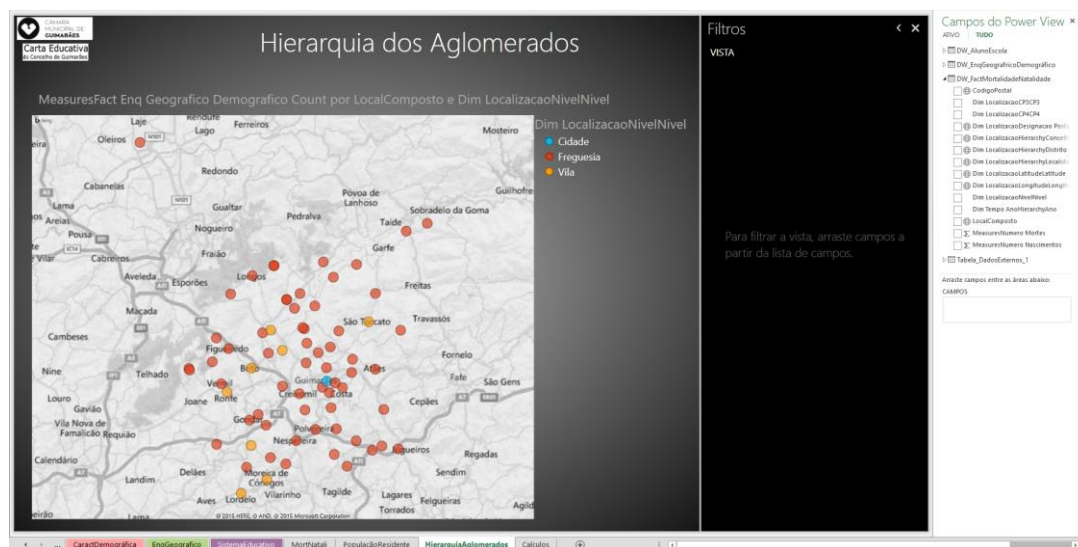


Figura 77 - Hierarquia dos Aglomerados (Excel 2013)

D12. Vista Geral do Relatório Mortalidade/Natalidade

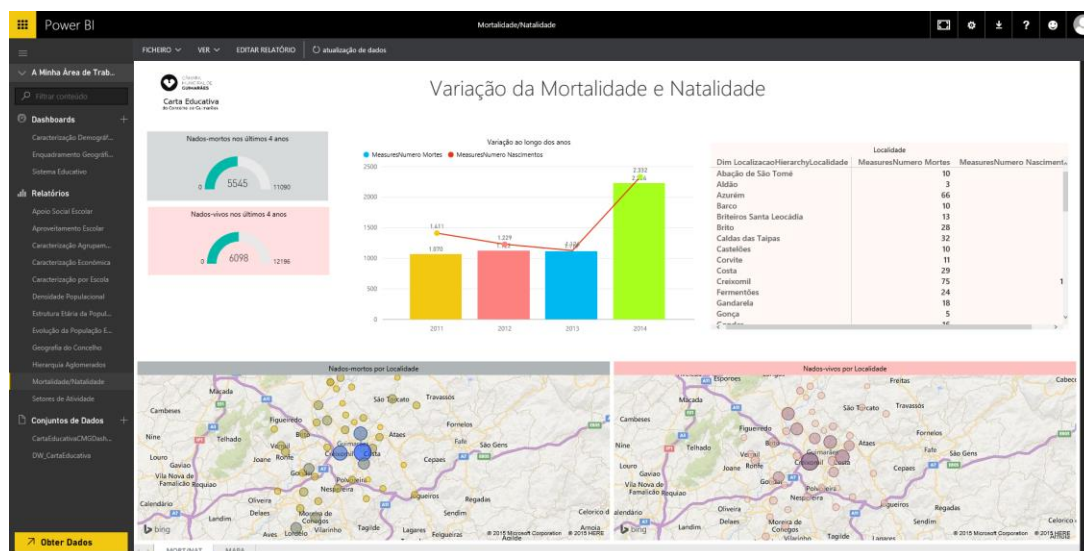


Figura 78 - Relatório Mortalidade/Natalidade 1

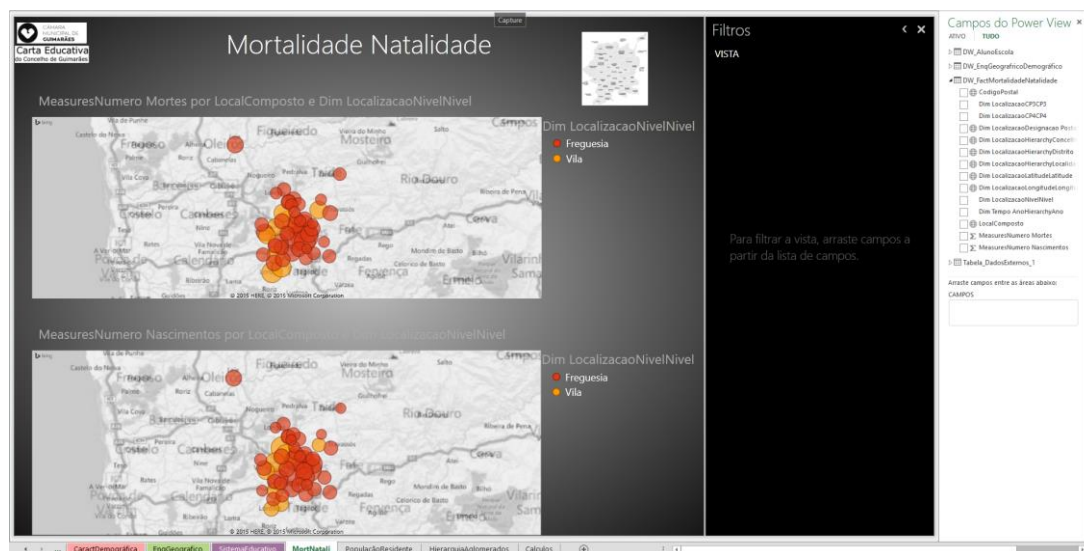


Figura 79 - Mortalidade/Natalidade (Excel 2013)

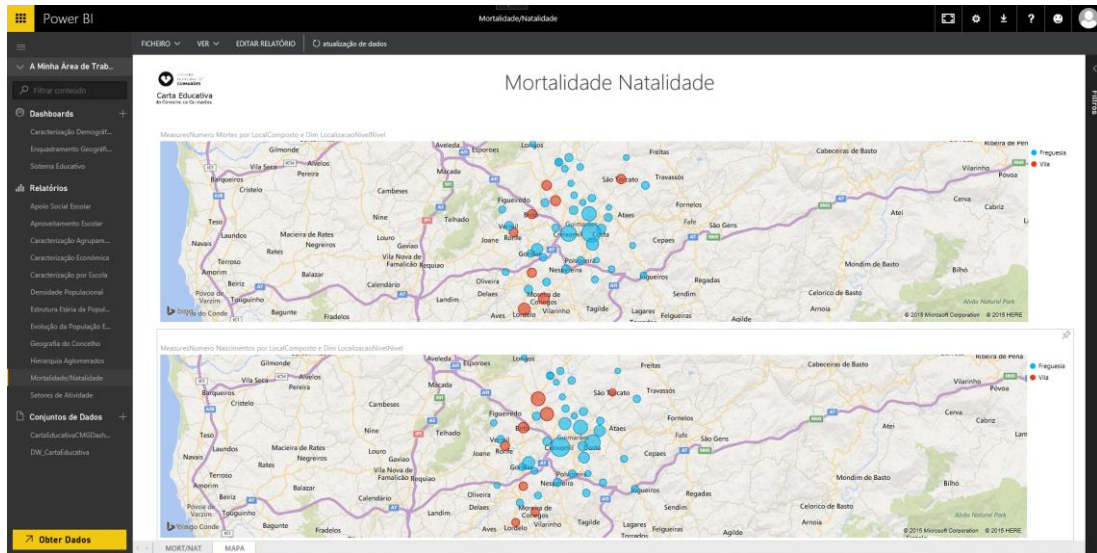


Figura 80 - Relatório Mortalidade/Natalidade 2

D13. Vista Geral do Relatório Setores de Atividade

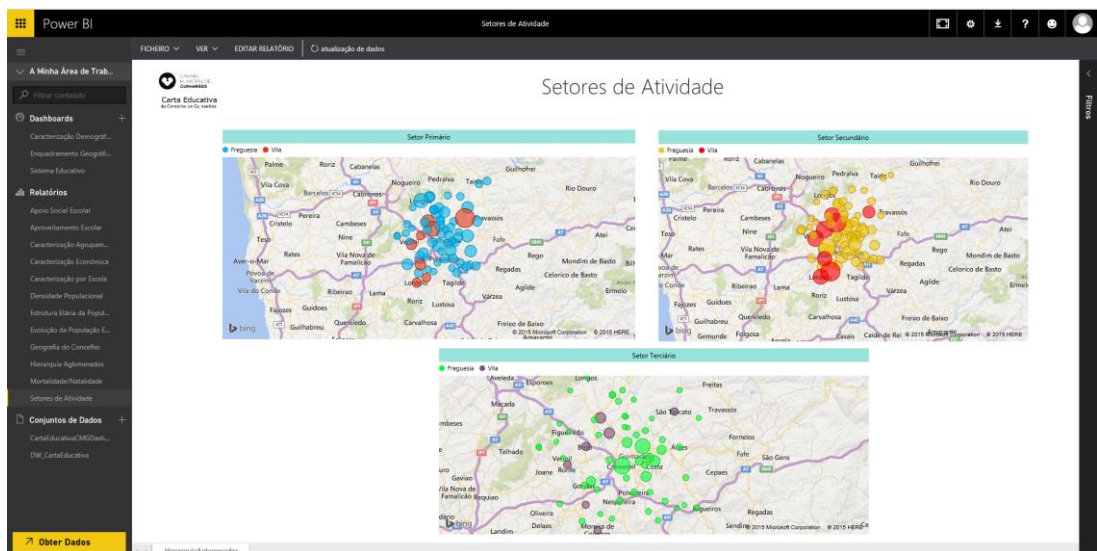


Figura 81 - Relatório Setores de Atividade