



Universidade do Minho
Escola de Engenharia

Ferramentas de Extração e Exploração de Dados para *Business Intelligence*

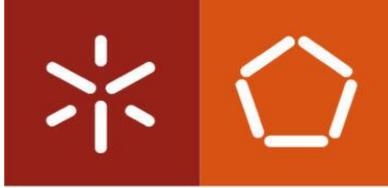
Mário João Gomes Cardoso

UMinho | 2018

Mário João Gomes Cardoso

**Ferramentas de Extração e Exploração de Dados para
*Business Intelligence***

Outubro 2018



Universidade do Minho
Escola de Engenharia

Mário João Gomes Cardoso

**Ferramentas de Extração e Exploração de
Dados para *Business Intelligence***

Dissertação de Mestrado Integrado
em Engenharia e Gestão de Sistemas de Informação

Trabalho efetuado sob a orientação de:
Professor Doutor Manuel Filipe Santos
e
Professor Doutor António Silva Abelha

Outubro 2018

DECLARAÇÃO

Nome: Mário João Gomes Cardoso

Endereço eletrónico: a66349@alunos.uminho.pt

Telefone: 931868714

Título de residência: 58V3F6132

Título da dissertação: Ferramentas de Extração e Exploração de Dados para *Business Intelligence*

Orientadores:

Professor Doutor Manuel Filipe Santos

Professor Doutor António Silva Abelha

Ano de conclusão: 2018

Mestrado Integrado em Engenharia e Gestão de Sistemas de Informação

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA DISSERTAÇÃO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, 14/12/2018

Assinatura: Mário João Gomes Cardoso

AGRADECIMENTOS

O desfecho deste capítulo da minha vida, não lograria realizar-se sem o apoio de várias pessoas. Pelo que, gostaria de agradecer primeiramente, à minha família, em particular, à minha irmã Manu, à minha tia-mãe Ludy e ao meu querido pai Lelé. Pela compreensão, pelo aconselhamento, pelo estímulo, pelo esforço e afeto, transmitidos durante esta fase da minha caminhada. Sou e serei eternamente grato!

Aos meus amigos e amigas para a vida toda, e aos meus colegas da Universidade do Minho, pelos instantes vividos em conjunto.

A todos os docentes e pessoal não docente da Universidade do Minho, que me auxiliaram neste percurso académico.

E ao meu orientador, o Professor Doutor Manuel Filipe Santos, pela orientação e disponibilidade ao longo desta dissertação.

Um muito obrigado, a todos!

Agradesidu di korason!

RESUMO

Os sistemas de *Business Intelligence* (BI) têm sofrido constantes mudanças, atualmente, devido ao surgimento de novas tecnologias, que são introduzidas para melhorar os processos inerentes à tomada de decisão nas organizações. Aumentando a competitividade, através da aplicação dos sistemas de BI, de forma eficiente e eficaz, obtendo mais agilidade nas tomadas de decisão, estratégicas e operacionais. Para tal, é necessário ter uma relação forte com os dados do negócio e que, os utilizadores do negócio tenham melhor acesso para efetuarem as suas análises.

No entanto, nem todos os utilizadores de uma organização estão familiarizados com as ferramentas de um sistema de BI, o que provoca uma enorme dependência da assistência dos técnicos de Tecnologia de Informação (TI) para as atividades de Extração e Exploração de Dados (EED) do negócio, na realização de análises *ad-hoc*. Tendo em conta este pressuposto, este trabalho de dissertação teve como objetivo identificar e explorar as ferramentas de EED, que permitissem dar resposta aos requisitos dos utilizadores mais dependentes da assistência de técnicos de TI.

Este documento apresenta a identificação dos utilizadores mais dependentes e os seus requisitos; os testes das ferramentas de EED (*Performance Point. Tableau. Power BI. Excel* e suplementos) dentro dum contexto de experimentação montado com recurso a várias tecnologias (*VMware Workstation Player, Windows Server, SQL Server*); o desenvolvimento da tipologia de catalogação das ferramentas para avaliar as ferramentas de EED e a determinação da ferramenta mais adequada aos utilizadores identificados, segundo a tipologia.

Neste documento é possível averiguar também, a contextualização do tema em estudo, a descrição dos objetivos e dos resultados esperados, assim como, as metodologias de investigação definidas e a estratégia de pesquisa adotada para a descrição do estado da arte associado.

Palavras-Chave: *Business Intelligence, Data Warehouse, Data Extraction Tools, Data Exploration Tools.*

ABSTRACT

Business Intelligence (BI) systems have undergone constant changes, due to the emergence of new technologies, which are introduced to improve the processes inherent in decision making in organizations. Increasing the competitiveness, through the application of BI systems, efficiently and effectively, getting more agile decision making, strategic and operational. To do this, demands a strong relationship with the business data and a better access for the business users to carry out their analyzes.

However, not all users in an organization are familiar with the tools of a BI system, which causes a great dependence on the assistance of Information Technology (IT) technicians to perform Data Extraction and Exploitation (EED) activities, while conducting ad-hoc analyzes. Given this assumption, this dissertation work had as objective to identify and to explore the tools of EED, that allowed to answer the requirements of the users more dependent on the assistance of IT technicians.

This document presents the identification of the most dependent users and their requirements; the tests of the tools of EED (Performance Point, Tableau, Power BI, Excel and supplements) within a context of experimentation implemented with several technologies (VMware Workstation Player, Windows Server, SQL Server); the development of the cataloging typology of the tools to evaluate the EED tools and the determination of the most appropriate tool to the users identified, according to the typology.

In this document it is possible to also ascertain, the contextualization of the subject under study, the description of the objectives and expected results, as well as the research methodologies defined, and the research strategy adopted to describe the associated state of the art.

KEYWORDS: *Business Intelligence, Data Warehouse, Data Extraction Tools, Data Exploration Tools.*

ÍNDICE

Agradecimentos.....	v
Resumo.....	vii
Abstract.....	ix
Índice.....	xi
Lista de Figuras.....	xv
Lista de Tabelas.....	xvii
Notação e Lista de Abreviaturas, Siglas e Acrónimos.....	xix
Notação.....	xix
Abreviaturas, Siglas e Acrónimos.....	xix
1 Introdução.....	1
1.1 Enquadramento e Motivação.....	1
1.2 Objetivos.....	2
1.2.1 Objetivo Principal.....	2
1.2.2 Objetivos Específicos.....	2
1.3 Questão de Investigação.....	3
1.4 Estrutura do Documento.....	3
2 Revisão de Literatura.....	5
2.1 Estratégia de Pesquisa.....	5
2.1.1 Fontes.....	5
2.1.2 <i>Keywords</i>	6
2.1.3 Gestão de Referências.....	6
2.1.4 Critérios de Seleção.....	6
2.2 Dados.....	7
2.2.1 Conceito.....	7
2.2.2 Tipos de Dados.....	7
2.2.3 Extração e Exploração de Dados.....	8

2.3	<i>Business Intelligence</i>	9
2.3.1	Origem.....	9
2.3.2	Conceito.....	11
2.3.3	Arquitetura BI.....	12
2.3.4	<i>Business Users</i>	14
2.3.5	Indicadores Chave de Desempenho.....	17
2.4	<i>Data Warehouse</i>	17
2.4.1	Conceito.....	17
2.4.2	Arquitetura de Sistemas de <i>DW</i>	19
2.4.3	Componentes do Sistema de <i>DW</i>	20
2.4.4	Arquiteturas do <i>DW</i>	21
2.4.5	Abordagens de Desenvolvimento.....	23
2.4.6	Representação dos Dados no <i>DW</i>	24
2.4.7	Metadados.....	27
2.4.8	Análise e Acesso.....	28
2.5	<i>Extract, Transform & Load</i>	31
2.5.1	Conceito.....	31
2.5.2	Fases do Processo ETL.....	32
2.6	Visão Crítica.....	33
3	Abordagem Metodológica.....	35
3.1	Metodologia de Investigação.....	35
3.2	Aplicação da Metodologia.....	37
4	Ambiente de Experimentação.....	39
4.1	Ferramentas de Desenvolvimento.....	39
4.1.1	<i>Vmware Workstation Player 14</i>	39
4.1.2	<i>MS Windows Server 2016</i>	40
4.1.3	<i>MS SQL Server 2016</i>	41
4.1.4	<i>MS SharePoint Server 2013</i>	41

4.2	Ferramentas de Extração e Exploração de Dados	44
4.2.1	Requisitos das Ferramentas.....	44
4.3	Cenário de Negócio	49
4.4	Arquitetura Tecnológica	49
4.5	Modelo Dimensional de Dados	51
4.5.1	Modelo Dimensional	51
4.5.2	Descrição das Tabelas de Factos	51
4.5.3	Descrição das Dimensões.....	53
4.6	Resumo do Contexto Elaborado	55
4.7	Ambiente de Testes.....	56
5	Tipologia de Catalogação	65
5.1	Métricas de Avaliação.....	65
5.2	Tipologia de Catalogação.....	66
5.3	Resultados da Tipologia.....	69
5.3.1	Perspetiva Individual.....	69
5.3.2	Perspetiva Global.....	70
5.4	Discussão dos Resultados	71
6	Conclusões, Limitações e Trabalho Futuro.....	73
6.1	Síntese do Trabalho.....	73
6.2	Análise dos Objetivos.....	73
6.3	Análise das Metodologias	74
6.4	Resultados Atingidos	74
6.5	Limitações e Trabalho Futuro	75
6.5.1	Limitações.....	75
6.5.2	Trabalho Futuro	75
	Referências	77
	Anexos	80
	Anexo I-Modelo Dimensional	80

Anexo II-Tabelas dos Factos	81
Anexo III-Tabelas das Dimensões	84

LISTA DE FIGURAS

Figura 1- História do Business Intelligence	10
Figura 2- Arquitetura de um sistema BI	12
Figura 3 - Arquitetura de um sistema de BI	13
Figura 4 - Classificação dos Business Users	15
Figura 5- Arquitetura de três camadas	19
Figura 6 - Arquitetura de Duas Camadas	20
Figura 7 - Arquitetura assente na web	20
Figura 8-Arquitetura de Data Marts independentes	22
Figura 9 – Data Marts dependentes	22
Figura 10 - Arquitetura Data Warehouse organizacional	23
Figura 11 - Esquema em Estrela	25
Figura 12 - Esquema em Floco de Neve	25
Figura 13 - Esquema em Constelação	26
Figura 14 - Exemplo de Dashboard	29
Figura 15 - Exemplo de scorecard	30
Figura 16 - Representação de um processo ETL genérico	32
Figura 17 - Metodologia DSR	35
Figura 18 - Arquitetura do VMware W. Player	40
Figura 19 - Arquitetura de um site no SharePoint	42
Figura 20 - Arquitetura da implementação "Farm Solution"	43
Figura 21-Arquitetura Tecnológica – Visão Global	50
Figura 22 – Arquitetura tecnológica	50
Figura 23 – Modelo dimensional vista simples	51
Figura 24 - Dashboard de vendas – Excel	57
Figura 25-Mashup de dados no Power Pivot	58
Figura 26 - Dashboard criado no Power View	58
Figura 27 - Dashboard criado no Excel	59
Figura 28 - Dashboard em ambiente SharePoint	60

Figura 29 – Performance Point Dashboard Designer	61
Figura 30 - Dashboard criado no Performance Point.....	61
Figura 31 – Dashboard criado no Power BI	62
Figura 32 - Dashboard criado no Tableau Desktop	63
Figura 33 - Resultados da Tipologia-Perspetiva Individual.....	70
Figura 34 - Resultados da Tipologia – Perspetiva Global	70
Figura 35 - Processo de Aplicação da Tipologia	71
Figura 36 - Modelo dimensional Adventure Works	80

LISTA DE TABELAS

Tabela 1-Critérios de seleção de artigos	6
Tabela 2-Enquadramento das ferramentas de Extração e Exploração de Dados.....	13
Tabela 3 - Classificação de utilizadores casuais.....	16
Tabela 4 - Tipos de metadados	27
Tabela 6 - Ferramentas de desenvolvimento.....	39
Tabela 7 - Características da máquina virtual	40
Tabela 8 - Classe dos utilizadores	44
Tabela 10 - Tabela das dimensões.....	53
Tabela 11 - Características do ambiente	56
Tabela 12 - Escala de classificação	65
Tabela 13-Peso das Classes dos utilizadores casuais	66
Tabela 14 – Tipologia de classificação das ferramentas	66
Tabela 15 - Descrição do Facto Internet Sales.....	81
Tabela 16 - Descrição do Facto ResellerSales.....	82
Tabela 17 - Descrição do Facto ProductInventory	83
Tabela 18- Descrição da Dimensão Sales Territory	84
Tabela 19 - Descrição da Dimensão Geography.....	84
Tabela 20 - Descrição da Dimensão Product	85
Tabela 21 - Descrição da Dimensão Product Category.....	86
Tabela 22 - Descrição da Dimensão Product SubCategory.....	86
Tabela 23 - Descrição da Dimensão Customer	87
Tabela 24 - Descrição da Dimensão Employee	88
Tabela 25 - Descrição da Dimensão Reseller	89
Tabela 26- Descrição da Dimensão Date.....	90
Tabela 27 - Descrição da Dimensão Currency	91
Tabela 28- Descrição da Dimensão Promotion	91

NOTAÇÃO E LISTA DE ABREVIATURAS, SIGLAS E ACRÓNIMOS

Notação

A notação utilizada ao longo deste documento segue o seguinte padrão:

- Texto em *itálico*, para palavras em língua estrangeira (e.g., Inglês);
- Texto em **negrito**, para realçar uma palavra ou expressão.

Abreviaturas, Siglas e Acrónimos

AW	<i>Adventure Works</i>
BI	<i>Business Intelligence</i>
BD	Base de dados
DCE	Desenvolvimento do Contexto de Experimentação
DEE	<i>Data Extraction and Exploration</i>
DMT	<i>Data Mart</i>
DM	<i>Data Mining</i>
DW	<i>Data Warehouse</i>
DWO	<i>Data Warehouse Organizacional</i>
ETL	<i>Extract, Transform & Load</i>
EED	Extração e Exploração de Dados
KPI	<i>Key Performance Indicators</i>
MS	<i>Microsoft</i>
OLAP	<i>Online Analytical Processing</i>
SI	Sistemas de Informação
SS	<i>SQL Server</i>
TI	Tecnologias de Informação

1 INTRODUÇÃO

O presente capítulo, desta dissertação apresenta: o enquadramento e a motivação para a concretização da dissertação e a estrutura do documento.

1.1 Enquadramento e Motivação

Ao longo dos anos, as aplicações informáticas permitiram às organizações ter melhor controlo sobre os dados que lidam diariamente, permitindo que, a informação gerada, faculte aos gestores um conjunto de indicadores sobre o negócio. Esta utilidade, auxilia a perceber o que aconteceu e o que acontece no negócio dentro das organizações, preparando-as na elaboração de cenários de previsão com mais eficácia e eficiência. Pelo que, a crescente dependência das organizações perante a utilização das aplicações de *Business Intelligence* (BI), para extrair, tratar e organizar os dados necessários, acarreta mais responsabilidades aos utilizadores destas aplicações.

Neste sentido, as capacidades de análise e adaptação dos utilizadores face às tecnologias adotadas nas organizações, é um fator decisivo para o sucesso na tomada de decisão. No entanto, nem todas as organizações levam em conta a inadaptação dos utilizadores mais dependentes, perante as tecnologias implementadas, sendo que, em certos casos, fazer corresponder a tecnologia com esses utilizadores através da identificação das suas necessidades e capacidades, traz melhorias no processo de tomada de decisão dentro das organizações. Pois, no caso dos sistemas de BI, estes, não podem simplesmente ser de exercício técnico do departamento de Sistemas de Informação (SI). Mas sim, devem servir como uma forma de alterar a conduta de gestão do negócio, melhorando os seus processos de negócio e transformando os processos de decisão, para serem orientados por dados (Turban et al., 2017). Com esta alteração pretende-se causar um bom impacto no ambiente de trabalho, sem a exclusão de nenhum tipo de utilizador.

Dentro desse âmbito, são escassos os trabalhos científicos que apresentam formas de avaliar as ferramentas de Extração e Exploração de Dados (EED) de acordo com requisitos de certos tipos de utilizadores, pelo que, a presente dissertação visa contribuir para colmatar esta lacuna, com o

desenvolvimento de uma tipologia de catalogação que possibilite identificar as ferramentas orientadas à EED estruturadas de sistemas de BI, adequadas aos utilizadores com menos conhecimentos técnicos e à organização no geral, no intuito de torná-los menos dependentes dos departamentos de SI.

Com este trabalho, pretende-se facultar novos parâmetros de avaliação das ferramentas de EED, com a elaboração de uma tipologia de catalogação das mesmas, assim como, esclarecer o propósito das atividades de EED. Isto, visa proporcionar aos utilizadores mais dependentes, a ferramenta mais adequada para uma realização eficiente e eficaz das atividades de EED, melhorando assim, a tomada de decisão dentro das organizações.

1.2 Objetivos

Este capítulo apresenta os objetivos definidos pelo autor para esta dissertação, assim como, as linhas orientadoras para o progresso do mesmo.

1.2.1 Objetivo Principal

Esta dissertação teve como objetivo principal: identificar e explorar as ferramentas orientadas à Extração e Exploração de Dados (EED), a partir de *Data Warehouses* ou *Data Marts* associados a sistemas de *Business Intelligence*, de acordo com os requisitos dos utilizadores mais dependentes.

1.2.2 Objetivos Específicos

Para esta dissertação, foi necessário ultrapassar um conjunto de desafios, de modo que resultado chegasse a um bom porto. Desta forma, os objetivos específicos para a realização desta dissertação foram:

- Levantar/identificar os utilizadores mais dependentes das ferramentas de EED e os seus requisitos;
- Experimentar as ferramentas de EED num contexto *ad-hoc*;
- Elaborar uma tipologia de catalogação e avaliação;
- Executar a avaliação das ferramentas, de acordo com a tipologia.

1.3 Questão de Investigação

Para elaborar um trabalho desta dimensão, existe uma questão de investigação para ajudar a perceber a problemática que se pretende resolver, pelo que, foi formulada a seguinte questão:

1. **“Em que medida as ferramentas existentes no mercado, correspondem aos requisitos dos utilizadores do negócio mais dependentes?”**

1.4 Estrutura do Documento

O presente trabalho está estruturado em seis capítulos, com o seguinte conteúdo:

1. **Revisão de Literatura** – Divulga-se o estudo do estado da arte dos conceitos pilares para o avanço do presente trabalho, tais como: *Business Intelligence*, *Data Warehouse*, *Extract, Transform & Load* (ETL), e a reflexão crítica da literatura estudada;
2. **Abordagem Metodológica** – Expõe as metodologias selecionadas para a execução da investigação e para a execução prática do trabalho, nomeadamente, o *Design Science Research*;
3. **Ambiente de Experimentação** – Exposição de toda a estrutura tecnológica montada, para apurar e testar as características das ferramentas selecionadas, assim como, apresentação dos requisitos identificados para avaliar as ferramentas orientadas à Extração e Exploração de Dados;
4. **Tipologia de Catalogação** – Capítulo que aglomera a definição das métricas, a tipologia elaborada e os resultados da avaliação das ferramentas;
5. **Conclusões, Limitações e Trabalho Futuro** – Reflexão sobre o concretizado desta dissertação, assim como, o enquadramento do conhecimento produzido e dos próximos desafios.

2 REVISÃO DE LITERATURA

Este trabalho insere-se na área de *Business Intelligence*, onde perante a proliferação de novas tecnologias (*cloud computing*, *web data*, *real-time Data warehouse*) utilizadas neste campo (Trujillo & Mat, 2012), torna-se necessário facultar ferramentas adequadas, que possibilitem aos utilizadores do negócio, concretizarem as suas atividades de análise e exploração dos dados, de forma simples, eficaz e eficiente.

Pelo que, este capítulo apresenta a revisão da literatura e o estado da arte dos conceitos associados ao desenvolvimento desta dissertação, sendo que uma boa revisão deve: Motivar o tópico da pesquisa e explicar as contribuições da revisão; descrever os conceitos-chave; delinear os limites da pesquisa; rever a literatura relevante em Sistemas de Informação e áreas relacionadas; desenvolver um modelo para orientar futuras pesquisas; justificar proposições apresentadas, explicações teóricas, achados empíricos passados e exemplos práticos; e apresenta conclusões para investigadores e profissionais da área (Webster & Watson, 2002).

2.1 Estratégia de Pesquisa

2.1.1 Fontes

No âmbito da revisão da literatura feita, foram acedidos os seguintes repositórios online, de modo a obter informações relevantes sobre os conceitos a serem abordados: *Scopus*¹; *Science Direct*²; *Web Of Science*³; Repositório da Universidade do Minho⁴; *Google Scholar*⁵.

¹ <http://scopus.com>

² <https://www.sciencedirect.com/>

³ <https://apps.webofknowledge.com>

⁴ <https://repositorium.sdum.uminho.pt/>

⁵ <https://scholar.google.com>

2.1.2 *Keywords*

A pesquisa da literatura foi realizada com as palavras-chaves: *Business Intelligence, Data, Data Warehouse, Data Extraction Tools, Data Exploration Tools*. Durante a pesquisa também foram efetuadas combinações entre as palavras-chave acima mencionadas, de modo a filtrar informação e encontrar as melhores fontes de informação.

2.1.3 Gestão de Referências

Para salvaguardar e organizar a seleção da literatura, em linha com os conceitos necessários para o cumprimento dos objetivos desta dissertação, recorreu-se à ferramenta *Mendeley Desktop*⁶.

2.1.4 Critérios de Seleção

De modo a obter uma base de literatura adequada ao trabalho desenvolvido, foi necessário definir critérios coerentes para a discernir os **artigos, livros, white papers, dissertações e conference proceedings**.

Tabela 1-Critérios de seleção de artigos

Critério	Relevância (1-3)
Ano de Publicação	2
Número de Citações	3
Jornal/Conferência	1
<i>Abstract</i>	3

Na tabela 1 encontram-se representados os critérios definidos, assim como, o grau de relevância dos mesmos (escala de 1 a 3, menos relevante para o mais relevante). O ano de publicação foi limitado no intervalo temporal: 2008 até 2017 (salvo algumas referências genéricas sobre o conceito de Business Intelligence, Dados e Data Warehouse), os *abstracts* foram avaliados destacando a ocorrência das palavras-chave, a situação do contexto e a relação do objetivo com o tema de dissertação.

⁶ *Mendeley* é um programa de desktop e web produzido pela Elsevier, para gerir e partilhar documentos de investigação, descobrir dados de investigação e colaboração online.

2.2 Dados

2.2.1 Conceito

“An ounce of information is worth a pound of data.

An ounce of knowledge is worth a pound of information.

An ounce of understanding is worth a pound of knowledge” (Ackoff, 1989).

Negash (2004) destaca, que um dos propósitos fundamentais dos sistemas de BI, é a capacidade de converter dados em informação útil e através de análises humanas, em conhecimento. No contexto de *Business Intelligence*, é importante distinguir dados, de informação, pois, ambos estão mutuamente interligados, mas provêm de processos e fontes distintas.

Dados são símbolos que representam as propriedades de objetos ou eventos (Ackoff, 1989; Davenport & Prusak, 2000). Davenport & Prusak (2000, pg. 2), acrescentam ainda, que, num “contexto organizacional, os dados são mais propriamente descritos, como registros estruturados de transações”. Por exemplo, um posto de gasolina registra, quando um cliente paga pelo abastecimento de combustível, a quantidade de gasolina, a quantia paga, e a hora do pagamento pelo serviço. Nesse contexto, os dados derivam da transação efetuada e a diferença entre esses dados e a informação, tem a ver com o propósito/utilidade, isto é, esses dados registrados podem ser apenas detalhes, sem traduzir em nenhum impacto para o negócio e, a informação por sua vez, contem um propósito e importância, permitindo retrair ilações ou respostas a determinadas questões (Ackoff, 1989; Davenport & Prusak, 2000).

Bellinger et al. (2004), defendem que, os dados representam um facto ou declaração de um evento sem relação com outros conceitos e, que a transição de dados para informação, informação para conhecimento, e conhecimento para sabedoria, acontece de acordo com a compreensão dos mesmos.

2.2.2 Tipos de Dados

No campo de BI, é requisitado a análise de dados estruturados, semiestruturados ou até não-estruturados, estando disponível em vários formatos (Grossmann & Rinderle-Ma, 2015; Negash, 2004).

Baars & Kemper (2008, pg.133) afirmam que a maioria das fontes informacionais dos sistemas de BI, são documentos eletrônicos (emails, páginas web, documentos pdf, etc), pelo que se torna imperativo considerar os dados estruturados e não-estruturados, para facultar conhecimentos válidos para desenvolvimento do negócio.

Os tipos de dados existentes, são:

- **Dados Semiestruturados:** no contexto de base de dados relacionais, são todos os dados que não podem ser armazenados orientado à coluna ou à linha. Não possuem formato pré-definido, não obedecendo aos esquemas tradicionais das base de dados relacionais, podendo possuir alguma estrutura parcial (Baars & Kemper, 2008; Negash, 2004). Exemplos: ficheiros XML e JSON, páginas web, documentos pdf, entre outros;
- **Dados Estruturados:** são dados organizados em registros com valores de dados simples (variáveis categóricas, ordinais e contínuas) e armazenados nos sistemas de gestão de base dados (Inmon, 2005; Turban et al. , 2017). Exemplos: números, transações financeiras, datas, etc.
- **Dados Não Estruturados:** são dados que não possuem nenhum formato e nem registo, sendo maioritariamente armazenados em textos e a maior parte dos dados disponíveis na internet, hoje, são os não estruturados (Turban et al., 2017). Os dados não estruturados segundo Inmon (2005), podem ser divididos em duas grandes categorias: **comunicações** e **documentos**, possuindo dados não estruturados em formas de texto. As comunicações tendem a ser relativamente curtas e são de distribuição muito limitada. Os documentos tendem a servir uma audiência mais ampla e geralmente, são maiores do que as comunicações. Exemplos: imagens, vídeos, documentos de texto, mensagens de email e de redes sociais, etc.

Para este trabalho, importam os dados estruturados, pois foram avaliadas as ferramentas que lidam com estes, dentro dum contexto *ad-hoc* e em conexão com sistemas de BI.

2.2.3 Extração e Exploração de Dados

O desenvolvimento do presente trabalho, implicou esclarecer o significado inerente às atividades de Extração e Exploração de Dados (EED), pois, as ferramentas estudadas desempenharam fundamentalmente, essas duas funções. Também, foi relevante distinguir os termos *Information Extraction* (Extração de Informação) e *Data Extraction* (Extração de Dados).

Na literatura existente, a diferença entre esses dois termos, reside no facto de ambas terem propósitos distintos, ou seja, a **Extração de Informação** é a transformação de dados não estruturados em dados estruturados, através de técnicas de *machine learning*, *text mining* e de processamento de linguagem natural, entre outras, sendo que a **Extração de Dados** lida com a aquisição/obtenção de dados de fontes informacionais, maioritariamente estruturados (Turban et al., 2017).

Data Extraction (Extração de Dados) no âmbito deste trabalho, é definido como o processo de extração/obtenção de dados estruturados representativos a estados ou objetos do negócio, a partir de *Data Warehouses* e *Data Marts* associados a sistemas de BI. Onde, após a obtenção desses dados, inicia-se o processo de **Data Exploration (Exploração de Dados)**, que se traduz na exploração dos dados (através de *dashboards*, *reports*, *scorecards*, etc) para dar respostas a análises *ad-hoc*, sobre indicadores do negócio.

2.3 Business Intelligence

2.3.1 Origem

A determinação exata da origem do termo *Business Intelligence*, é algo que muitos autores não puderam estabelecer com precisão, embora o termo esteja presente no artigo de Luhn (1958, pg. 314), o termo *Intelligence* foi definido como “a capacidade de apreender as inter-relações dos factos apresentados, de forma a orientar a ação para um objetivo pretendido” e o termo *Business*, como “uma coleção de atividades efetuadas para qualquer finalidade, seja ciência, tecnologia, comércio, indústria, direito, governo, defesa, etc”. Luhn (1958, pg. 314), também especifica o termo *Business Intelligence System*, como “um sistema automático, que está a ser desenvolvido para disseminar informação para os vários setores de qualquer indústria, organização governamental ou científica”.

No entanto, o termo BI só ganhou ênfase 30 anos depois, e em 1989, Howard Dresner conotou-o novamente, como um termo genérico para um conjunto de conceitos e métodos, com objetivo de melhorar a tomada de decisões empresariais, utilizando sistemas baseados em factos (Grossmann & Rinderle-Ma, 2015, pg. 1). Os autores Turban et al. (2017) confirmam que, o conceito é ainda mais antigo, tendo suas raízes nos sistemas de relatórios dos Sistemas de Informação de Gestão (SIG), da década de 1970.

A publicação (Macaulay, 2016), da revista CIO (*Chief Information Office*), apresentou a seguinte cronologia (figura 1), sobre a origem e evolução do termo BI, de forma resumida:

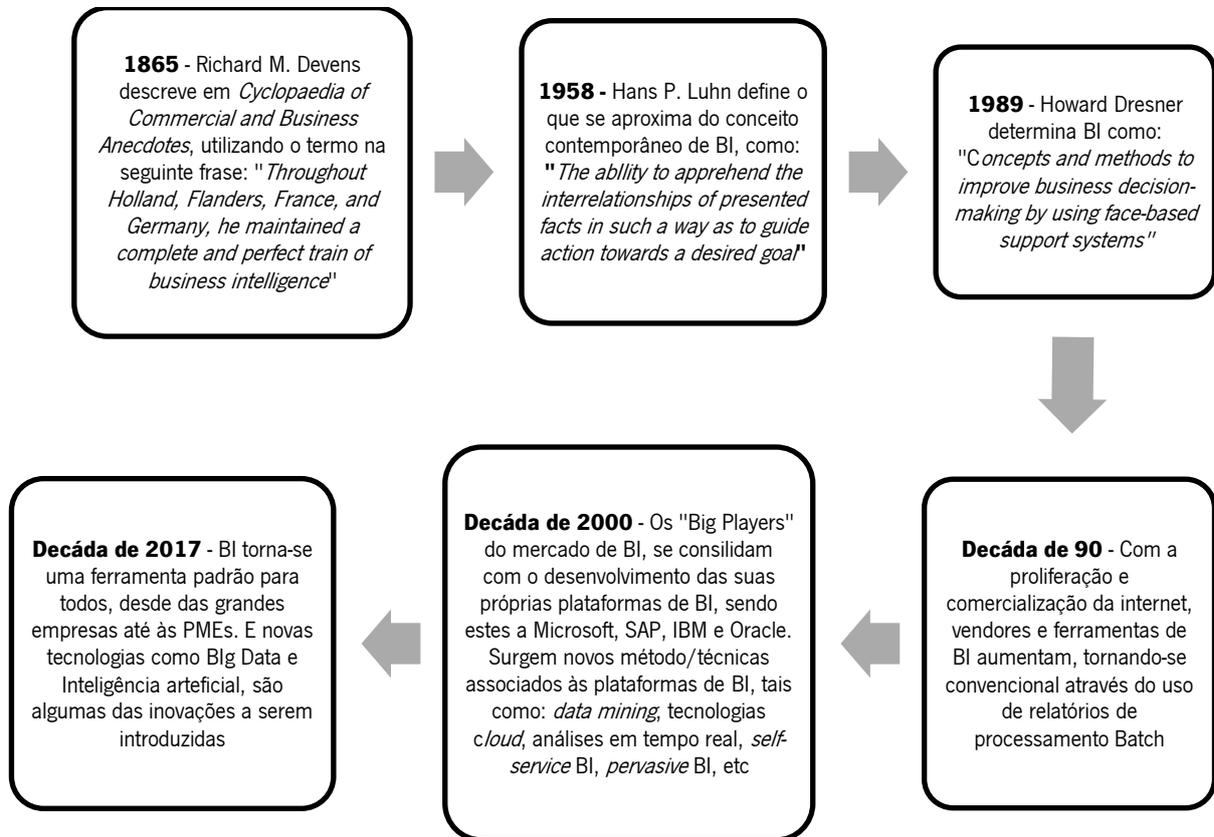


Figura 1- História do Business Intelligence (Adaptado de Macaulay, 2016)

2.3.2 Conceito

Turban et al. (2017, pg. 8) afirma que, “*Business Intelligence* é um termo genérico que combina arquiteturas, armazenamento de dados, ferramentas analíticas, aplicações informáticas, e metodologias”. Enfatizando, que esta, é uma expressão livre, pelo que pode ter diferentes significados para diferentes pessoas. Para estes autores, o principal objetivo do BI, “é habilitar o acesso interativo (às vezes em tempo real) aos dados, possibilitar a manipulação de dados e facultar aos gestores e analistas do negócio a capacidade de efetuar análises adequadas”.

Grossmann & Rinderle-Ma (2015) com base em várias definições, caracterizaram BI como detentora das seguintes características:

- **Serviço:** fornecer apoio à tomada de decisão para objetivos específicos, definidos no contexto de atividades comerciais em diferentes áreas de domínio, levando em consideração o quadro organizacional e institucional;
- **Alicerce:** depende de informações empíricas baseadas nos dados e o uso de diferentes tipos de conhecimento e teorias para produção de informação;
- **Conceção:** o apoio à decisão deverá ser concebido como um sistema, empregando as capacidades facultadas pelas Tecnologias de Informação e Comunicação (TIC);
- **Apresentação:** um sistema de BI deverá fornecer informações no momento certo, às pessoas certas, de forma apropriada.

Um sistema de BI, é uma coleção de tecnologias de suporte à decisão para a empresa, com o objetivo de capacitar trabalhadores do conhecimento, tais como executivos, gestores e analistas a tomar decisões melhores e mais rápidas (Aruldoss, Lakshmi Travis, & Prasanna Venkatesan, 2014; Chaudhuri, Dayal, & Narasayya, 2011). Negash (2004, pg. 178) afirma, de forma mais específica que: “Os sistemas de BI combinam a recolha de dados, armazenamento de dados e gestão do conhecimento com ferramentas analíticas, para amostrar informações internas, complexas e competitivas, aos gestores e decisores”.

Um aspecto fundamental e atual, que alguns dos autores referidos anteriormente não destacaram foi os benefícios que um sistema de BI poderá proporcionar. No entanto, Aruldoss et al. (2014) e Turban et al. (2017), abordam esses aspectos, afirmando que, os sistemas de BI analisam o desempenho de uma organização e aumentam as receitas e a competitividade, facultando a informação certa, na hora certa, à pessoa certa, sendo que a competitividade é um imperativo para sobrevivência do negócio.

2.3.3 Arquitetura BI

De acordo com Turban et al. (2017, pg. 10), um sistema de *Business Intelligence* possui 4 componentes principais, exibidas na figura 2: um **Data warehouse** e as suas fontes de dados; **Análise do negócio**, uma coleção de ferramentas para manipulação, *mining* e análise, dos dados presentes no DW; **Gestão do desempenho empresarial**, para analisar e monitorar o desempenho; e a interface do utilizador (*dashboards*, portal e relatórios).

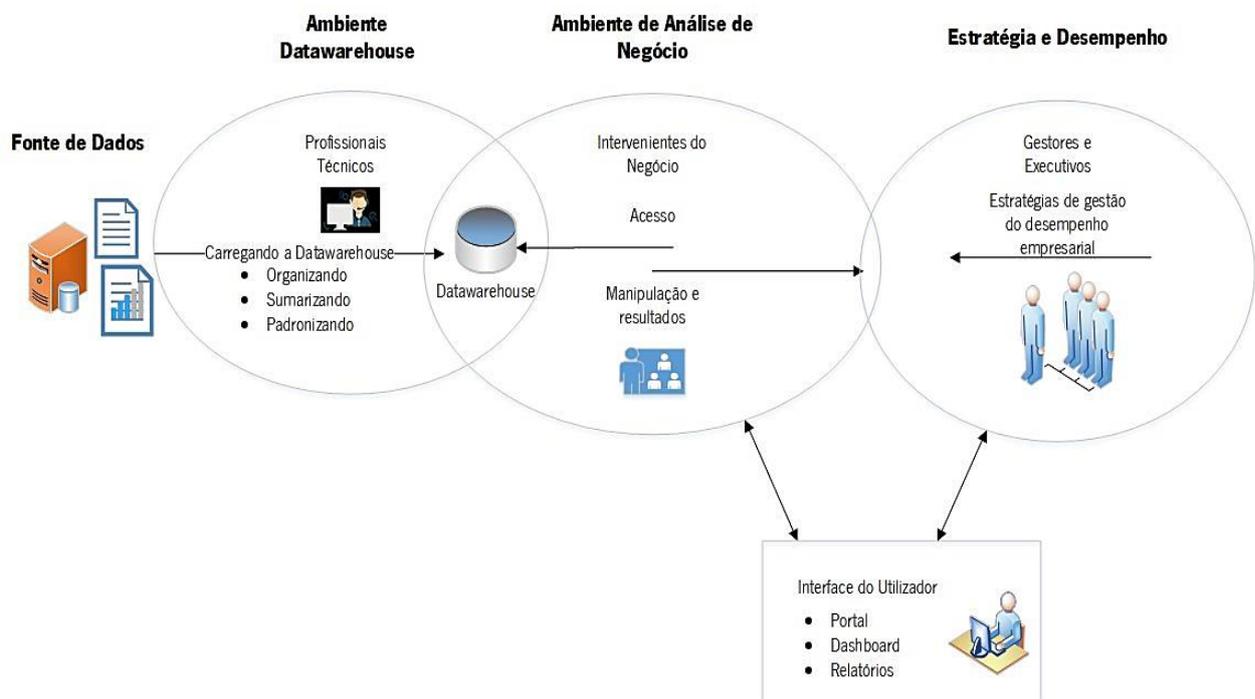


Figura 2- Arquitetura de um sistema BI (Adaptado de Turban et al., 2017)

Em Chaudhuri et al. (2011, pg. 89) é exposto uma arquitetura típica do sistemas BI, onde para além das componentes exibidas por Turban et al. (2017), esta, representa-as, de forma minuciosa, como podemos observar na figura 3.

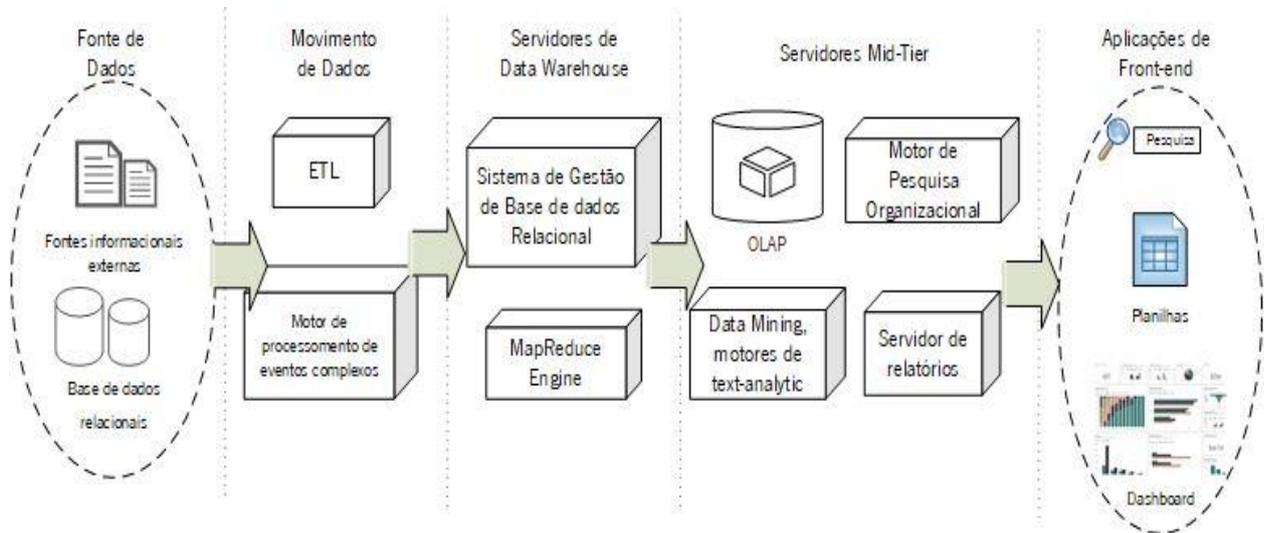


Figura 3 - Arquitetura de um sistema de BI (Adaptado de Chaudhuri et al., 2011)

Tendo em conta as arquiteturas de BI apresentadas anteriormente, as ferramentas de Extração e Exploração de Dados (EED) podem ser enquadradas no formato demonstrado na tabela 2.

Tabela 2-Enquadramento das ferramentas de Extração e Exploração de Dados

Arquitetura	Camada	Ferramentas de EED	Justificação
Turban et al. (2017)	Análise do Negócio e Gestão do Desempenho Empresarial	Presente	As atividades de Extração correspondem ao acesso e manipulação, assim como, a atividade de Exploração incide sob a apresentação/visualização dos dados. As ferramentas que desempenham estas atividades podem estar presentes nas duas camadas da arquitetura apresentada por Turban et al. (2017)

Arquitetura	Camada	Ferramentas de EED	Justificação
Chaudhuri et al. (2011)	Mid-Tier e Aplicações de Front-end	Presente	As ferramentas que desempenham as atividades de Extração e Exploração de Dados, assumem as funcionalidades de acesso (Mid-Tier) e de interface de visualização dos dados (Front-end).

2.3.4 Business Users

Um dos aspetos mais importantes de uma implementação de *Business Intelligence*, bem-sucedida, é que este seja benéfico para a empresa como um todo (Turban et al., 2017). Um sistema de BI é gerido por um conjunto de utilizadores do negócio (*Business users*), pertencentes a diferentes grupos de trabalho, isto é, a sua utilização não se destina somente a um grupo restrito dentro de uma organização. Segundo Turban et al. (2017), um sistema de BI não pode simplesmente ser de exercício técnico, do departamento de sistemas de informação. Este, deve servir como uma forma de alterar a conduta de gestão do negócio, melhorando os seus processos de negócio e modificando os processos de decisão, para serem orientados por dados.

Na área dos sistemas de BI, Dyché, (2007) afirma que, “o termo *easy-of-use* é relativo, pois um relatório mensal, por exemplo, de um utilizador, é o projeto de análise de previsão, de outro e o parecer/percepção de cada utilizador do negócio, pode ser distinto. Classificar e compreender as necessidades de cada utilizador, torna-se elementar para facilitar: a identificação dos requisitos dos dados e do negócio, a seleção apropriada das tecnologias no mercado, assim como, a formação dos utilizadores”.

2.3.4.1 Classificação

No estudo apresentado por Dyché (2007), foram classificados os diferentes tipos de utilizadores finais de uma companhia telefónica, contrastando as diferentes categorias de análise e os diferentes tipos de utilizadores. A autora afirma que maior parte dos utilizadores da companhia telefónica, eram utilizadores casuais, sem conhecimento técnico suficiente das ferramentas de BI, onde se limitavam a

trabalhar com relatórios, *dashboards* e *scorecards*, que refletissem os indicadores chave de desempenho (KPI) da organização. A classificação foi feita através de uma pirâmide, como ilustra a figura 4:



Figura 4 - Classificação dos Business Users (Adaptado de Dyché, 2007)

Eckerson (2010), apresenta uma classificação distinta, agrupando os utilizadores em duas categorias: *Information Producers* (Produtores de Informação) e *Information Consumers* (Consumidores de Informação). De acordo com o autor, os Produtores de Informação são os **utilizadores avançados** (*power users*), que criam blocos de informação para outros consumirem. Podem ser: desenvolvedores de tecnologias de informação que elaboram/constroem *dashboards* e relatórios complexos; superutilizadores que criam relatórios e *dashboards* para os colegas de departamento; analistas de negócio que utilizam uma variedade de ferramentas de análise para explorar; ou estatísticos e analistas, que criam modelos preditivos. Os Consumidores de Informação, são os ditos **utilizadores**

casuais (*casual users*) que regularmente consultam relatórios, mas não executam análises profundas ou de previsão diariamente. Alguns exemplos destes são: executivos, gestores, clientes, funcionários e fornecedores, que utilizam maioritariamente relatórios e *dashboards*; e analistas que trabalham com ferramentas de exploração dos dados (OLAP, *Excel*, ferramentas de *data mining*, etc). Eckerson (2013), agrupou os utilizadores casuais em três níveis, como demonstra a tabela 3.

Tabela 3 - Classificação de utilizadores casuais (Adaptado de Eckerson, 2013)

Classe	Descrição	Papel/Cargo	Necessidades Analíticas	Preferências de Layout (Apresentação)	Canais Preteridos
Visualizador	Visualiza relatórios e <i>dashboards</i> estáticos	Executivos, responsáveis por vendas e funcionários	Efetua questões aos analistas de suporte	Tabelas e gráficos suplementares, de apresentação estática	Email, documentos pdf, dispositivos móveis
Navegador	Navega e executa operações sobre os dados presentes nos relatórios e <i>dashboards</i> , procurando mais detalhe	Gestores que precisam informação sobre o desempenho do negócio	Operações sobre os dados (<i>drill-down</i> , <i>pivot</i> , <i>ranking</i> , modificar, etc) e solicita apoio dos analistas	Gráficos e tabelas de dados complexos e dinâmicos	Plataformas web e dispositivos móveis
Explorador	Explora os dados da camada semântica dos sistemas de BI e elabora relatórios complexos	Analistas	Exploração <i>ad-hoc</i> e elaboração de relatórios complexos	Camada semântica e interfaces de <i>point-and-click</i>	Computadores Desktop

No contexto deste trabalho, interessa avaliar as necessidades dos utilizadores casuais, pois são os utilizadores finais, que mais dependem do suporte dos técnicos de TI, para aceder, solucionar e trabalhar os dados nas ferramentas de exploração/extração, pelo que a classificação de Eckerson

(2017; 2013), apresenta estratégias de como identificar as necessidades e requisitos desses utilizadores, que vão de encontro com o âmbito desta dissertação.

2.3.5 Indicadores Chave de Desempenho

Os sistemas de *Business Intelligence* utilizam os KPI para avaliar o estado atual do negócio e para prescrever o curso da ação face aos objetivos estratégicos das organizações. Para um negócio atingir um determinado objetivo, as organizações devem ter uma combinação de métricas quantitativas, que ajudam a avaliar o seu sucesso e progresso, em direção a um determinado objetivo. Estas métricas, são os KPI's (*Key performance indicators* – Indicadores chave de desempenho), que suportam os decisores na avaliação do estado atual do negócio e de como adotar atividades, atitudes e tarefas que possam ter impacto nesse estado (Sauter, 2011).

De acordo com Grossman et al. (2015), um KPI interliga as atividades de negócio aos objetivos, definindo uma meta quantitativa. Estes, podem referir a aspetos de desempenho de um processo de negócio ou do negócio em geral. Os KPI podem ser: quantitativos (apresentados como números); direcionais que indicam o progresso (ou a ausência deste); práticos, que indicam os processos de negócio; acionáveis, para controlar as mudanças a serem feitas, se necessário; financeiros, relativos a custos (Grossmann & Rinderle-Ma, 2015; Sauter, 2011).

2.4 Data Warehouse

2.4.1 Conceito

Os sistemas de *Business Intelligence* alimentam-se dos dados existentes nos sistemas transacionais das organizações. Pelo que, para o armazenamento dos dados são utilizados *Data Warehouses* ou *Data Marts*. El-Sappagh et al. (2011, pg. 91) define DW, como “uma coleção de tecnologias destinadas a possibilitar que, os decisores tomem decisões melhores e mais rápidas”. Outros autores como Turban et al. (2017, pg. 32), definiram DW, como “um conjunto de dados, produzidos para apoiar a tomada de decisão; e também, um repositório de dados históricos de relevância para gestores da organização”.

Os *Data Warehouses* são orientados a um determinado assunto, integrados, variantes no tempo e não-voláteis, para apoiar a decisão (Inmon, 2005, pg. 29; Turban et al., 2017, pg. 32). Pelo que diferem das base de dados operacionais, por possuir os atributos mencionados anteriormente, assim como, por apresentarem maior capacidade de armazenamento e os seus dados serem sumarizados (El-Sappagh et al., 2011).

Inmon (2005), explica os atributos fundamentais de um DW, tais como:

- **Orientado ao assunto:** os dados são organizados por assuntos, tais como vendas, clientes ou produtos, contendo apenas informações relevantes para apoiar a tomada de decisão;
- **Integrado:** devem converter os dados de diferentes fontes, em um formato consistente, ou seja, desde extração até ao carregamento, os dados devem ser uniformizados e integrados;
- **Variante no tempo:** detêm uma qualidade temporal, pois devem armazenar dados históricos;
- **Não volátil:** após o carregamento dos dados no DW, estes permaneceram inalteráveis por norma, não podendo ser alterados ou atualizados. Qualquer alteração ou atualização, será assumida como um novo registo.

A causa de apoio à decisão de um DW, que até aos dias de hoje, é extramente valorizado, é realçado por Sá (2009) ao afirmar que, DW é um “repositório de registos informacionais integrados, oriundo de várias fontes internas ou externas da organização, onde estes registos representam eventos ou factos de um determinado período de tempo, que satisfazem os requisitos informacionais de uma organização”.

O significado do termo *Data Warehousing*, que é muitas vezes mal-entendida como *Data Warehouse*, explicado por Sá (2009), este, é equivalente a Sistemas de *Data Warehouse*, que significam, ambos, o processo de criação de um DW, que corresponde, por sua vez, ao conteúdo do Sistema de *Data Warehouse*.

Existem em 3 tipos principais de DW, sendo que cada um existe num determinado contexto e para um fim específico, podendo variar nas suas características. Turban et al. (2017) apresentam-nas, da seguinte forma:

- **Data Marts:** um DMT é relativamente menor que um DW e foca-se num assunto/departamento em particular. Este, pode ser dependente de um DW, isto é, a fonte de dados provém deste, ou independente quando é projetado para responder as necessidades de um departamento/assunto em específico, sem ligações a nenhum DW;
- **Armazenamento de Dados Operacionais:** é um tipo de base de dados, que normalmente é utilizada como uma área de “estágio”, antes da construção de um DW. O seu conteúdo é dinâmico, sofrendo constantes alterações/atualizações para dar resposta a objetivos/problemas de curto prazo;
- **Data Warehouse Organizacional:** é um DW de grande escala, utilizada em grandes organizações para armazenar uma elevada quantidade de dados, provenientes de várias fontes e padronizadas, facultando um melhor suporte à tomada de decisão.

2.4.2 Arquitetura de Sistemas de DW

Turban et al. (2017), definem as seguintes arquiteturas possíveis para um sistema de *Data Warehouse*:

- **Arquitetura de Três Camadas:** os sistemas operacionais contêm os dados e o software para aquisição de dados em um nível (ou seja, o servidor), o DW é outro nível, e o terceiro nível inclui o mecanismo de BI (ou seja, o servidor) e o terminal do cliente/utilizador. Os dados armazenados são processados duas vezes e depositados em uma base de dados multidimensional adicional, organizado para análise e apresentação multidimensional simplificado, ou replicado em *Data Marts*. Vantagem desta arquitetura, é a separação das funções do DW, removendo as restrições de recursos e facilitando a criação de DMT. A figura 5 exemplifica esta arquitetura;

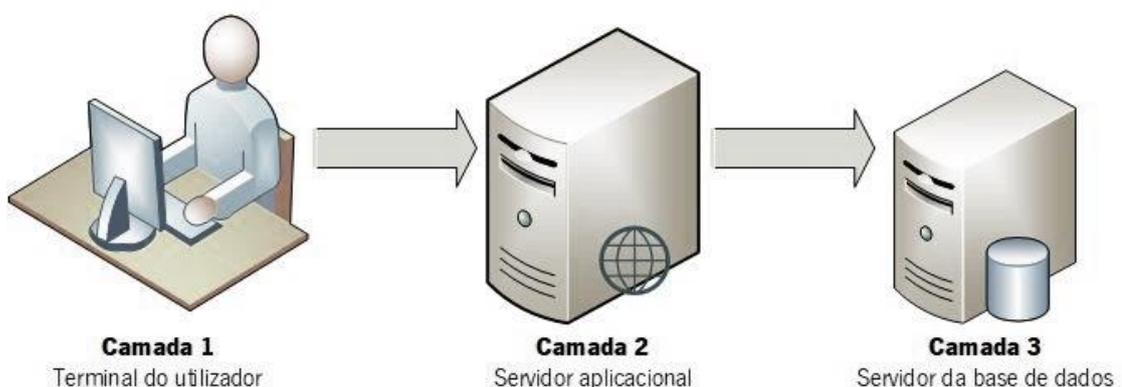


Figura 5- Arquitetura de três camadas (Adaptado de Turban et al., 2017)

- **Arquitetura de Duas Camadas:** na primeira camada o cliente/utilizador e na segunda camada, a aplicação do sistema de apoio à decisão, que por sua vez, funciona na mesma plataforma de hardware que o DW, tal como demonstra a figura 6. Esta, pode ser o mais

económico, no entanto, poderá acarretar problemas de desempenho ao lidar com grandes quantidades de dados;

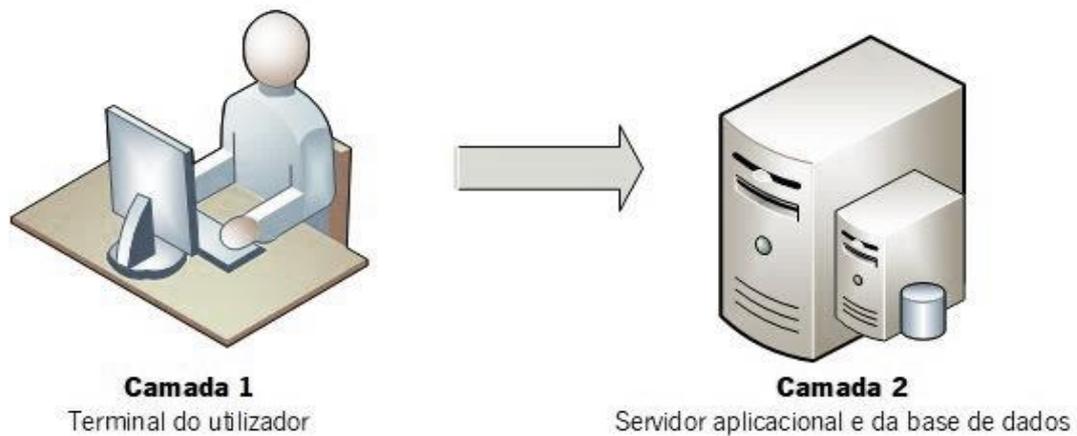


Figura 6 - Arquitetura de Duas Camadas (Adaptado de Turban et al., 2017)

- **Arquitetura Assente na Web:** representa um servidor cliente (browser), que através da internet, intranet ou extranet se conecta a um servidor Web, que por sua vez possibilita o acesso ao DW, demonstrada na figura 7. Esta arquitetura é de baixo custo, proporciona fácil acesso e independência das plataformas associadas.

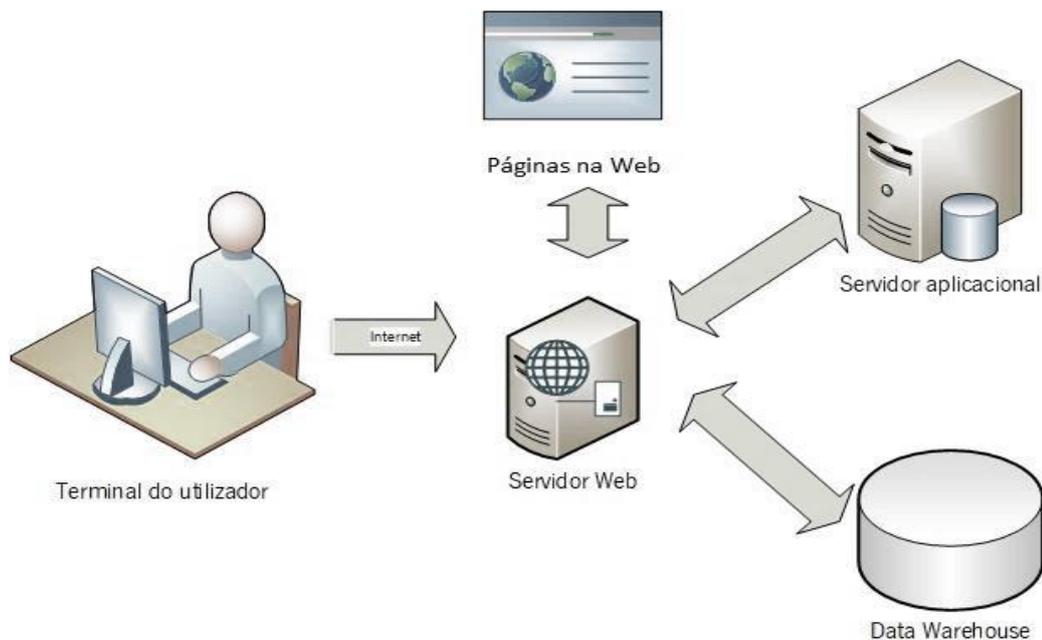


Figura 7 - Arquitetura assente na web (Adaptado de Turban et al., 2017)

2.4.3 Componentes do Sistema de DW

Um sistema de *Data Warehouse*, segundo Sá (2009, pg. 18) é constituído por diversos componentes, não se resumindo ao repositório do DW, sendo usualmente constituído por:

- **Fontes informacionais** – podendo ser internas ou externas, no primeiro caso, estes, resultam das atividades diárias das organizações, materializando-se nos repositórios de registos operacionais que irão alimentar o DW. No segundo caso, estes, originam de informações de mercado, de concorrentes, localizações geográficas, de cotações na bolsa;
- **Software para Extração, Transformação e Carregamento** – são ferramentas que permitem extrair os dados das diversas fontes informacionais, transformá-los e carregá-los no DW;
- **Repositórios** – esta componente é constituída por diversos elementos, pelo DW, vários DMT e os metadados (Secção 4.5);
- **Ferramentas de análise** – permitem aos utilizadores acederem à informação armazenada nos repositórios, podendo esse acesso, ser efetuado através de: linguagem de consulta estruturada (*SQL-Structure Query Language*), gerador de relatórios, *data mining*, ferramentas OLAP, entre outros.

2.4.4 Arquiteturas do *DW*

Num nível mais alto, a conceção da arquitetura do *Data Warehouse*, pode ser realizada nas categorias de DW Organizacional (figura 10) e *Data Marts*, segundo Golfarelli & Rizzi (Golfarelli & Rizzi, 2009, citado em Turban et al., 2017) . Na categoria DMT, existem duas opções, DMT independentes ou dependentes, tal como apresenta a figura 8 e 9. Para o primeiro caso, teremos um DMT orientado a um determinado assunto/departamento, contendo dados necessários para suportar e fornecer informações úteis aos utilizadores/gestores. No segundo caso, teremos DMT dependentes de um DW.

- **Categoria DMT - *Data Marts* Independentes:**

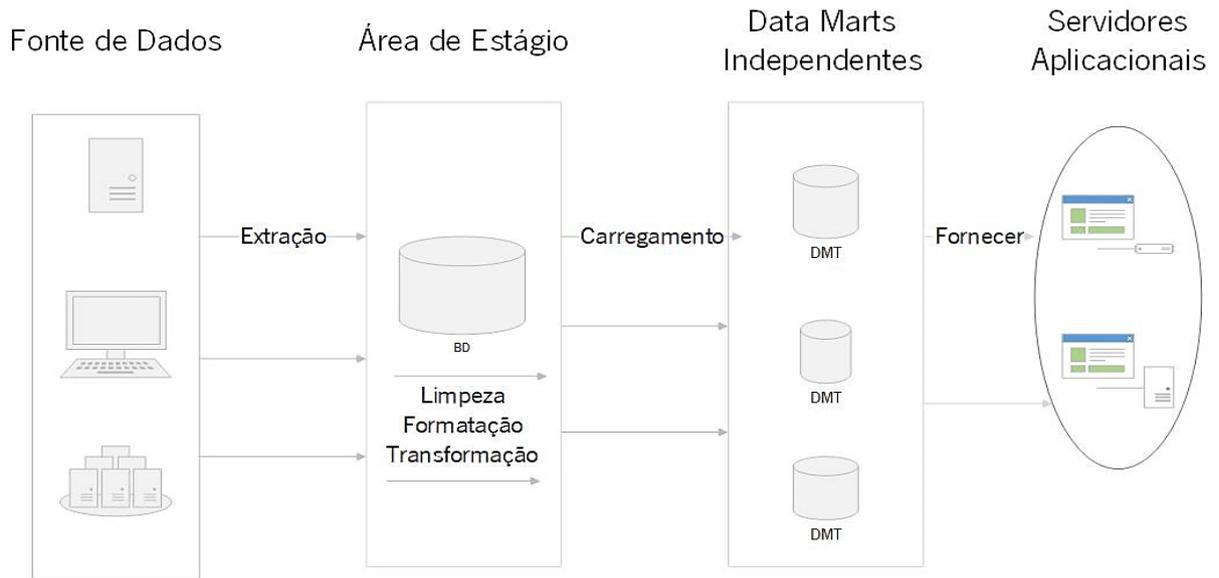


Figura 8-Arquitetura de Data Marts independentes (Adaptado de Turban et al., 2017)

- **Categoria DMT - *Data Marts Dependentes*:**

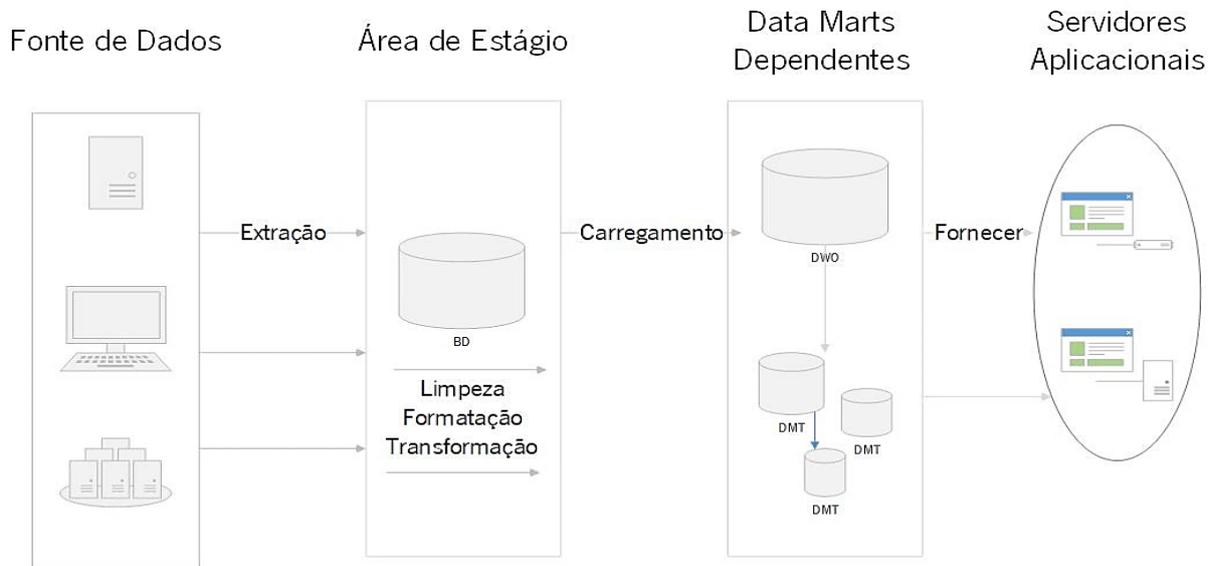


Figura 9 – Data Marts dependentes (Adaptado de Turban et al., 2017)

- **Categoria *Data Warehouse Organizacional*:**



Figura 10 - Arquitetura Data Warehouse organizacional (Adaptado de Turban et al. 2017)

2.4.5 Abordagens de Desenvolvimento

Os métodos frequentemente utilizados para no processo de construção de um DW, são os preconizados por Inmon e Kimball, respetivamente, a abordagem *Top-Down* e *Bottom-Up*.

- **Top-Down:** Inmon defende esta abordagem, pois esta, adapta-se as ferramentas tradicionais de base de dados relacionais ao desenvolvimento das necessidades de um DWO, requerendo inicialmente, uma análise global das necessidades da organização e posteriormente, o planeamento do desenvolvimento, conceção e implementação do DW (Turban et al., 2017). De uma forma geral, esta, possui duas etapas, de acordo com Malinowski e Zimány (Malinowski e Zimány, 2008, citado em Sá, 2009, pg. 69): a “primeira consiste em definir o esquema do conteúdo de todo o *Data Warehouse*. A segunda consiste em implementar *Data Marts* de acordo com as características particulares de cada departamento ou área de negócio”;
- **Bottom-Up:** esta abordagem, também conhecida por estratégia de DMT, é uma abordagem de "*Plan Big, Build Small*", isto é, realizar um bom e rigoroso planeamento, permitindo simplificar o processo de construção do DW. Segundo Sá (2009, pg. 69), esta abordagem “tem como objetivo modelar e construir esquemas dos conteúdos de cada *Data Mart*, tendo em conta as necessidades informacionais existentes. Os esquemas de cada *Data Mart* devem ser modelados com o objetivo de, posteriormente, serem unificados para assim se conseguir obter um esquema global de todo o *Data Warehouse*.”

A abordagem preconizada por Inmon, apresenta alguns benefícios para organização, na medida que permite ter uma visão global dos objetivos, mas implica muito tempo de desenvolvimento e custos elevados. Já a abordagem de Kimball, reduz significativamente o tempo de desenvolvimento, permitindo obter resultados mais tangíveis (Golfarelli & Rizzi, 2009).

2.4.6 Representação dos Dados no DW

Segundo Turban et.al., (2017, pg. 55), “Qualquer que seja a arquitetura do sistema de DW, a concepção da representação dos dados no DW tem sido baseada no conceito de modelação dimensional”. Turban et al., (2017), afirmam que, a modelação dimensional é um sistema baseado em consulta, que suporta o acesso a um grande volume de dados. Kimball & Ross, (2013), afirmam que, a modelação dimensional é aceite como uma das técnicas preteridas para apresentar os dados analíticos, por responder dois requisitos em simultâneo: **fornecer dados que sejam compreensíveis para os intervenientes do negócio; fornecer um desempenho de consulta rápida**. O armazenamento e representação dos dados num DW, deverá ser concebido, não só de forma a acomodar, mas também, a melhorar o processamento das consultas multidimensionais complexas (Turban et al., 2017).

Para a modelação dimensional, segundo Sá (2009), existem três estruturas principais para a representação dos dados, tais como:

- **Esquema em Estrela:** representa o estilo mais simples e o mais exercido na modelação dimensional. Como ilustrado na Figura 11, é constituído por uma tabela de factos e um conjunto de tabelas de dimensão. As tabelas de dimensão contêm atributos e apresentam-se desnormalizadas, facilitando a navegação na pesquisa. As tabelas de factos contêm um conjunto de chaves estrangeiras, uma por cada tabela de dimensão associada e apresentam-se normalizadas, ou seja, sem redundância;

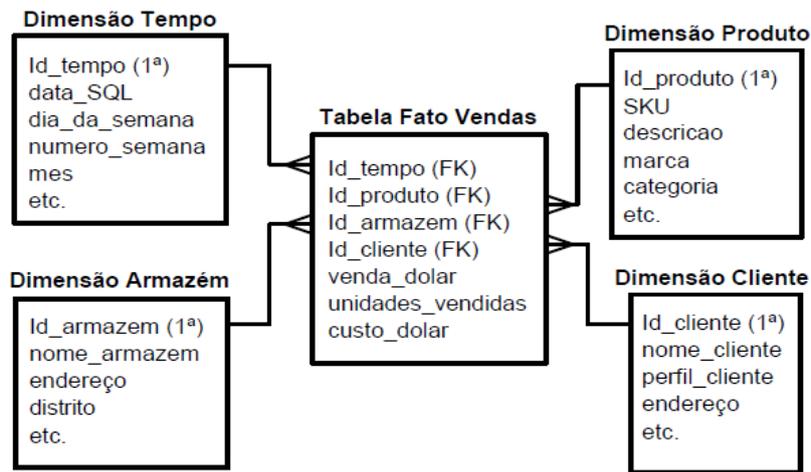


Figura 11 - Esquema em Estrela

- Esquema em Floco de Neve:** o esquema distingue-se do esquema em estrela, na medida em que, as tabelas de dimensões encontram-se normalizadas, com representação clara, da hierarquia de atributo nas dimensões, ilustrado na figura 12. Kimball & Ross, (2013), afirmam que, só se justifica a utilização deste modelo, na construção de um DW quando compensar a perda de rapidez e facilidade de acesso aos dados, relativamente ao espaço ocupado pelos mesmos;

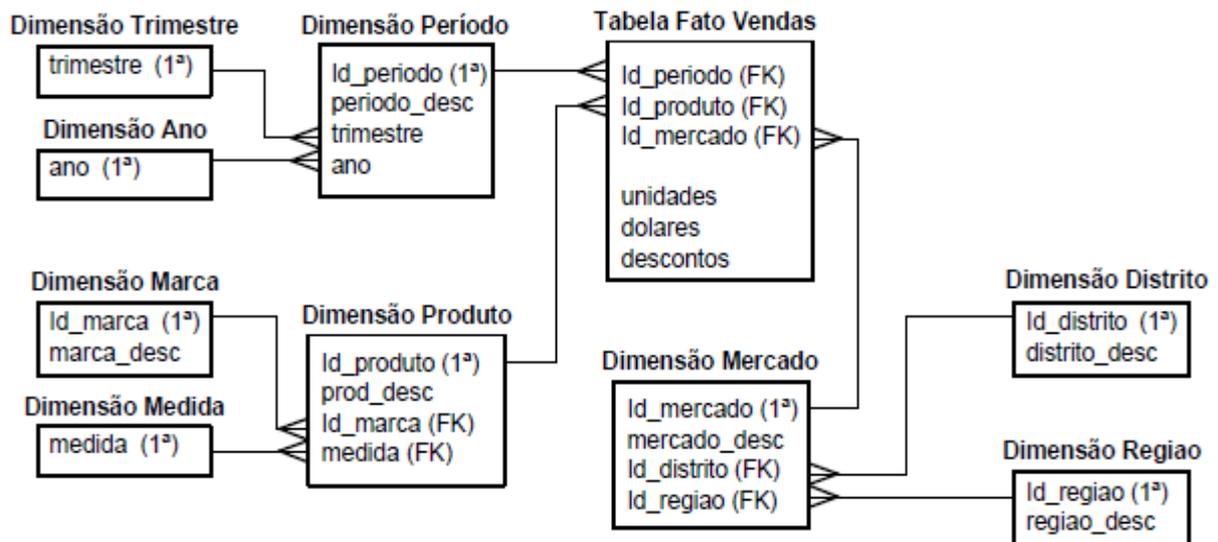


Figura 12 - Esquema em Floco de Neve

- Esquema em Constelação:** esta, ilustrada na figura 13, resulta da combinação de várias estruturas (em estrela, em floco de neve ou em constelação), garantindo a conformidade das dimensões existentes, de forma a permitir a sua partilha.

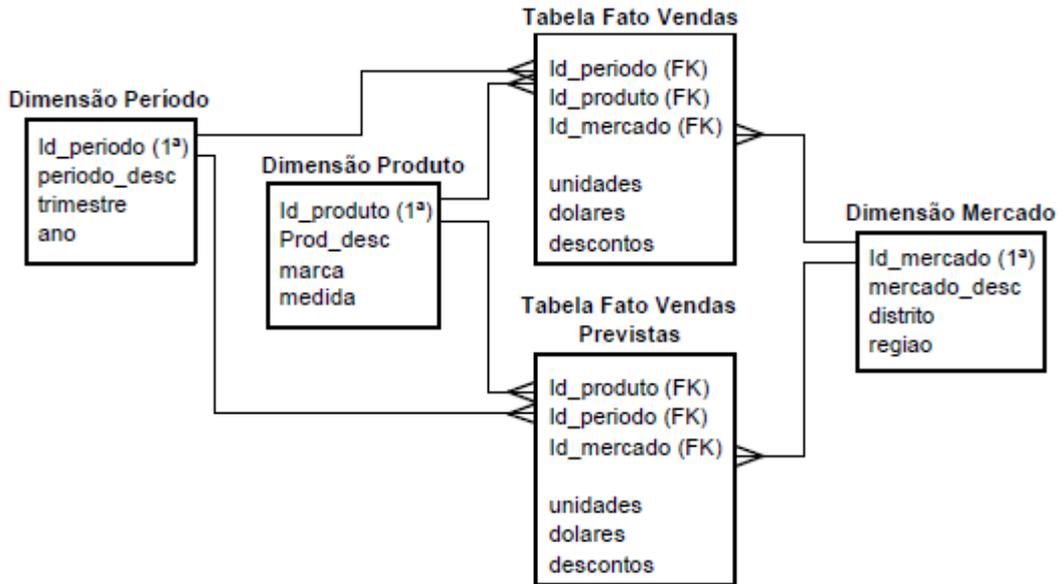


Figura 13 - Esquema em Constelação

2.4.7 Metadados

“Another consideration of metadata is that every technology in the business intelligence environment has its own metadata” Inmon (2005).

A importância dos metadados para um sistema de Business Intelligence, é crucial para a compreensão e controlo sobre os dados durante os vários processos que sofrem no sistema. De acordo com Sá (2009, pg. 19), os metadados “constituem informação sobre os registos informacionais armazenados no *Data Warehouse* e nos *Data Marts*, identificando a origem de cada registo informacional, o processo de transformação e limpeza que sofreu, e o seu significado”. Os autores Turban et al. (2017, pg. 38) e Kimball & Ross (2013), afirmam que, os metadados são “dados sobre os dados” (discordado por Loshin 2012, pg. 120, por não transpor a total essência dos metadados), estes, descrevem a estrutura e algum significado sobre os dados, contribuindo assim para o seu uso efetivo ou ineficaz.

De certa forma, os metadados são dicionários de dados, como é referido por Loshin (2012, pg. 120), com toda a informação necessária para os analistas do negócio. Sendo que embora os autores Turban et al. (2017, pg. 38), refiram que, os metadados podem ser diferenciados por: **sintáticos** (descrevem a sintaxe dos dados); **estruturais** (descrevem a estrutura dos dados) e **semânticos** (descrevem o significado dos dados, em um domínio específico). Loshin (2012) apresenta sete tipos de metadados, presentes na tabela seguinte.

Tabela 4 - Tipos de metadados (Adaptado de Loshin 2012)

Tipo	Descrição
Estrutural	Refere aos aspetos da forma e formato, os nomes das tabelas, o número de colunas e o tamanho e tipo de elemento dos dados.
Técnico	Engloba as informações necessárias para o desenvolvimento e execução de aplicações.
Referência	Representa os conjuntos partilhados das tabelas de referência, unidades de medidas associadas, enumerações e mapeamentos
Operacional	Corresponde aos registos de relatórios operacionais e dados estatísticas.
Informacional	Inclui: modelos conceptuais e lógicos de dados e as descrições das entidades

Tipo	Descrição
Analítica	Diz respeito à entrega de relatórios e análises, incluindo nomes de relatórios, colunas não-persistentes para relatórios ou tipos de modelos preditivos
Semântica	Envolve o glossário de termos de negócios que catalogam e mapeiam os conceitos de elementos dos dados.
Negócio	Correspondem a descrição da estrutura dos dados de relevância para os clientes e os utilizadores de negócio.

2.4.8 Análise e Acesso

➤ **Online Analytical Processing**

Para se efetuar a análise dos dados contidos num DW, é preciso ter em conta as técnicas de análise de dados existentes. De acordo com Turban et al. (2017), o processamento analítico online (OLAP – *Online Analytical Processing*) é o mais utilizado.

- **Online Analytical Processing (OLAP)** - Estes autores, Turban et al. (2017), definem OLAP, como uma abordagem para responder rapidamente às questões *ad-hoc*, executando consultas analíticas multidimensionais aos repositórios organizacionais. Também, afirmam que a estrutura principal da operação no OLAP, é baseado no conceito de “cubos”, que é uma estrutura de dados multidimensional para permitir consultas analíticas com maior rapidez e eficácia. Através da utilização do OLAP, o utilizador poderá navegar sobre os dados contidos no DW, efetuando pesquisas de um conjunto específico destes, esta navegação aglomera operações de *drill-down/up* (*agregação e desagregação*), *roll-up*, *slice* (*dividir*), *dice* (*separar*) e *pivot*.

Os servidores OLAP facultam de forma eficiente, a perspetiva multidimensional dos dados, às aplicações e utilizadores, permitindo também efetuar operações sobre os mesmos (Chaudhuri et al., 2011, pg. 90). As formas mais comuns de implementação desses servidores, são:

1. **MOLAP (*Multidimensional Online Analytical Processing*)**: são implementados em base de dados multidimensionais, o sistema é baseado em modelos lógicos *ad-hoc*, que podem ser utilizados para representar dados multidimensionais e executar operações sobre as mesmas (Golfarelli & Rizzi, 2009). Este tipo de arquitetura apresenta como principal vantagem o seu excelente desempenho e uma rápida indexação;

2. **ROLAP (Relational Online Analytical Processing):** são implementados em base de dados relacionais e geram consultas SQL (*Structured Query Language*) para efetuar as operações e cálculos requisitados pelo utilizador. Sendo que, as bases de dados relacionais não foram estruturadas tradicionalmente para lidar com os dados num formato dimensional. As vantagens desta arquitetura tem haver com a eficiência que proporciona na consulta e carregamento dos dados (Chaudhuri et al., 2011);

3. **HOLAP (Hybrid Online Analytical Processing):** é a divisão do armazenamento de dados em um MOLAP e uma base de dados relacional. A divisão dos dados pode ser feita de diferentes formas. Um deles é armazenar os dados detalhados numa base de dados relacional, como fazem os servidores ROLAP e pré-computar dados agregados no MOLAP. Outro método é armazenar dados mais recentes no MOLAP, para obter acesso rápido e dados antigos no ROLAP (Chaudhuri et al. 2011).

➤ **Dashboard e Scorecards**

Dashboards ou painéis, são métodos práticos de apresentar e aceder à informação armazenada num *Data Warehouse*. Este, é um GUI (*Guide User Interface*), que apresenta uma quantidade limitada de dados num formato legível. Os indicadores de performance, são os principais conteúdos apresentados nos *dashboards*, contendo informação abrangente sobre a organização, exemplificado na figura 14. Isto, permite aos utilizadores do negócio visualizar, consultar e analisar as operações da organização (Golfarelli & Rizzi, 2009).



Figura 14 - Exemplo de Dashboard (retirado de Hart & Saxton, 2018)

Scorecards, também são formas de acesso aos dados, fornecendo arranjos visuais da informação, que é consolidada e disposta em uma única tela para que a informação possa ser interpretada em um numa visão e a estratégia global de uma organização, em um conjunto de objetivos, medidas, metas e iniciativas financeiras e não financeiras inter-relacionadas, tal como demonstra a figura 15 (Sauter, 2011; Turban et al., 2017).

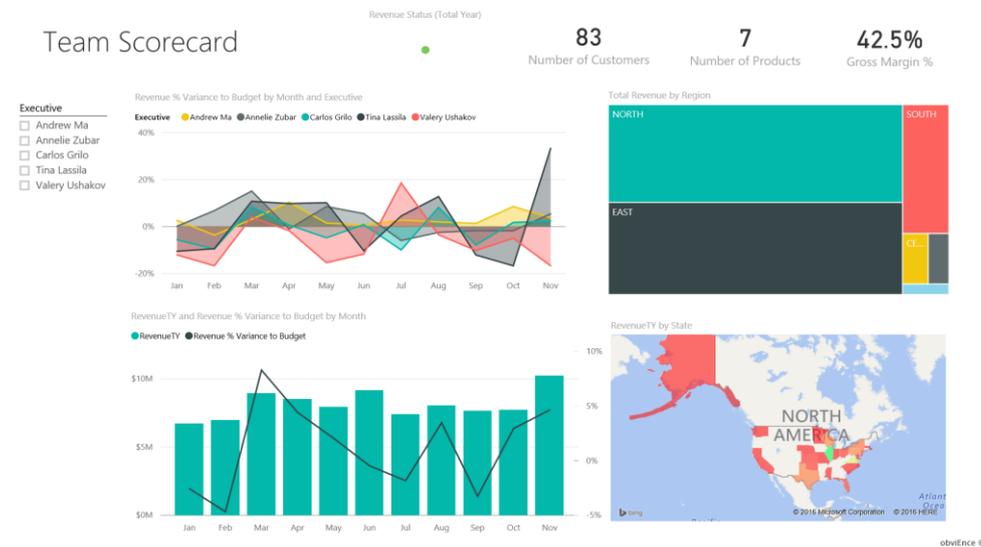


Figura 15 - Exemplo de scorecard (retirado de Hart & Saxton, 2018)

Segundo Kimball & Ross (2013, pg. 423), “alavancar portais e *dashboards* informativos personalizáveis, são as estratégias dominantes para disseminar o acesso”. Mas, Golfarelli & Rizzi (2009, pg. 30), alertam que, os *dashboards* devem ser vistos como um *add-on* eficaz do DW, e não como o objetivo final deste.

➤ **Reports**

Aplicações de *Reporting (relatórios)* possibilitam aceder aos dados de um *Data Warehouse* e construir modelos preditivos para ajudar a responder questões como: “Quais os clientes com maior probabilidade de responder à minha nova campanha de marketing via email?”. Golfarelli & Rizzi (2009, pg. 28), afirmam que, “é uma abordagem de acesso, através de consultas de dados multidimensionais ou *layouts* de histogramas e gráficos, apresentando como vantagens: a apresentação de informação fiável e correta; a separação de operações de transação e análise, melhorando a performance do DW”.

➤ **Data Mining**

Data Mining (DM) é uma das outras formas de acesso aos dados de um DW, sendo que este, é o processo de descoberta de padrões em grandes quantidades de dados, através da aplicação de técnicas/modelos matemáticos, estatísticos e de inteligência artificial (Turban et al. 2017). O uso de DM nas aplicações de BI, tem sido considerado como uma das características essenciais dos sistemas de BI (Chen et al., 2012). Pelo que, tem permitido aos utilizadores do negócio, analisar os dados ao detalhe, incluindo a capacidade de construir modelos preditivos. Esta abordagem, passa por selecionar um subconjunto de dados do DW, efetuar análises de dados sofisticados, no subconjunto de dados selecionados, identificando características estatísticas de relevância e, em seguida, construir modelos preditivos (Chaudhuri et al., 2011).

2.5 Extract, Transform & Load

2.5.1 Conceito

De acordo com Ferreira et al. (2017), o “ETL (Extração, Transformação e Carregamento) e as ferramentas de limpeza de dados consomem um terço do orçamento num projeto de DW, podendo, no que respeita ao tempo de desenvolvimento de um projeto de DW, chegar a consumir 80% desse valor.” El-Sappagh et al. (2011), afirmam que o processo ETL, é normalmente uma combinação complexa de tecnologias e processos, que consomem uma grande percentagem de esforço, do desenvolvimento de um DW e exige as capacidades técnicas de várias equipas de trabalho, tais como: os analistas de negócio, os programadores, os designers da base de dados. E não sendo um processo linear, isto é, um processo a ser executado uma vez, pois as fontes informacionais são alteradas e o DW atualizado periodicamente, sendo que o processo ETL deverá ser adaptável às mudanças.

Na figura 16, é apresentado um processo ETL genérico, resumido e estruturado com as suas fases principais:

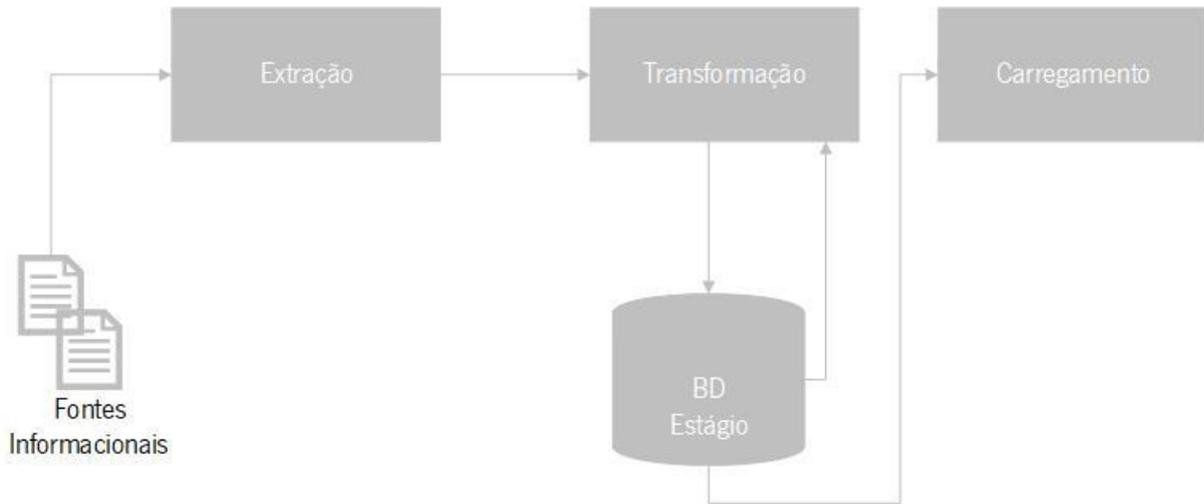


Figura 16 - Representação de um processo ETL genérico (Adaptado de El-Sappagh 2011)

Os autores Turban et al. (2017, pg. 47), definem o ETL como, o “processo de extração (capturar os dados das várias fontes informacionais), transformar (converter o formato dos dados extraídos para o formato adequado, para ser inserido no DW ou DMT), carregar (efetuar o carregamento dos dados no DW)”. Chaudhuri et al. (2011), realçam o papel crucial, que as ferramentas ETL desempenham na garantia da qualidade dos dados e no carregamento eficiente de grandes quantidades de dados no DW. Outros autores como Shilakes & Tylman (Shilakes & Tylman, 2009, citado em El-Sappagh et al., 2011) referem que, as ferramentas ETL pertencem a uma categoria de ferramentas especializadas, com a tarefa de lidar com os problemas da homogeneidade, limpeza, e transformação do DW.

2.5.2 Fases do Processo ETL

O sistema ETL consiste em três etapas funcionais consecutivas: extração, transformação e carregamento (El-Sappagh et al., 2011, pg. 93).

- **Extração:** esta etapa é responsável pela extração dos dados das fontes informacionais. Cada fonte tem as suas características específicas, que precisam ser geridas, de modo que a extração ocorra de forma eficaz. O processo precisa eficientemente integrar os sistemas que possuem diferentes plataformas tecnológicas, tais como, sistemas de gestão de base de dados, sistemas operativos, e diferentes protocolos de comunicação (El-Sappagh et al., 2011);
- **Transformação:** Os autores Golfarelli & Rizzi (2009), defendem que esta etapa deverá dividir-se em duas fases distintas: Limpeza e Transformação. No entanto, de acordo com El-Sappagh et al. (2011), a segunda etapa diz respeito à transformação dos dados, esta, tem como objetivo a limpeza e transformação dos dados extraídos, de forma a obter a acuidade dos mesmos, isto

é, dados corretos, completos, consistentes e sem ambiguidade. Inclui atividades como: definição da granularidade das tabelas de factos e das tabelas de dimensões; o esquema DW (em estrela ou floco de neve), factos derivados, dimensões com alteram-se lentamente, tabelas de factos sem factos;

- **Carregamento:** nesta etapa, os dados extraídos e transformados são inseridos nas estruturas dimensionais, acedidas pelos utilizadores finais e pelas aplicações do sistema. A etapa de carregamento engloba, tanto o carregamento das tabelas de dimensão, assim como o carregamento das tabelas de factos (El-Sappagh et al., 2011). Golfarelli & Rizzi (2009), associam duas técnicas a esta etapa, respetivamente o refrescamento e a atualização. O refrescamento trata-se de substituir os dados antigos, por novos dados, e a atualização, é a adição de novos registos, sem modificar ou eliminar os dados antigos.

De acordo com Turban et al. (2017, pg. 49), a execução extensiva de um processo de ETL, poderá ser um sinal de má gestão dos dados e falta de uma estratégia coerente de gestão dos dados. Quando os dados são geridos corretamente como um bem organizacional, os esforços de ETL são significativamente menores e os dados redundantes são completamente removidos. Isto traduz-se em poupanças na manutenção, maior eficiência no desenvolvimento do projeto e melhor qualidade dos dados. Pelo que, é crucial efetuar as escolhas apropriadas em termos de tecnologias e ferramentas para suportar o desenvolvimento e manutenção do processo ETL.

2.6 Visão Crítica

A revisão de literatura possibilitou reconhecer e dissecar, os conceitos associados ao tópico e os vários autores na área de *Business Intelligence*. Proporcionou constatar que, existe uma convergência entre os vários autores, na classificação de sistemas de *Business Intelligence*, como um agregador de informação competitiva, com o intuito de acrescentar valor nas decisões dos gestores/intervenientes no negócio. Este valor proativo dos sistemas BI, é transversal, sendo que, existe um vasto conjunto de tecnologias associadas que influenciam a implementação dos sistemas de BI e o ambiente em que é inserido.

É interessante também, realçar a importância dos metadados, que muitas vezes passam ao lado nas organizações, são poucas as organizações que sabem como conceber e implementar uma estratégia para os metadados e poucas capturam os vários tipos de metadados, sendo que, os que capturam,

não lhe proporcionam o devido uso (Loshin 2012; Turban et al. 2017). Concordando com a afirmação de Wang (2009, pg. 1845), sobre a vantagem da implementação de modelo de metadados integrado (comum e compartilhado), como dissipador da lacuna, entre a falta de representação dos metadados semânticos e sintáticos, nos vários componentes do sistema de DW.

No campo de ação deste capítulo, foi também pertinente identificar os vários perfis/tipos de utilizadores do negócio presentes numa organização e sua importância para o ambiente de sistemas de BI, no desígnio de discernir os requisitos para ferramentas, assim como, desmitificar os significados inerentes à Extração e Exploração de Dados, devido à divergência existente na categorização das ferramentas existentes, pois estas podem estar presentes em diferentes camadas na arquitetura dos sistemas de BI atuais. Pelo que houve muita dificuldade em encontrar trabalhos publicados, no contexto de avaliação de ferramentas enquadradas com objetivo deste trabalho.

3 ABORDAGEM METODOLÓGICA

A realização desta dissertação seguiu diretrizes referenciais no âmbito de Sistemas de Informação, de forma a sustentar uma base sólida para a investigação e materialização do mesmo. Pelo que este capítulo apresenta a metodologia empregue para a conclusão deste trabalho.

3.1 Metodologia de Investigação

Para a elaboração desta dissertação, a metodologia de investigação aplicada, tratou-se da *Design Science Research (DSR)*, sendo que esta, consiste num paradigma de *problem-solving* (resolução de problemas). Segundo Denning (1997) e Tsichritzis (1998) (citado em Hevner et. all, 2004), esta metodologia, procura criar inovações que definam ideias, métodos, capacidades técnicas e artefactos através dos quais a análise, o design, a implementação e o proveito dos Sistemas de Informação, possam ser concretizados de forma eficiente e efetiva.

Peppers et al. (2008) propôs a metodologia de DSR (DSRM) para a produção e apresentação de pesquisas *Design Science* em Sistemas de Informação. Este esforço contribuiu para a pesquisa de SI, fornecendo uma *framework* para a efetuação exitosa de DSR e um modelo mental para sua apresentação. O processo inclui seis etapas: identificação e motivação do problema; definição dos objetivos para uma solução, design e desenvolvimento; demonstração; avaliação; e comunicação, representada na figura 17.

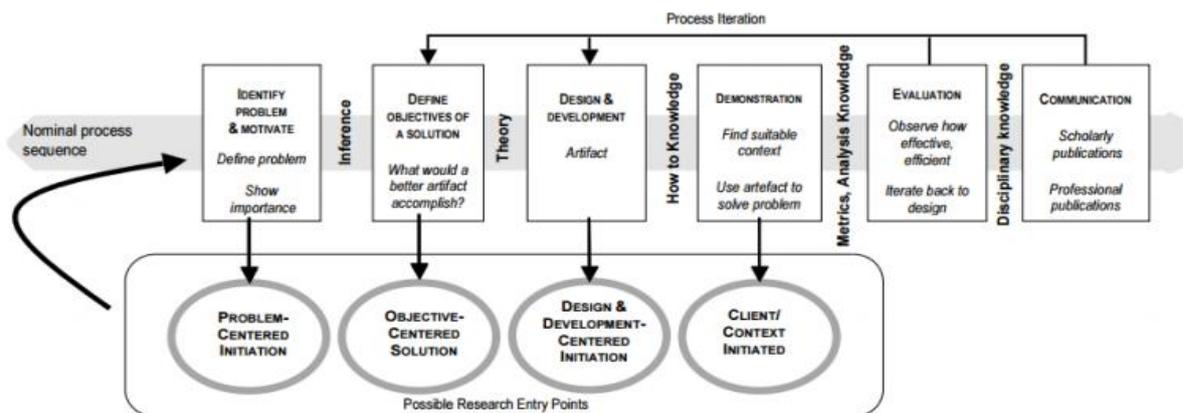


Figura 17 - Metodologia DSR (Adaptado de Peppers et al., 2008)

Hevner & Chatterjee (2010), descrevem as fases do DSR como:

1. A primeira etapa da metodologia diz respeito à **Identificação do problema e motivação**, este, consiste em definir o problema que será utilizado para desenvolver um artefacto que pode efetivamente fornecer uma solução, pode ser útil ir ao detalhe do problema conceptualmente, para que a solução possa capturar a sua complexidade. Justificar o valor de uma solução implica duas coisas: motivar o investigador e a comunidade científica a irem ao encontro da solução entendendo o raciocínio do associado ao problema. Os recursos necessários para esta atividade incluem o conhecimento do estado da arte do problema e a importância da sua solução.
2. A segunda etapa, corresponde com a **Definição dos objetivos da solução**, onde pretende-se identificar os objetivos para solução, a partir da definição do problema e inferir o viável. Os objetivos podem ser quantitativos, para medir o desempenho de uma solução através de indicadores; ou qualitativos, quando se deseja interpretar algum fenómeno específico, através da sua observação, descrição, compreensão e o significado do seu comportamento. Os recursos necessários para esta atividade incluem o conhecimento e a eficácia, do estado dos problemas e das soluções atuais, se existirem.
3. A terceira etapa refere-se à fase de **Conceção e desenvolvimento**. Esta, consiste na criação do artefacto, podendo ser: modelos, métodos, instâncias, “novas propriedades de recursos informacionais e/ou sociais”. Conceptualmente, um artefacto de pesquisa de conceção pode ser qualquer objeto projetado, no qual uma contribuição de pesquisa está incorporada no projeto. Esta atividade inclui a determinação da funcionalidade desejada do artefacto e a sua arquitetura, e de seguida criação do artefacto real. Os recursos necessários para passar de objetivos à conceção e desenvolvimento, incluem o conhecimento da teoria que pode servir como base para uma possível solução.
4. A quarta etapa incide na **Demonstração**, onde pretende-se apresentar o uso do artefacto para resolver uma ou mais instâncias do problema. Este passo pode envolver o seu uso na experimentação, simulação, estudo de caso, prova, ou outra atividade apropriada. Os recursos necessários para a demonstração, incluem o conhecimento efetivo de como usar o artefacto para resolver o problema.

5. A quinta etapa passa pela **Avaliação**, que envolve analisar como o artefacto suporta uma solução para o problema definido. Esta atividade envolve a comparação dos objetivos de uma solução com os resultados observados, através da utilização do artefacto na demonstração. Isto requer conhecimento de métricas e técnicas de análise relevantes. Dependendo da natureza do problema e do artefacto, a avaliação pode assumir diversas formas. Poderia incluir itens para servir de comparação da funcionalidade do artefacto com os objetivos da solução, medidas de desempenho quantitativo, como orçamentos ou itens produzidos, resultados de pesquisas de satisfação, feedback de clientes ou simulações. Poderia incluir medidas quantitativas de desempenho do sistema, como o tempo de resposta ou disponibilidade. Conceptualmente, tal avaliação poderia incluir qualquer tipo de prova empírica apropriada ou lógica. No final desta atividade, os investigadores podem decidir avaliar novamente a etapa três e tentar melhorar a eficácia do artefacto, ou continuar com a comunicação e deixar melhorias futuras para projetos subsequentes. A natureza do ponto de investigação pode determinar se tal iteração é viável ou não.

6. A sexta e última etapa consiste na **Comunicação**, do problema e a sua importância, o artefacto, a sua utilidade e inovação, o rigor da sua conceção e a sua efetividade para investigadores e outro público relevante, como profissionais na área, quando apropriado. Em pesquisas académicas, os investigadores podem usar a estrutura desse processo para estruturar um artigo, assim como, a estrutura nominal de um processo de pesquisa empírica (definição do problema, revisão de literatura, desenvolvimento de hipóteses, coleta de dados, análise, resultados, discussão e conclusão) consiste numa estrutura comum para trabalhos empíricos de investigação. A comunicação requer conhecimento da cultura disciplinar.

3.2 Aplicação da Metodologia

Para o desenvolvimento deste trabalho teve-se como diretriz a metodologia DSR, sendo esta aplicada da seguinte forma:

- A primeira etapa consistiu na definição do problema e motivação estabelecidos na secção 1.1;
- A segunda etapa que diz respeito à definição de um conjunto de objetivos para responder ao problema definido, corresponde ao conjunto de objetivos (principal e específicos) definidos na secção 1.2;
- A terceira etapa correspondente à conceção e desenvolvimento do artefacto. Aqui, pretende-se provar que é possível identificar e explorar as ferramentas orientadas à atividade de Extração e

Exploração de Dados (EED), de acordo com as necessidades de um determinado grupo de utilizadores, neste caso, os mais dependentes, tal como demonstra nos capítulos 4 e 5;

- A quarta e quinta etapa, circunscrevem-se à avaliação das ferramentas de Extração e Exploração de Dados (EED) através da tipologia de catalogação e avaliação elaborada. Para demonstrar que ao determinar um conjunto de requisitos pertencentes a um conjunto de utilizadores específicos, o artefacto (a tipologia) é capaz de determinar a ferramenta de EED que melhor se adequa. Apresentada no capítulo 5;
- Para última etapa, referente à comunicação, espera-se a apresentação dos principais resultados obtidos do estudo deste problema, demonstrando a importância da elaboração deste trabalho, a sua utilidade, e o contributo que proporcionará, para a comunidade científica. Serão apresentados os resultados obtidos e a prova de que é possível dar resposta à questão de investigação e elaboração de um artigo científico para submissão em uma conferência internacional.

4 AMBIENTE DE EXPERIMENTAÇÃO

Este capítulo apresenta o desenvolvimento do ambiente de experimentação para a avaliação e experimentação das ferramentas de Extração e Exploração de Dados (EED), assim como, a descrição e justificação das ferramentas utilizadas, quer para o desenvolvimento do contexto de experimentação, quer para a realização das atividades de EED. Para além de agregar todos fatores que estiveram envolvidos na determinação dos requisitos das ferramentas de EED.

4.1 Ferramentas de Desenvolvimento

As principais ferramentas utilizadas durante este projeto, podem ser agrupadas em dois grupos, **Desenvolvimento do Contexto de Experimentação (DCE)** e **Extração e Exploração de Dados (EED)**, apresentadas na tabela 5. Nesta secção apenas serão descritas as ferramentas de DCE.

Tabela 5 - Ferramentas de desenvolvimento

DCE	EED
<i>VMware Workstation Player 14</i>	<i>Microsoft Excel</i>
	<i>Microsoft Excel & Power View</i>
	<i>Microsoft Excel & Power Pivot</i>
<i>MS Windows Server 2016</i>	<i>Tableau Public</i>
<i>MS SQL Server 2016</i>	<i>Power BI Desktop</i>
<i>MS SharePoint Server 2013 Enterprise Edition</i>	<i>Performance Point</i>

4.1.1 *VMware Workstation Player 14*

O software *VMware Workstation Player*, antigamente *VMware Player*, é um software proprietário de virtualização para sistemas operativos Microsoft Windows e Linux, desenvolvido pela empresa *VMware, Inc.* O software permite criar ambientes virtuais através da manipulação dos recursos da máquina local, sendo um *hypervisor* do tipo dois, ou seja, esta depende da disponibilidade da máquina local, para providenciar os serviços do ambiente virtual, que são partilhados em simultâneo com outras aplicações.

O *VMware Workstation Player* está disponível com uma licença gratuita, que se revelou crucial na escolha deste software. A figura a seguir ilustra bem o exemplo de uma arquitetura do *VMware Workstation Player*.

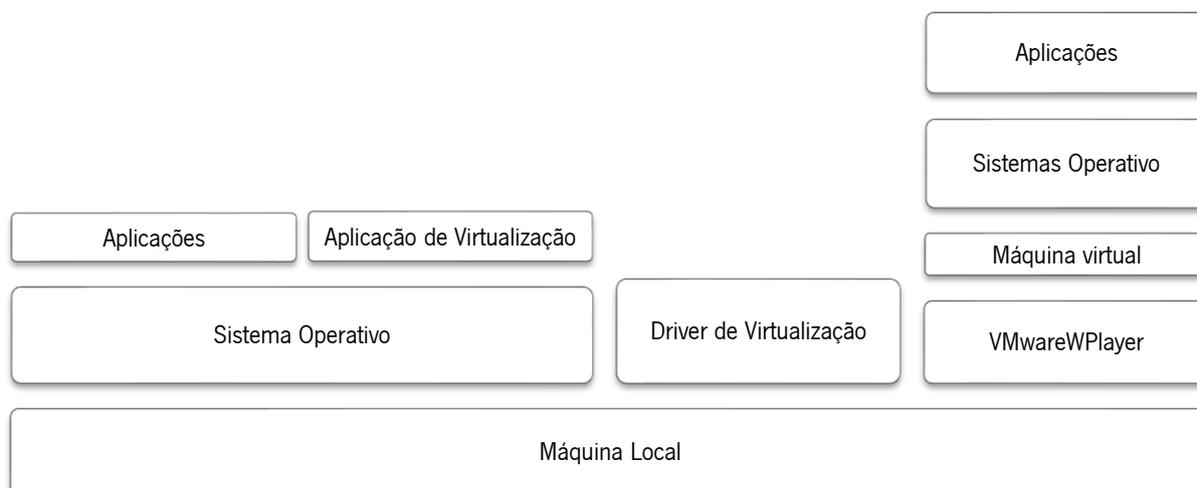


Figura 18 - Arquitetura do VMware W. Player (Elaborado pelo autor)

O ambiente de experimentação desenvolvido está armazenado numa máquina virtual montada através *VMware Workstation Player 14*, com os recursos de hardware descritos na tabela 6.

Tabela 6 - Características da máquina virtual

Requisitos de Hardware	
Sistema Operativo	Windows Server 16
Processador	Intel Core I5-3337U CPU @ 1.80 GHz
Memória	8 GB
Armazenamento	200 GB

4.1.2 MS Windows Server 2016

Windows Server é um sistema operativo (SO) para servidores desenvolvido pela Microsoft, para ser executado em cenários locais e na *cloud*. Como SO, o Windows Server proporciona dois modos de instalação: GUI (*Graphical User Interface*) ou CLI (*Command Line Interface*). Este sistema operativo

inclui funcionalidades propícias a uma boa gestão do servidor em termos de segurança e para ambiente de desenvolvimentos aplicativos, pelo que, esses benefícios influenciaram na escolha deste SO.

4.1.3 *MS SQL Server 2016*

O *Microsoft SQL Server* é um sistema de gestão de bases de dados relacional desenvolvido pela Microsoft. Como servidor de base de dados, a sua função principal consiste em armazenar e recuperar dados conforme solicitado por outras aplicações de software. *SQL Server* lançada pela primeira vez em 1989 como SS 1.0 e permitiu a entrada da Microsoft no mercado de base de dados a nível organizacional, concorrendo contra líderes como Oracle ou IBM (Szmajdych, 2016). Durante a década seguinte, a Microsoft desenvolveu e lançou a versão 7.0 que já suportava as soluções OLAP.

Atualmente, o SS 2016 (empregue no DCE) comparativamente às versões anteriores, fornece suporte para escalonamento de infraestruturas de *cloud computing* e acresce funcionalidades como a integração de R ou *polybase*. A escolha da versão do SS recaiu sobre a 2016 *Enterprise*, por ser disponibilizada de forma gratuita pela universidade e também a versão mais recente.

4.1.4 *MS SharePoint Server 2013*

O *SharePoint* evoluiu dos projetos *Office Server* e *Tahoe*, durante o ciclo de desenvolvimento do *Office XP*. A primeira versão lançada foi o *Microsoft SharePoint Portal Server 2001*, seguido do *Microsoft SharePoint Team Services 2002*. Seguido pelo *SharePoint 2003*, que incluiu o *Windows SharePoint Services 2.0* (licença livre) e o *Microsoft SharePoint Portal Server 2003* (versão comercial). Em seguida, surgiu uma nova versão em 2007, que incluía o *Windows SharePoint Services 3.0* (licença gratuita) e o *Microsoft Office SharePoint Server 2007* (extensão comercial). O próximo na fila esteve o *Microsoft SharePoint Foundation 2010* (gratuito) com o *Microsoft SharePoint Server 2010* (extensão comercial para *Foundation*) e o *SharePoint Enterprise 2010* (extensão comercial para Servidores) (Lauer et. al, 2012). De momento, existe o *Microsoft SharePoint Foundation 2016* e o *Microsoft SharePoint Server 2016*.

É uma tecnologia de portais, concebido, desenvolvido e testado com a estratégia *SaaS*⁷ (*Software as Service*) da Microsoft, fornecendo um grande conjunto de ferramentas para organizar conteúdos, gerir documentos, partilhar conhecimento, criar ambientes de colaboração, identificando informações e pessoas. Uma tecnologia que pode ser usado para fornecer portais de intranet/extranets, redes sociais, sites, pesquisas empresariais e serviços de *Business Intelligence*. Também possui integração de sistemas, integração de processos e recursos de automação do fluxo de trabalho (Mendes, 2018).

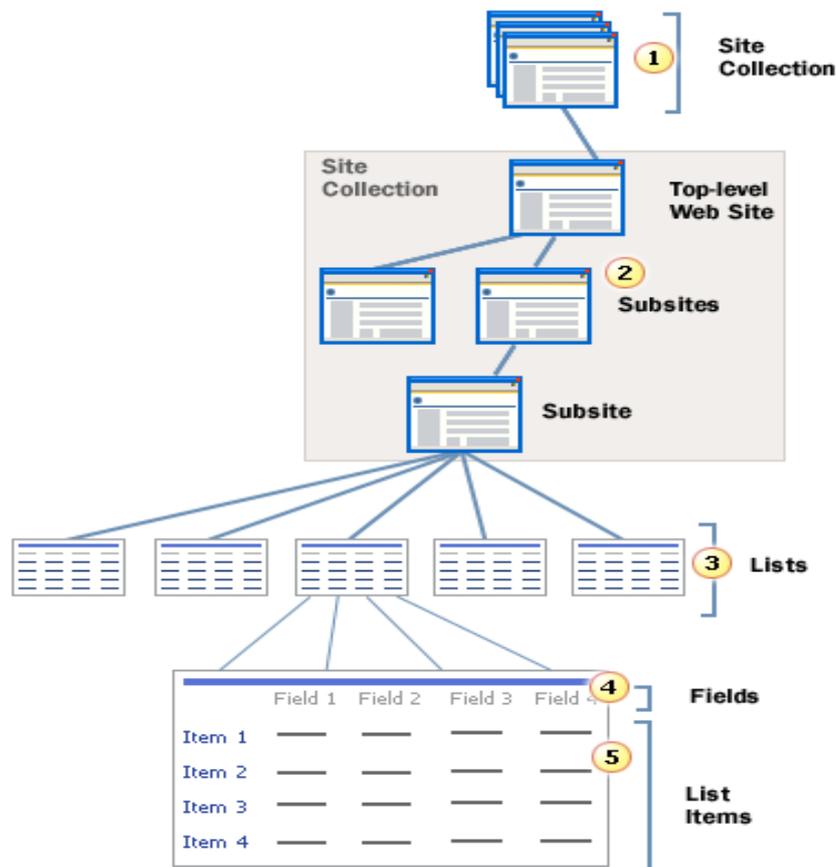


Figura 19 - Arquitetura de um site no SharePoint (Retirado de Mendes, 2018)

A arquitetura de um site *SharePoint*, segue uma estrutura hierárquica semelhante a ilustrada na figura 19, separada e organizada por funcionalidades. “No topo, está a coleção de sites que permite isolar configurações entre duas coleções diferentes. As configurações que estão isoladas são: definição de metadados; navegação das páginas; grupos de permissões; *templates* e estilos das páginas. Dentro de

⁷ <https://azure.microsoft.com/pt-br/overview/what-is-saas/>

uma coleção há um site principal que pode conter páginas ou *subsites*. Cada *subsite* pode conter, tal como o anterior, um conjunto de páginas e *subsites*. Deste modo é possível montar uma estrutura no *SharePoint* que reflita a estrutura da organização que necessita de uma ferramenta de colaboração. Dentro de cada site da estrutura é possível criar listas. As listas têm como objetivo guardar informação que podem ser documentos, ficheiros, endereços ou valores (texto, números, datas, etc.)” (Mendes, 2018).

Para a concretização desta dissertação, optou-se pelo *SharePoint Server Enterprise Edition*, por ser a versão mais recente, disponibilizada gratuitamente pela universidade, para além de agregar serviços de BI, não disponíveis em outras edições. A estrutura da implementação efetivada neste trabalho, é designada por *Farm Solution* ou *Full-Trust Solution*, sendo o *SharePoint* instalado num único servidor. A figura 20 ilustra o exemplo da arquitetura desse tipo de implementação.

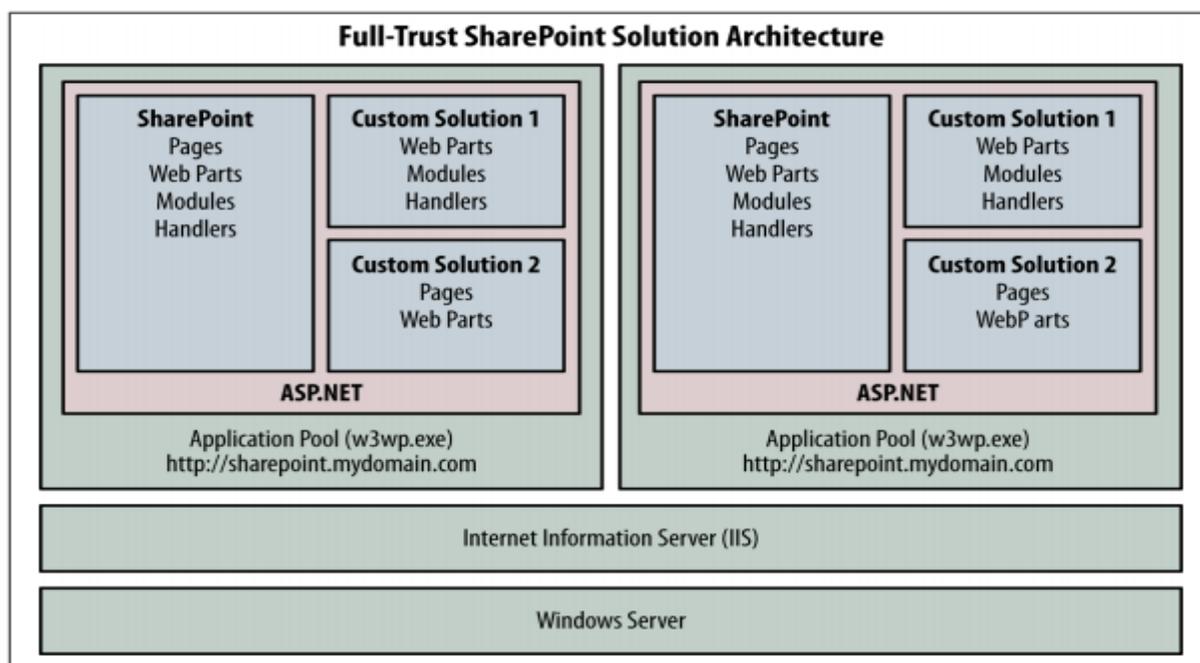


Figura 20 - Arquitetura da implementação "Farm Solution" (retirado de Feldman & Himmelstein, 2013)

4.2 Ferramentas de Extração e Exploração de Dados

Para determinar a utilidade de uma determinada ferramenta perante os sistemas de *Business Intelligence* atuais, é necessário compreender o ambiente/contexto envolvente. Pois, dependendo do negócio, estes sistemas podem estar organizados de várias formas.

O tipo de utilizador avaliado é transversal nos vários contextos de implementação dos sistemas de BI. De acordo com classificação apresentada no Capítulo 2 (secção 2.3.4), foi preciso identificar um conjunto de utilizadores finais, neste caso, os utilizadores casuais, para efetuar o levantamento das ferramentas existentes no mercado e a avaliação dos mesmos, como demonstra a tabela 7.

Tabela 7 - Classe dos utilizadores (Adaptado de Eckerson, 2013)

Classe	Necessidades
Visualizador	Qualidade dos dados, relatórios, <i>scorecards</i> e <i>dashboard</i> intuitivos e estáticos; permitir a colaboração interativa com analistas (real-time); exportação de dados para ficheiros pdf ou csv;
Navegador	Rápida execução das operações sobre os dados (drill-down/up, slice, dice, etc) presentes em gráficos e tabelas dinâmicas; permitir a colaboração interativa com analistas (real-time); exportação de dados para ficheiros pdf, word e csv;
Explorador	Rapidez na execução de consultas <i>ad-hoc</i> e na extração dos dados; apresentação de dados relevantes para análises dos indicadores do negócio; permitir a colaboração interativa com analistas (real-time); exportação de dados para ficheiros pdf, word e csv;

4.2.1 Requisitos das Ferramentas

Para a avaliação e catalogação das ferramentas, foi necessário determinar um conjunto de requisitos/características, de modo que fossem cumpridas as exigências identificadas dos utilizadores casuais. Para que isto seja cumprido, o modo com que os utilizadores interagem com os sistemas de BI, para analisar, manipular e partilhar os dados, pode variar de acordo com as infraestruturas, capacidades e o grupo de utilizadores.

Para efeitos deste trabalho, foi considerado o termo **estilo**, como a forma/intenção de análise do utilizador presentes em Lauer et. al (2012), onde são destacados os seguintes estilos, por enquadrarem com os grupos de utilizadores identificados neste trabalho:

- **Análise e Relatórios *Self-Service*** – análise e relatórios *Self-Service* capacitam utilizadores que não possuem habilidades específicas, para explorar os dados e manipular as informações disponíveis. Estes podem identificar padrões, elaborar relatórios e partilhá-los com os colegas de trabalho, sem depender dos técnicos de Tecnologia de Informação (TI);
- ***Mashups de dados Self-Service*** - Os *mashups* de dados são criados combinando dados de várias origens, por ex.: *Data Warehouse*, relatórios, planilhas, sites e assim por diante. Depois dos dados serem reunidos, os analistas vão querer obter *insights* mais profundos manipulando e analisando as informações. Estes utilizadores precisam ser capazes de trabalhar independentemente dos departamentos de TI, pois os *mashups* de dados geralmente devem ser executados rapidamente, podendo ser utilizados apenas para uma única análise;
- ***Scorecarding*** – O *Scorecarding* é um estilo que descreve visualizações altamente resumidas com indicadores-chave de desempenho (KPIs) referentes a objetivos predefinidos, como um *Balanced Scorecard*.

De acordo com os estilos de interação mencionados anteriormente, foram selecionados os requisitos destes que permitissem avaliar as ferramentas de Extração e Exploração de Dados (EED) em linha com as classes dos utilizadores identificados. Pelo que, foram considerados os seguintes:

➤ **Análise e Relatórios *Self-Service*:**

- Permite a criação rápida de relatórios e visualizações de dados;
- Permite que os utilizadores façam atividades *Data Driven* (*drill-down*, *drill-up* filtragem e *pivot*) e apresentem os dados em tabelas, gráficos e outras visualizações;
- Permite a utilização de interfaces *drag-and-drop* para *design* de relatórios;
- Permite que os utilizadores trabalhem independentemente dos técnicos de TI;
- Permite a análise de grandes quantidades de dados;
- Permite ser utilizado sem que os utilizadores estejam familiarizados com linguagens de consulta, como *SQL (Structured Query Language)* e *MDX (Multidimensional Expressions)*;
- Permite que os relatórios sejam exportados em múltiplos formatos com PDF, Excel, Microsoft Word e HTML.

➤ ***Mashups de dados Self-Service*:**

- Permite o acesso a fontes de dados empresariais, departamentais, externas e pessoais;

- Permite que os utilizadores trabalhem independentemente dos técnicos de TI;
- Permite a análise de grandes quantidades de dados;
- Permite que os dados sejam atualizados automaticamente;
- Permite a criação avançada de relatórios usando *SQL*, *MDX* ou outras linguagens de consulta.

➤ **Scorecarding:**

- Permite a análise de grandes quantidades de dados;
- Permite que os dados sejam atualizados automaticamente;
- Permite que as informações sejam distribuídas regularmente;
- Permite elaborar relatórios que forneçam pontuações para o desempenho de uma organização, departamento ou indivíduo;
- Permite elaborar relatórios que apresentem os objetivos de negócio e *KPIs* discriminados hierarquicamente, podendo ser filtrados para auxiliar na identificação de *outliers*;
- Permite agregações e *KPIs* personalizados;
- Permite KPIs que direcionam a navegação para outros estilos de relatório no *dashboard*.

As ferramentas elegidas se alinham com as necessidades dos utilizadores casuais e os estilos de interação, sendo que, a facilidade de utilização e integração com os sistemas atuais de BI, licenças facultadas pela Universidade do Minho e a interoperabilidade das ferramentas da Microsoft, influenciaram na seleção das mesmas. Posto isto, foram selecionadas as seguintes ferramentas:

1. Microsoft Excel

O *Excel*⁸ é um programa de software da *Microsoft*, que faz parte do *Microsoft Office*. Foi desenvolvido e lançado pela primeira vez, em 30 de setembro de 1985, este, é capaz de criar e editar planilhas que são guardadas com uma extensão de arquivo *.xls* ou *.xlsx*. A utilização do *Excel* engloba cálculos baseados em células, tabelas dinâmicas e várias ferramentas gráficas. Por exemplo, com uma planilha, podemos criar um orçamento mensal, rastrear despesas comerciais ou classificar e organizar grandes quantidades de dados.

O *Microsoft Excel* é a ferramenta de *Business Intelligence* mais amplamente utilizada no mundo para o estilo de relatório e análise *self-service*. Praticamente, todos os utilizadores estão familiarizados de

⁸ <https://www.computerhope.com/jargon/e/excel.htm>

algum modo com o *Excel*, sendo um ambiente confortável para analisar dados e desenvolver relatórios *ad-hoc*. O *Excel* pode ser usado para explorar dados e criar relatórios em conexão como por exemplo: o *SQL Server Analysis Services*, o *SQL Server* e outras fontes de dados.

Durante a fase experimentação, foi utilizado o *Microsoft Excel 2013*, conjuntamente com o *Excel Services*, que foi introduzido no *SharePoint 2007* com objetivo de permitir aos utilizadores carregarem e compartilharem, o conteúdo de uma pasta de trabalho do *Excel*, com outros membros do site da equipa. Também, o *Power View* e o *Power Pivot*, que são um conjunto de tecnologias da MS, funcionando com suplementos do *MS Excel 2013*.

O *Excel Services* é um componente nativo do *SharePoint 2017 Enterprise Edition*, apresentando uma versão HTML, interativa da experiência de trabalho nativa do *Excel* e é construída em par com o servidor do mecanismo de cálculo do *Excel*, na máquina local. Isto, permite que os utilizadores interajam com as pastas de trabalho do *Excel* diretamente em seu navegador da *Web*, sem precisar descarregar a aplicação.

2. Power View

O *Power View* é uma experiência interativa de exploração, visualização e apresentação de dados que proporciona a conceção de relatórios *ad-hoc*. O *Power View* é um suplemento do *Microsoft Excel 2013* e do *Microsoft SharePoint Server 2017* e 2013 como parte do suplemento do *SQL Server 2012 Service Pack 1 Reporting Services* para o *Microsoft SharePoint Server Enterprise Edition*.

O *Power View* oferece suporte à análise de *self-service*, capacitando os utilizadores a desempenhar tarefas de relatório e análise, que antes exigiam habilidades de um desenvolvedor de relatórios profissional. Um utilizador pode facilmente analisar dados do *PowerPivot* ou do *Analysis Services* através de técnicas de *drag-and-drop*. Os relatórios do *Power View* são criados em um ambiente familiar e de fácil gestão.

3. Power Pivot

O *Power Pivot* é um suplemento do *Excel* que permite executar uma análise de dados eficiente e criar modelos de dados sofisticados. Com o *Power Pivot*, pode-se facilmente extrair, interagir e manipular

dados, exercer *mashups* de grandes volumes de dados de várias fontes, ultrapassando o limite do *Excel*, que é cerca de 1 milhão de linhas por folha. O *Power Pivot* pode ser dimensionado para milhões e até centenas de milhões de linhas, permitindo executar análises de informações rapidamente e compartilhar *insights* com facilidade.

4. Performance Point

Performance Point é um serviço do *Microsoft SharePoint Server 2017*, sendo uma ferramenta de gestão de desempenho utilizada para criar *dashboards* e *scorecards*. Ao fornecer uma ferramenta flexível e fácil de aplicar, sendo que para este projeto, serviu para criar *dashboards*, *scorecards*, relatórios e indicadores-chave de desempenho (KPIs), o serviço *Performance Point*, pode ajudar os utilizadores a tomar melhores decisões de negócios, alinhadas aos objetivos e à estratégia da empresa. Oferecendo a capacidade de extrair vários tipos de visualizações de dados em um único *dashboard*, podendo interagir com relatórios do *Reporting Services*, gráficos baseados em cubos OLAP, mapas de desempenho e diagramas do *Microsoft Visio*.

5. Power BI Desktop

O *Power BI* é um serviço de análise do negócio da Microsoft, onde a versão *Desktop*⁹ proporciona análises visuais através da criação intuitiva de relatórios, com uma interface de funcionalidades *drag-and-drop* para obter o conteúdo desejado, representado de uma forma fluida. O ambiente de trabalho permite ainda manipular, visualizar e partilhar *online* o trabalho produzido, recorrendo facilmente à publicação num site do *SharePoint Server*, na versão paga e ao *Power BI online*, na versão normal.

O *Power BI Desktop* simplifica e agiliza o processo de conceção e criação de repositórios, para sistemas de BI e elaboração de relatórios, permitindo analisar o negócio através da apresentação de informações sobre a organização. Com a facilidade de conectar-se a diversas fontes de dados, simplificando a preparação de dados. Podendo criar *dashboards* personalizados com uma vista única sobre o negócio.

6. Tableau Desktop

⁹ <https://powerbi.microsoft.com/en-us/features/>

Tableau é um software baseado na *Web*, cuja versão *open-source* é o *Tableau Desktop*¹⁰. Este, permite criar visualizações interativas e integrá-las em um *website*, publicá-las no *Tableau Public Gallery* ou partilhá-las. A aplicação gera um código que pode ser incorporado em qualquer página *Web* e, variadas formas de representação dos dados, tais como gráficos, tabelas e mapas. Possibilita a conexão com diferentes fontes de dados. Ao estabelecer ligação a uma fonte de dados, identifica o papel de cada campo, determinando a existência de dimensões e factos, e o modelo tridimensional da estrutura onde os dados são armazenados.

4.3 Cenário de Negócio

O cenário de negócio foi estabelecido de acordo com o *Data Warehouse* utilizado, *Adventure Works*¹¹. Este, foi elaborado pelo projeto de “código aberto” *CodePlex*, pertencente à Microsoft, para o seguinte cenário fictício: “A *Adventure Works Cycles* é uma grande empresa multinacional de produção e distribuição de componentes de bicicletas de metal para mercados comerciais na América do Norte, Europa e Ásia. A sede da *Adventure Works Cycles* é *Bothell*, Washington, onde a empresa emprega 500 trabalhadores. Além disso, a *Adventure Works Cycles* emprega várias equipas regionais de vendas em toda a sua base de mercado. De momento, a *Adventure Works Cycles* quer ampliar sua quota de mercado, através da extensão da disponibilidade de produtos através de um site externo e redução do custo de vendas, reduzindo os custos de produção.” (Duncan, 2018).

4.4 Arquitetura Tecnológica

Através da integração de várias tecnologias e ferramentas, de modo a providenciar os dados do *Data Warehouse - Adventure Works*, foi disponibilizado o acesso aos dados necessários para testar as ferramentas de Extração e Exploração de Dados EED. Na figura 21, pode se verificar a visão global do ambiente elaborado para dar suporte ao DW, ao cubo OLAP, às visualizações e à apresentação dos dados.

¹⁰ <https://www.tableau.com/products/desktop>

¹¹ <https://github.com/Microsoft/sql-server-samples/releases/tag/adventureworks>

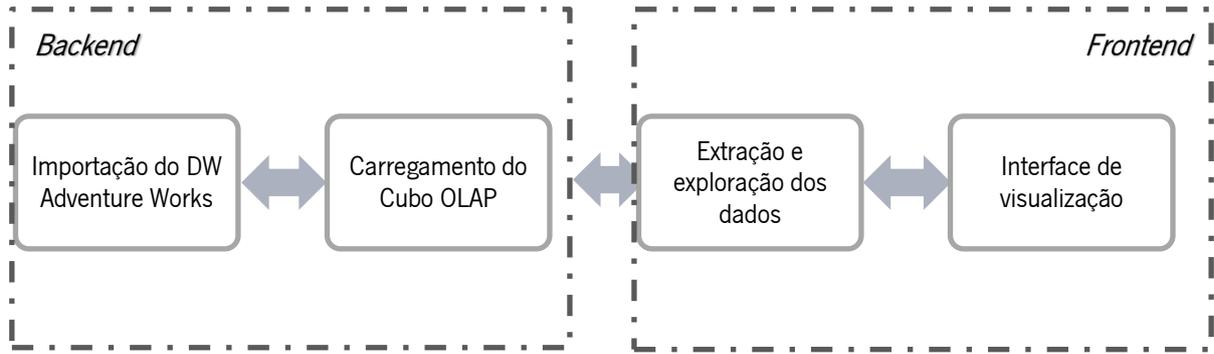


Figura 21-Arquitetura Tecnológica – Visão Global (Elaborado pelo autor)

Para a montagem do ambiente de experimentação, ilustrado na figura 22, foi utilizado o *VMware Workstation Player*, para a instalação de um servidor virtual (SV) com o sistema operativo *Windows Server 2016*, onde foram instalados: o *SQL Server* no intuito de alojar o *Data Warehouse*, através *Database Engine* e o cubo OLAP através do *Analysis Service*, sendo o cubo carregado através do *Visual Studio*; as ferramentas de Extração e Exploração de Dados (*Excel*, *Power View*, *Power Pivot*, *Tableau*, *Power BI*, *Performance Point*); e o *SharePoint Server* para disponibilização e apresentação dos dados. Assim, foram estabelecidas as condições de *hardware* necessárias para testar as ferramentas de EED, num contexto *ad-hoc*.

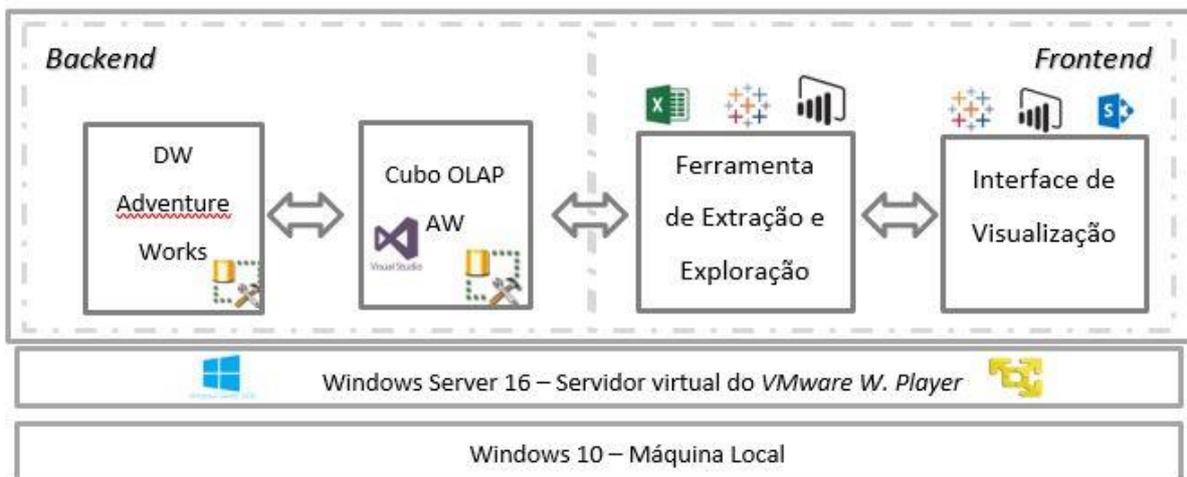


Figura 22 – Arquitetura tecnológica (Elaborado pelo autor)

4.5 Modelo Dimensional de Dados

Para experimentação no contexto deste projeto, apenas foram selecionados as dimensões e tabelas de factos mais relevantes, para a realização das atividades de Extração e Exploração de Dados (EED), neste caso, do negócio de vendas (*Internet Sales*, *Reseller Sales* e *Product Inventory*). Pode-se consultar nos pontos a seguir, a descrição e as respetivas ligações do *Data Warehouse*, assim como, o modelo com as ligações entre as entidades.

4.5.1 Modelo Dimensional

O DW *Adventure Works* é constituído por três tabelas de factos (*InternetSales*, *ResellerSales* e *ProductInventory*) com uma granularidade à transação e onze dimensões, num esquema de constelação, como se pode observar na figura 23 (e em Anexo I, detalhadamente).

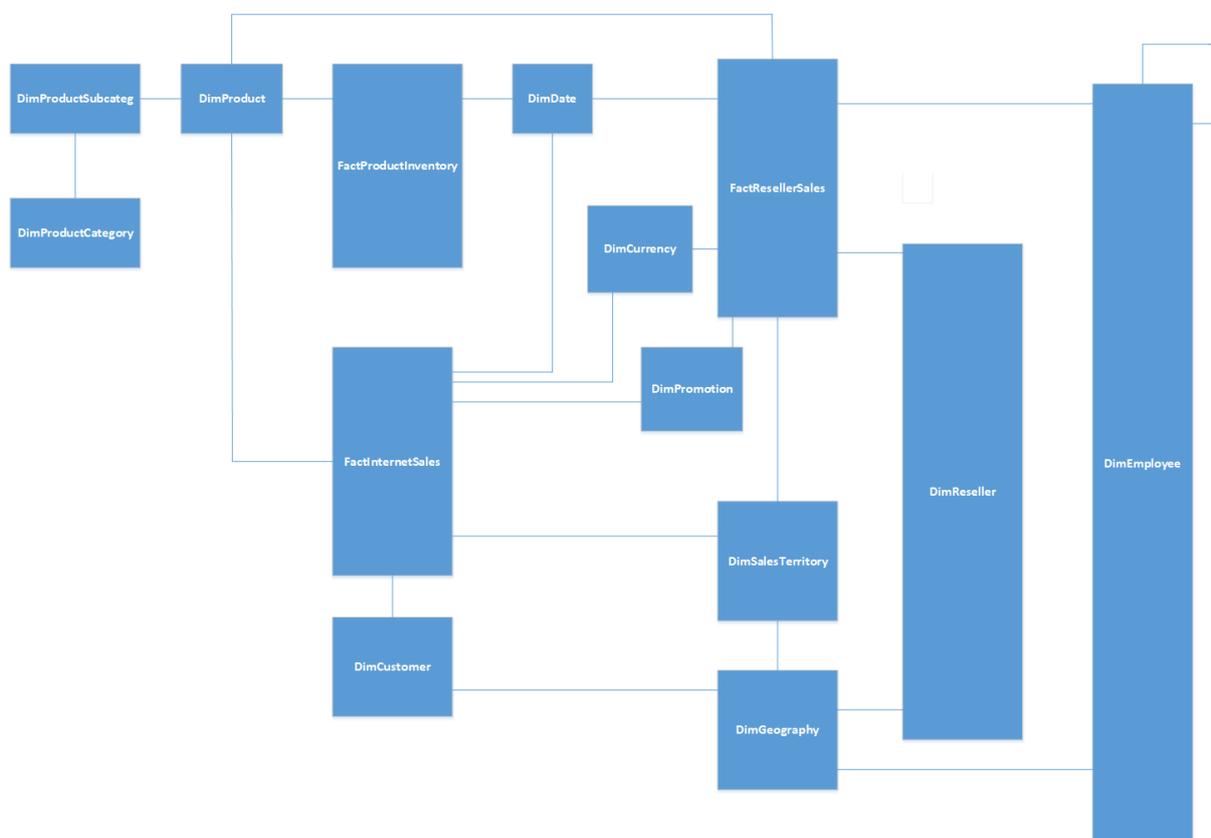


Figura 23 – Modelo dimensional vista simples (Elaborado pelo autor)

4.5.2 Descrição das Tabelas de Factos

A Tabela 9 lista as tabelas de factos do DW, os atributos e o número de registos de cada uma. Identifica também, a granularidade em causa para cada tabela de factos e um resumo da informação armazenada. A descrição completa dos atributos pode ser consultada nos anexos.

Tabela 9- Tabelas de Factos

Tabelas de Factos			
Nome da tabela (atributos)	Granularidade	Registos	Informação
FactInternetSales (<i>ProductKey, OrderDateKey, DueDateKey, ShipDateKey, CustomerKey, PromotionKey, CurrencyKey, SalesTerritoryKey, SalesOrderNumber, SalesOrderLineNumber, RevisionNumber, OrderQuantity, UnitPrice, ExtendAmount, UnitPriceDiscountPct, DiscountAmount, ProductStandardCost, TotalProductCost, SalesAmount, TaxAmt, Freight, CarrierTrackingNumber, CustomerPONumber, OrderDate, DueDate, ShipDate</i>)	Daily (ao dia)	60398	Guarda todos os dados correspondentes às transações de vendas efetuadas através da internet, como por ex.: o tipo de produto vendido, data da venda, data de expedição, preço, quantidade vendida, local da venda, etc.
FactResellerSales (<i>ProductKey, OrderDateKey, DueDateKey, ShipDateKey, ResellerKey, EmployeeKey, PromotionKey, CurrencyKey, SalesTerritoryKey, SalesOrderNumber, SalesOrderLineNumber, RevisionNumber, OrderQuantity, UnitPrice, ExtendedAmount, UnitPriceDiscountPct, DiscountAmount, ProductStandardCost, TotalProductCost, SalesAmount, TaxAmt, Freight, CarrierTrackingNumber, CustomerPONumber, OrderDate, DueDate, ShipDate</i>)	Daily (ao dia)	60855	Guarda todos os dados correspondentes às vendas efetuadas através dos revendedores, como por ex.: o tipo de produto vendido, data da venda, data de expedição, preço, quantidade vendida, local da venda, etc.

Nome da tabela (atributos)	Granularidade	Registos	Informação
FactProductInventory (<i>ProductKey, DateKey, MovementDate, UnitCost, UnitsIn, UnitsOut, UnitsBalance</i>)	Daily (ao dia)	776286	Guarda os dados sobre os produtos que fazem parte do inventário/armazém, disponibilizando informações sobre as quantidades existentes, quantidades saídas/entradas, data de transação, custo por unidade, etc.

4.5.3 Descrição das Dimensões

A tabela 8 lista as dimensões do DW, os atributos, número de registos e um resumo da informação armazenada em cada dimensão.

Tabela 8 - Tabela das dimensões

Nome da tabela (atributos)	Registos	Informação
DimCustomer (<i>CustomerKey, GeographyKey, CustomerAlternateKey, Title, FirstName, MiddleName, LastName, NameStyle, BirthDate, MaritalStatus, Suffix, Gender, EmailAddress, YearlyIncome, TotalChildren, NumberChildrenAtHome, EnglishEducation, SpanishEducation, FrenchEducation, EnglishOccupation, SpanishOccupation, FrenchOccupation, HouseOwnerFlag, NumbersCarsOwned, AddressLine1, AddressLine2, Phone, DateFirstPurchase, CommuteDistance</i>)	18484	Guarda todos os dados relativos aos clientes do negócio.
DimGeography (<i>GeographyKey, City, StateProvinceCode, StateProvinceName, CountryRegionCode, EnglishCountryRegionName, SpanishCountryRegionName, FrenchCountryRegionName, PostalCode, SalesTerritoryKey, IpAddressLocator</i>)	655	Guarda dados sobre as cidades.
DimSalesTerritory (<i>SalesTerritoryKey, SalesTerritoryAlternateKey, SalesTerritoryRegion, SalesTerritoryCountry, SalesTerritoryGroup, SalesTerritoryImage</i>)	11	Guarda dados das regiões territoriais dos clientes, tais como, continente, país, região, etc.

Nome da tabela (atributos)	Registos	Informação
<i>DimProduct</i> <i>(ProductKey, ProductAlternateKey, ProductSubcategoryKey, WeightUnitMeasureCode, SizeUnitMeasureCode, EnglishProductName, SpanishProductName, FrenchProductName, StandardCost, FinishedGoodsFlag, Color, SafetyStockLevel, ReorderPoint, ListPrice, Size SizeRange, Weight, DaysToManufacture, ProductLine, DealerPrice, Class, Style, ModelName, LargePhoto, EnglishDescription, FrenchDescription, ChineseDescription, ArabicDescription, HebrewDescription, ThaiDescription, GermanDescription, JapaneseDescription, TurkishDescription, StartDate, EndDate, Status)</i>	606	Guarda dados relativos aos produtos, tais como, o tipo, a categoria, peso, nome, etc.
<i>DimProductCategory</i> <i>(ProductCategoryKey, ProductCategoryAlternateKey, EnglishProductCategoryName, SpanishProductCategoryName, FrenchProductCategoryName)</i>	4	Guarda dados relativos a categoria dos produtos.
<i>DimProductSubCateg</i> <i>(ProductSubcategoryKey, ProductSubcategoryAlternateKey, EnglishProductSubcategoryName, SpanishProductSubcategoryName, FrenchProductSubcategoryName, ProductCategoryKey)</i>	37	Guarda dados relativos a subcategoria dos produtos, pois existem produtos que para além da categoria, necessitam de uma especificação mais detalhada em termos do material.
<i>DimDate</i> <i>(DateKey, FullDateAlternateKey, DayNumberOfWeek, EnglishDayNameOfWeek, SpanishDayNameOfWeek, FrenchDayNameOfWeek, DayNumberOfMonth, DayNumberOfYear, WeekNumberOfYear, EnglishMonthName, SpanishMonthName, FrenchMonthName, MonthNumberOfYear, CalendarQuarter, CalendarYear, CalendarSemester, FiscalQuarter, FiscalYear, FiscalSemester)</i>	3652	Guarda todos os dados correspondentes à data, tais como, o ano, o dia do mês, dia da semana, nome do mês em vários idiomas, etc.
<i>DimPromotion</i> <i>(PromotionKey, PromotionAlternateKey, EnglishPromotionName, SpanishPromotionName, FrenchPromotionName, DiscountPct, EnglishPromotionType, SpanishPromotionType, FrenchPromotionType, EnglishPromotionCategory, SpanishPromotionCategory, FrenchPromotionCategory, StartDate, EndDate, MinQty, MaxQty)</i>	16	Guarda todos os dados relativos as campanhas promocionais dos produtos vendidos pelo negócio.

Nome da tabela (atributos)	Registos	Informação
<i>DimReseller</i> <i>(ResellerKey, GeographyKey, ResellerAlternateKey, Phone, BusinessType, ResellerName, NumberEmployees, OrderFrequency, OrderMonth, FirstOrderYear, LastOrderYear, ProductLine, AddressLine1, AddressLine2, AnnualSales, BankName, MinPaymentType, MinPaymentAmount, AnnualRevenue, YearOpened)</i>	701	Guarda os dados relativos os revendedores.
<i>DimEmployee</i> <i>(EmployeeKey, ParentEmployeeKey, EmployeeNationalIDAlternateKey, ParentEmployeeNationalIDAlternateKey, SalesTerritoryKey, FirstName, LastName, MiddleName, NameStyle, Title, HireDate, BirthDate, LoginID, EmailAddress, Phone, MaritalStatus, EmergencyContactName, EmergencyContactPhone, SalariedFlag, Gender, PayFrequency, BaseRate, VacationHours, SickLeaveHours, CurrentFlag, SalesPersonFlag, DepartmentName, StartDate, EndDate, Status, EmployeePhoto)</i>	296	Guarda dados relativos aos funcionários do negócio/empresa
<i>DimCurrency</i> <i>(CurrencyKey, CurrencyAlternateKey, CurrencyName)</i>	105	Guarda todos os dados sobre o tipo de moeda (EUR, USD, CVE) utilizadas nas vendas.

4.6 Resumo do Contexto Elaborado

Para formalizar a avaliação das ferramentas de EED, o *stack* aplicacional elaborado, permite o acesso aos dados de um *Data Warehouse* de tamanho considerável, facultando informações concretas de negócio. Pelo que o objetivo abrange testar as ferramentas, que poderão proporcionar uma melhor resposta aos utilizadores casuais.

Foi criado um ambiente de experimentação e de visualização, de forma o aceder a um DW, facultando um conjunto de relatórios que permitem a utilização intuitiva e autónoma pelos utilizadores, através da aplicação de diferentes filtros. Para criar estes ambientes, foi necessário definir e importar diversos componentes, tal como, o esquema multidimensional. A tabela 9 lista as principais características do ambiente de experimentação desenvolvido, como o número de tabelas de factos, dimensões,

indicadores e relatórios, entre outros. Trata-se de uma visão geral do ambiente que suportou a sua conceção, estrutura e interface.

Tabela 9 - Características do ambiente

Número	Particularidade	Ferramenta
3	Tabela de Factos no DW	<i>SQL Server</i>
11	Dimensões no DW	<i>SQL Server</i>
1	Carregamento do Cubo OLAP	<i>Visual Studio – Analysis Service</i>
5	Indicadores do negócio	<i>Visual Studio – Analysis Service</i>
7	Relatórios/ <i>Dashboards</i> / <i>Scorecards</i>	<i>Excel; Excel & Power Pivot; Excel & Power View</i>
		<i>Power BI</i>
		<i>Tableau</i>
		<i>Performance Point</i>

4.7 Ambiente de Testes

Nesta atividade foi elaborado um ambiente de teste com base no DW, permitindo avaliar as ferramentas que processaram as atividades de EED. Esta experimentação permite facultar informações sobre a capacidade de resposta de cada ferramenta, em relação aos requisitos estabelecidos, nesta perspetiva, os relatórios e *dashboards*, a seguir, apresentados ilustram o que foi alcançado durante a experimentação.

➤ **Microsoft Excel**

O *dashboard* ilustrado na figura 24, exemplifica os tipos de *dashboards* elaborados com recurso ao *Excel*, através das funcionalidades bastante intuitivas e fácil usabilidade, *pivot table* e *pivot chart*. Este, apresenta as vendas por distribuição geográfica e vendas por categoria de produto, onde é possível **filtrar** os mesmos por anos, por canais de vendas (*channel of sales*) e categorias/subcategorias dos produtos vendidos.

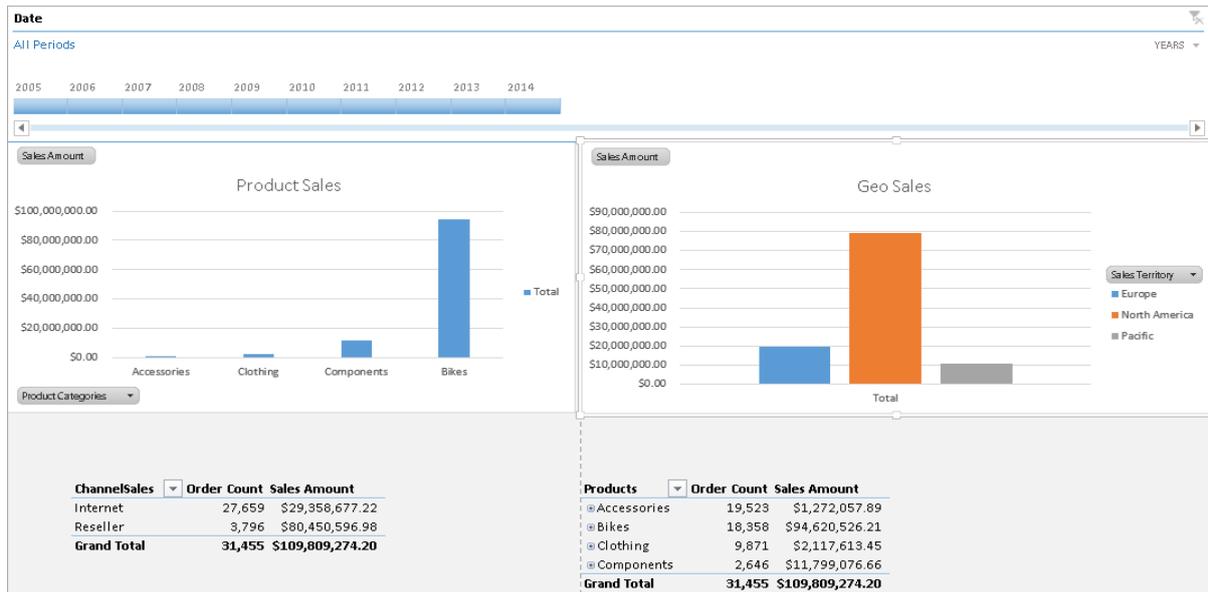


Figura 24 - Dashboard de vendas – Excel (Elaborado pelo autor)

➤ **MS Excel & Power View & Power Pivot**

Através do suplemento *Power View*, é possível tornar o Excel uma ferramenta ainda mais poderosa, agregando funcionalidades de *Data Driven (drill-up/down)* que auxiliam ter uma perspectiva dinâmica dos dados de DW. No entanto, o *Power Pivot*, outro suplemento do Excel, é um facilitador na construção de modelos de dados, importando dados de várias fontes distintas e viabilizando convergi-las num único modelo de dados, ilustrado na figura 25. Isto, aumenta a quantidade de dados, assim como, contribui para análises detalhadas e de várias perspectivas. O *dashboard* da figura 26, ilustra as vendas por categoria de produto e regiões dos Estados Unidos da América, este é dinâmico e **permite aumentar/diminuir o nível de detalhe**, graças ao modelo de dados elaborado no *Power Pivot*.

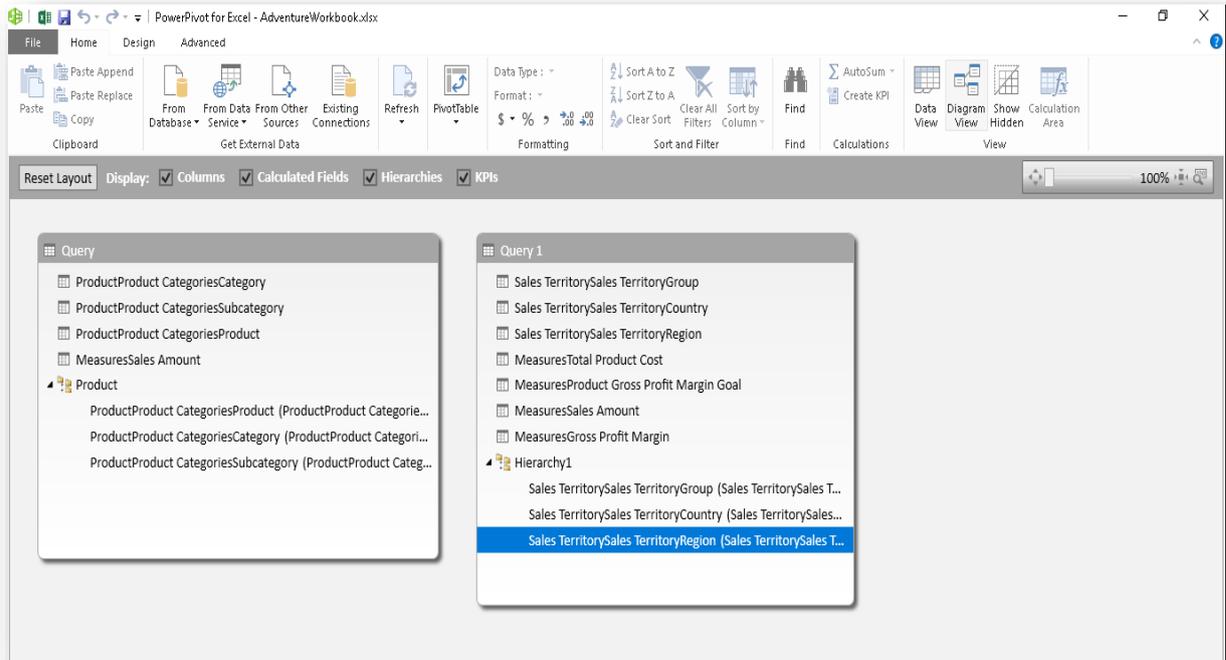


Figura 25-Mashup de dados no Power Pivot

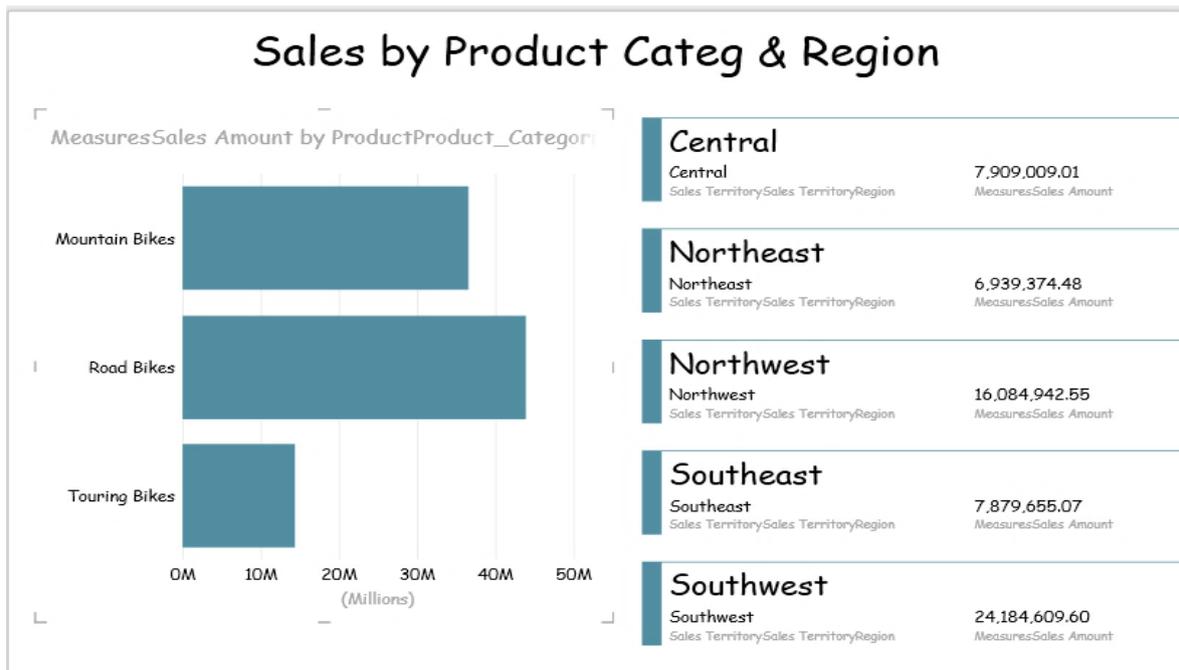


Figura 26 - Dashboard criado no Power View (Elaborado pelo autor)

Na figura 27, utilizando o mesmo modelo de dados criado no *Power Pivot*, elaborou-se um relatório com as vendas discriminadas por categorias de produtos e dados sobre custos/vendas, por regiões dos Estados Unidos da América.

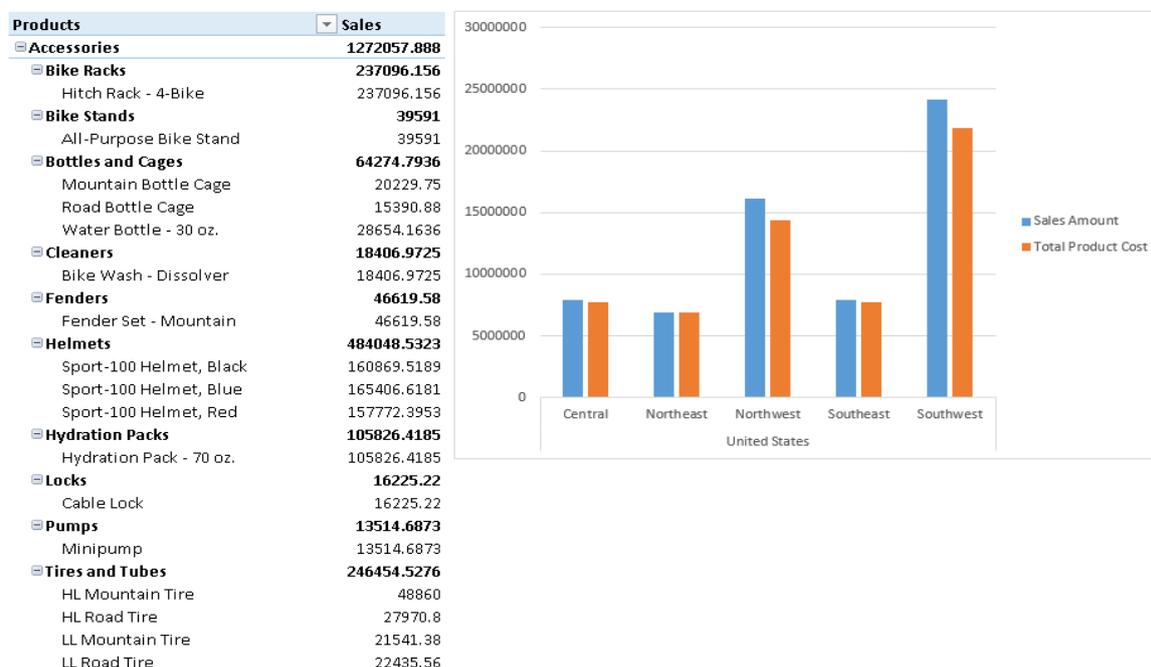


Figura 27 - Dashboard criado no Excel (Elaborado pelo autor)

➤ **MS Excel & Microsoft SharePoint**

Dentro do ambiente colaborativo do *SharePoint*, foi criado um site (no modo *Business Intelligence Center*¹²) onde foi possível importar os *dashboards* criados através do Excel. Esse ambiente foi utilizado no intuito de testar as funcionalidades de partilha em tempo real de relatórios, *dashboards* e gráficos, sendo que perante uma necessidade *ad-hoc*, o *SharePoint* facilita o acesso aos mesmos, incorporando as capacidades (filtros, *KPI's*, *slicers*) do Excel.

A figura 28 ilustra um dos *dashboards* criados e importados, para ambiente colaborativo do *SharePoint*, contendo os dados referentes às vendas (através da internet e revendedores), organizados por país. Onde também, é possível filtrar as vendas por tipo de promoção exercido.

¹² <https://support.office.com/en-us/article/What-is-a-Business-Intelligence-Center-dcd208d0-f50b-46fb-ac8a-3c9a8b2ab357>

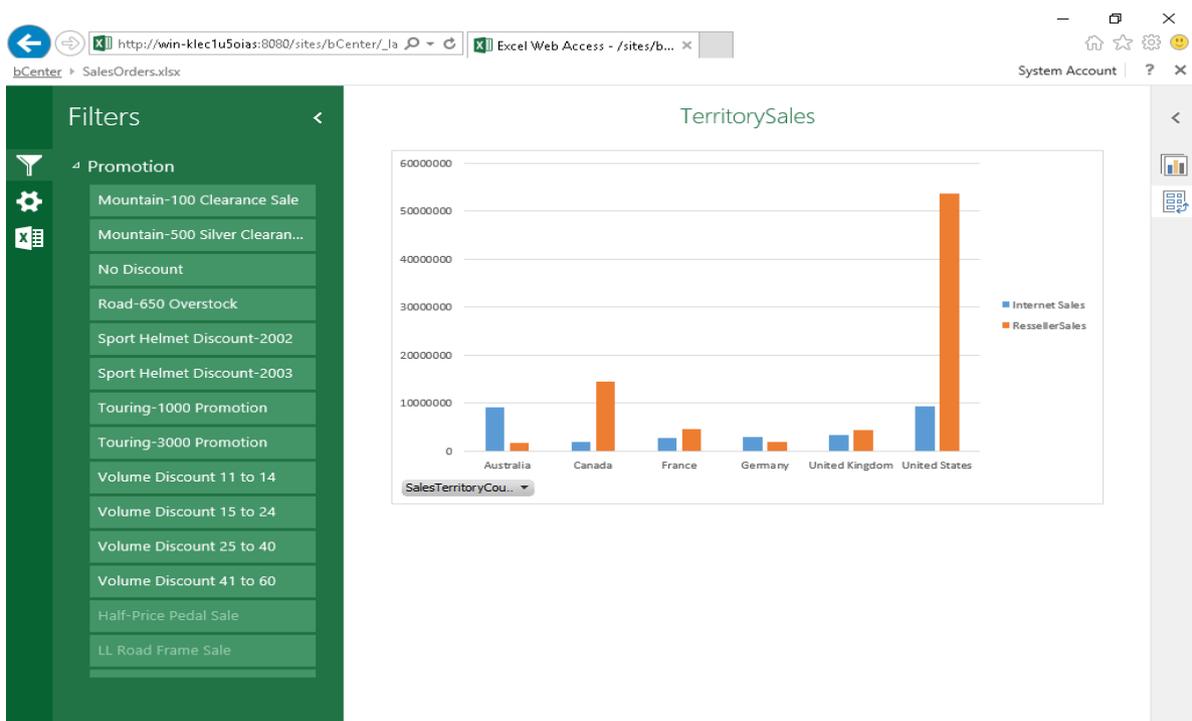


Figura 28 - Dashboard em ambiente SharePoint (Elaborado pelo autor)

➤ **MS SharePoint & Performance Point**

Os serviços do *Performance Point* ajudam os utilizadores na tomada de decisões de negócio, que se alinham com os objetivos e à estratégia da empresa. Estes serviços, oferecem a capacidade de extrair vários tipos de visualizações de dados. Os utilizadores podem interagir com os relatórios, criar gráficos baseados em cubos OLAP, KPIs, árvores de decomposição e diagramas do *Microsoft Visio* (Lauer et al., 2012). Outros tipos de conteúdo podem ser integrados, exibindo-os em qualquer página *Web*, como um *dashboard*.

Com gráficos sobre dados assentes em modelos multidimensionais, os utilizadores podem executar análises *ad-hoc*, manipular dados dimensionais, navegar pelas hierarquias e dinamizar/manipular os gráficos, de maneira rápida e eficaz em apenas alguns cliques, como ilustra a figura 29. Esse tipo de análise permite que os utilizadores interajam com os dados, capacitando-os na obtenção de respostas que podem estar ocultas.

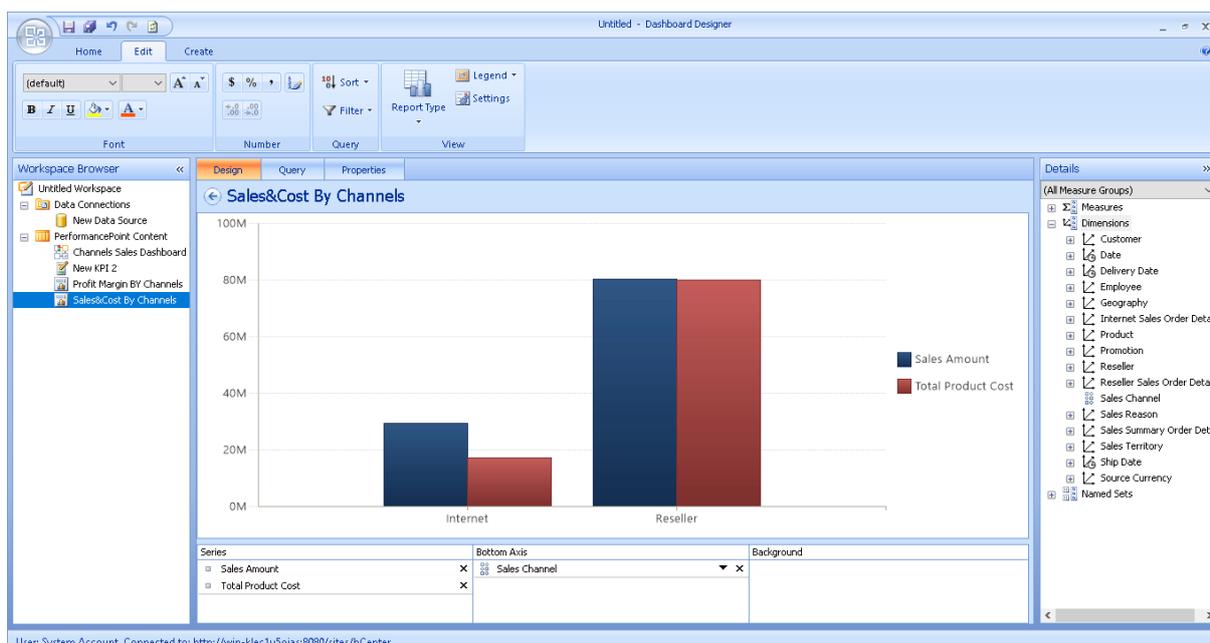


Figura 29 – Performance Point Dashboard Designer (Elaborado pelo autor)

Através do *Performance Point Dashboard Designer* é possível criar uma variedade de *dashboards*, relatórios, filtros, KPIs, etc., de forma simples. A figura 30 apresenta um *dashboard* criado, com os dados sobre as vendas e margem de lucro, agregadas por meios de venda. Este *dashboard* foi integrado no ambiente colaborativo do *SharePoint*, onde é possível expor e partilhar numa rede intranet ou extranet em tempo real. Isto é propício para assegurar uma dinâmica de trabalho interativa.

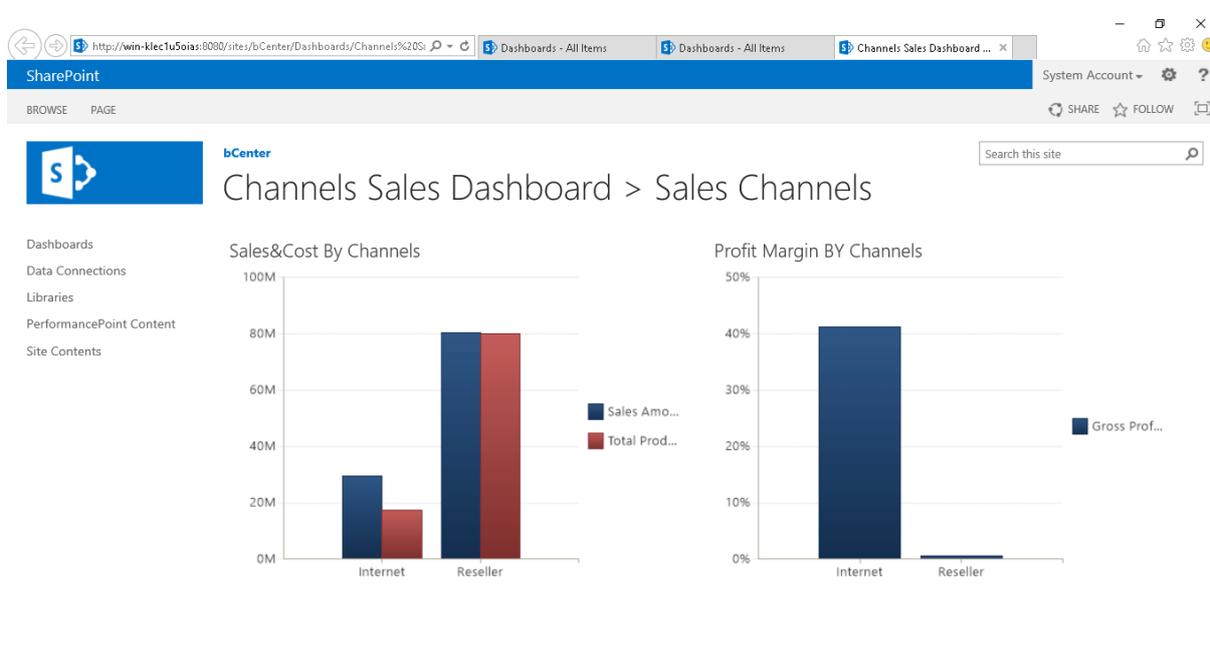


Figura 30 - Dashboard criado no Performance Point (Elaborado pelo autor)

➤ **Power BI Desktop**

No *Power BI Desktop* elaborou-se o relatório ilustrado na figura 31, para apresentar as vendas por categoria de produto e o KPI relativo à margem de lucro, estes podem ser automaticamente filtrados por continente (Norte América, Europa e Pacífico) no mapa. Embora, a conceção de relatórios/*dashboards* no *Power BI* seja rápida e intuitiva, o mesmo não possibilita exportação para outros tipos de ficheiro (word, html, jpg) para além de pdf e da funcionalidade de publicação no *Power BI online*.

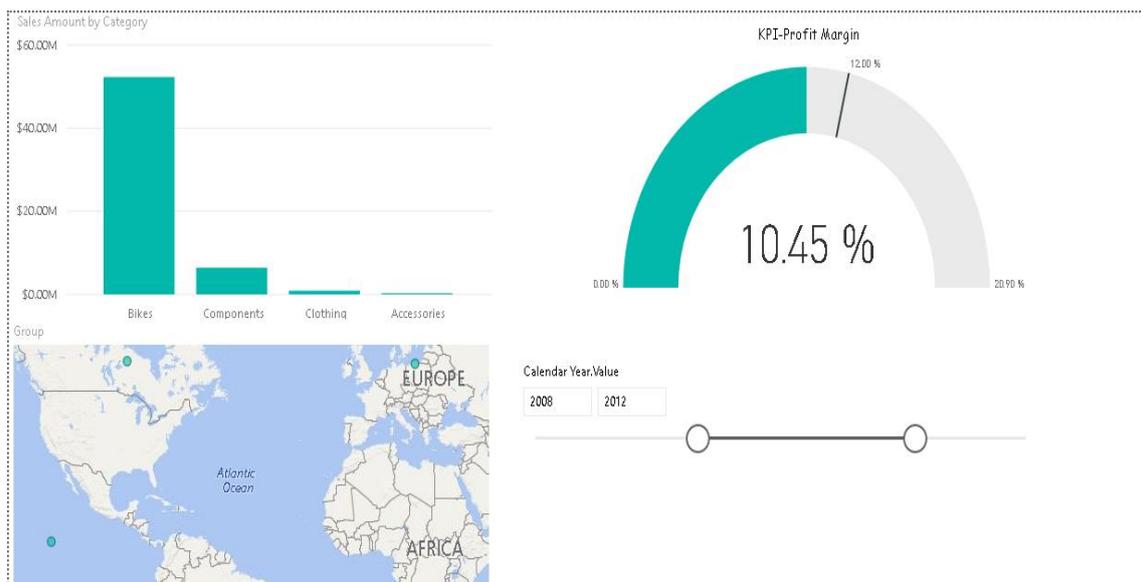


Figura 31 – Dashboard criado no Power BI (Elaborado pelo autor)

➤ **Tableau Desktop**

O *Tableau Desktop* viabiliza a criação de gráficos e *dashboards* de forma eficiente, graças as funcionalidades automatizadas (como interpretação dos tipos de dados e formação de gráficos/filtros). A figura 32 ilustra um *dashboard* criado nesse ambiente, representando as vendas por categoria de produto e continente, com a discriminação da margem de lucro por continente. Também, contendo um filtro dos anos em que se processaram as vendas.

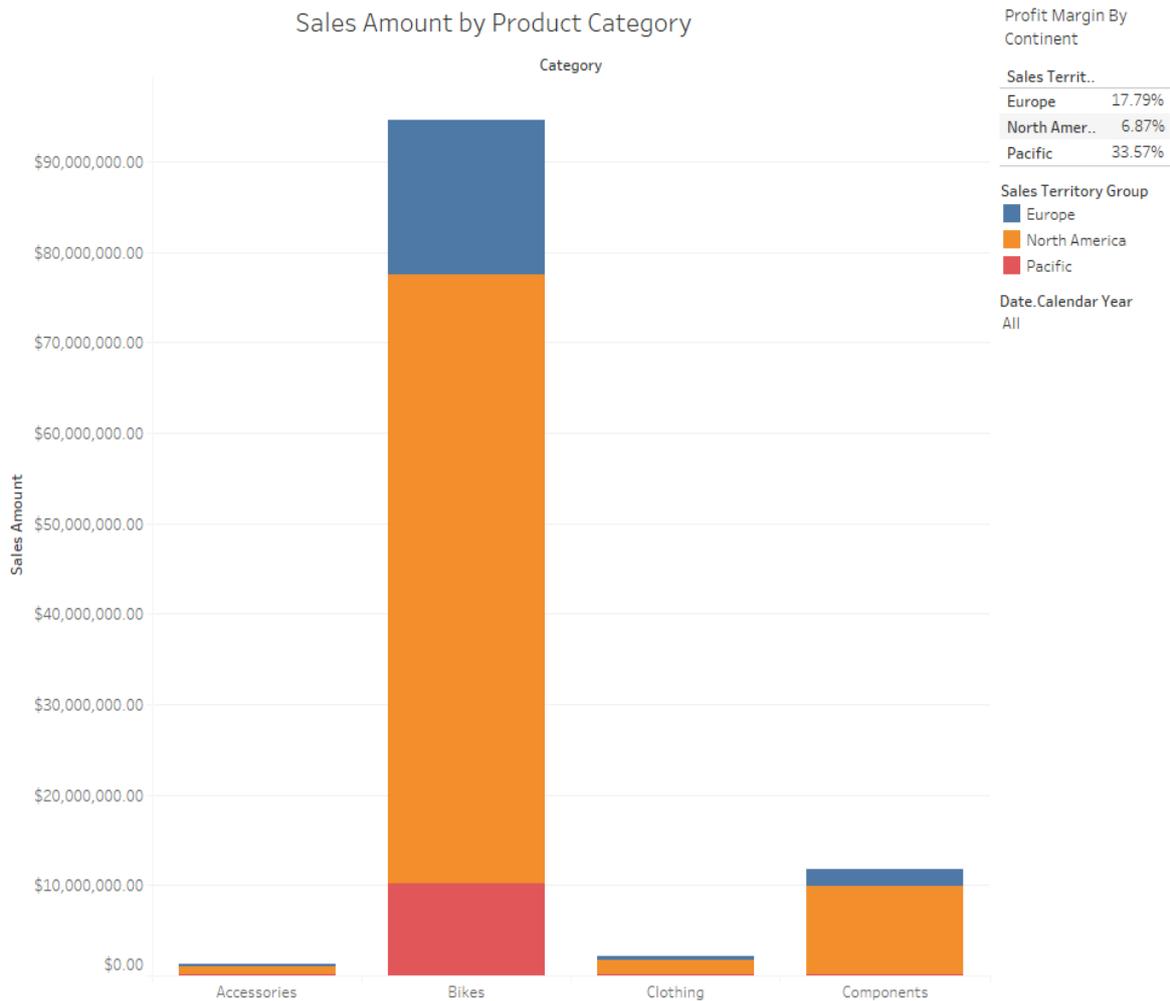


Figura 32 - Dashboard criado no Tableau Desktop (Elaborado pelo autor)

5 TIPOLOGIA DE CATALOGAÇÃO

Após a discussão dos conceitos e o cumprimento das tarefas, inerentes aos objetivos desta dissertação, este capítulo apresenta a descrição do artefacto construído (a tipologia de catalogação) e os seus resultados.

5.1 Métricas de Avaliação

O artefacto desenvolvido possibilita identificar as funcionalidades das ferramentas selecionadas, face aos requisitos exigidos dos utilizadores casuais, sendo que, estes requisitos dizem respeito a um contexto genérico, com o pressuposto de utilização em atividades de **Extração e Exploração de Dados (EED)** de sistemas de *Business Intelligence*.

De acordo com os requisitos de cada ferramenta, atribui-se uma classificação baseada em uma **escala qualitativa ordinal de 0 a 3**, presente na tabela 10, aplicada a cada ferramenta, de acordo com a sua correspondência perante os requisitos.

Tabela 10 - Escala de classificação

Escala	Significado	Informação
0	Ausente	A característica não é evidente ou está em falta
1	Insuficiente	A característica existe, mas responde de forma exígua
2	Suficiente	A característica existe e responde de forma razoável
3	Ótimo	A característica existe e responde de forma cabal

Também, **cada classe dos utilizadores casuais tem um peso**, conforme a sua habilidade de utilização das ferramentas, como demonstra a tabela 11.

Tabela 11-Peso das Classes dos utilizadores casuais

Classe dos Utilizadores Casuais	Tipo de Utilizador	Peso
Visualizador (V)	Executivos	30%
Navegador (N)	Gestores	60%
Explorador (E)	Analistas	10%
Todas (VNE)	Todos	100%

5.2 Tipologia de Catalogação

Para avaliação desta tipologia de catalogação, apresentada na tabela 12, teve-se em conta os requisitos/características das ferramentas de acordo com as formas de interação identificadas, enquadrando-os com as classes dos utilizadores casuais levantados, para posteriormente cruzar com as ferramentas de EED, de modo a avaliar como as mesmas respondem a cada requisito.

Tabela 12 – Tipologia de classificação das ferramentas

Características	Classe	Excel	Excel & Power View	Excel & Power Pivot	Performanc e Point	Power BI	Tableau
Permite ser utilizado sem que os utilizadores estejam familiarizados com linguagens de consulta, como <i>SQL</i> e <i>MDX</i> (<i>Multidimensional Expressions</i>)	V	3	2	2	2	3	3
Permite a análise de grandes quantidades de dados	NE	1	3	3	3	3	3
Permite a criação avançada de relatórios usando <i>SQL</i> , <i>MDX</i> ou outras linguagens de consulta	NE	0	0	3	3	2	2

Características	Classe	Excel	Excel & Power View	Excel & Power Pivot	Performanc e Point	Power BI	Tableau
Permite a criação rápida de relatórios e visualizações de dados	NE	3	2	2	2	3	3
Permite agregações e KPIs personalizados	NE	1	1	3	3	3	3
Permite elaborar relatórios que forneçam pontuações para o desempenho de uma organização, departamento ou indivíduo	VN	2	3	1	3	3	3
Permite KPIs que direcionam a navegação para outros estilos de relatório no <i>dashboard</i>	NE	1	1	1	2	2	3
Permite o acesso a fontes de dados empresariais, departamentais, externas e pessoais	NE	3	1	3	1	3	3
Permite a utilização de interfaces <i>drag-and-drop</i> para design de relatórios	NE	2	2	2	2	3	3
Permite que as informações sejam distribuídas regularmente	VNE	0	0	0	3	3	1
Permite que os dados sejam atualizados automaticamente	NE	0	2	3	3	4	0

Características	Classe	Excel	Excel & Power View	Excel & Power Pivot	Performanc e Point	Power BI	Tableau
Permite que os relatórios sejam embebidos em outras aplicações do negócio	VNE	3	2	2	2	1	2
Permite que os relatórios sejam exportados em múltiplos formatos com PDF, Excel, Microsoft Word e HTML	VNE	3	3	0	2	1	1
Permite que os utilizadores façam atividades <i>Data Driven</i> (<i>drill-down</i> , <i>drill-up</i> , filtragem e <i>pivot</i>) e apresentem os dados em tabelas, gráficos e outras visualizações	VN	2	3	3	3	3	3
Permite que os utilizadores incluam seus relatórios e visualizações em apresentações ou compartilhem informalmente com outros colegas de trabalho	VN	3	1	1	3	3	3
Permite que os utilizadores trabalhem independentemente dos técnicos de TI	VNE	3	2	1	2	3	3
Permite relatórios que apresentem os objetivos de negócio e KPIs discriminados hierarquicamente, podendo ser filtrados para auxiliar na	VNE	1	2	3	3	3	2

identificação de “outliers”							
-----------------------------	--	--	--	--	--	--	--

5.3 Resultados da Tipologia

Após a incorporação das ferramentas na tipologia elaborada, procedeu-se à avaliação da classificação de cada uma destas, de acordo com as métricas e os valores obtidos. Esta avaliação, abrange duas perspetivas:

- 1) **Individual** - tendo em conta as classes dos utilizadores casuais, adjudicados à avaliação das ferramentas, a tipologia possibilitará delimitar qual a ferramenta adequada para cada classe, conseqüentemente para cada tipo de utilizador. A fórmula a seguir, onde ***n*** é **número de requisitos/características** e ***X*** é a **escala**:

$$\sum_{i=1}^n X_i$$

- 2) **Global** - Como cada característica encontra-se associada às classes dos utilizadores e por conseguinte ao peso destes. Ao atribuir-se uma escala a uma ferramenta, esta, é multiplicada pelo peso (da classe de utilizador), ou seja, a pontuação final é determinada pelo somatório do produto - escala e peso. A fórmula seguinte transcreve esta perspetiva, onde ***n*** é **número de características/funcionalidades**, ***X*** é a **escala** e ***Y*** é o **peso das classes dos utilizadores casuais**:

$$\sum_{i=1}^n X_i * Y_i$$

5.3.1 Perspetiva Individual

Após aplicar a tipologia de catalogação das ferramentas, analisaram-se as pontuações obtidas, de acordo com as classes dos utilizadores casuais (visualizador, navegador e explorador). Isto produziu resultados que apontam o caminho, para identificar a ferramenta apropriada aos requisitos desses utilizadores. A figura 33, representa os resultados obtidos segundo as classes, onde pode-se observar as ferramentas propícias para cada classe de utilizador casual, sendo que, na primeira classe acontece um empate técnico entre duas ferramentas, *Performance Point* e *Power BI*, e em relação as demais classes, uma clara vantagem do *Power BI*.

Perspetiva Individual

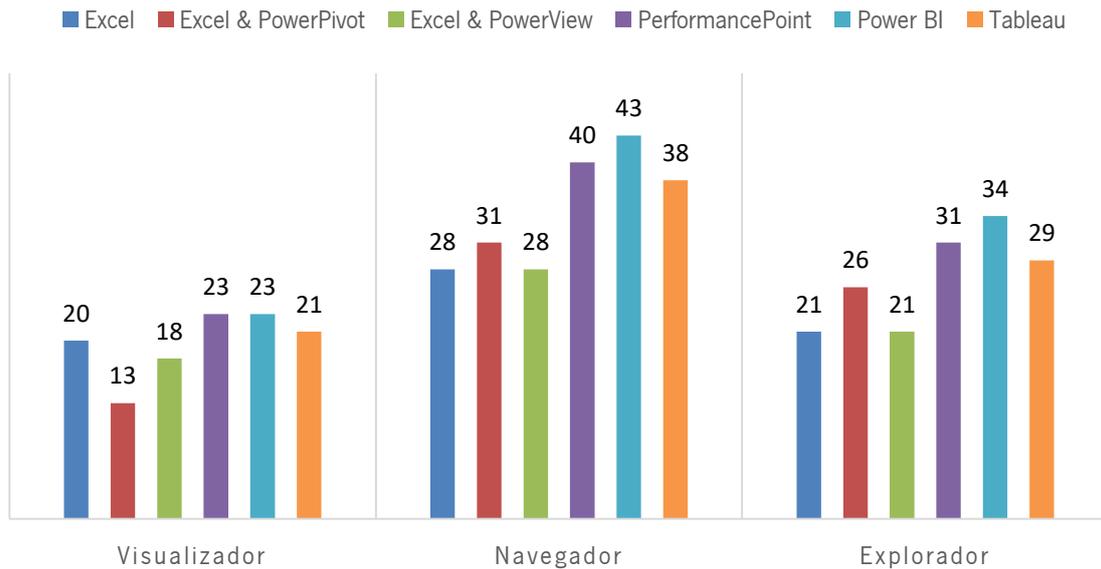


Figura 33 - Resultados da Tipologia-Perspetiva Individual (Elaborado pelo autor)

5.3.2 Perspetiva Global

Dentro desta perspetiva é levado em conta as variáveis que influenciam a utilização generalizada das ferramentas, ou seja, uma utilização feita por vários tipos de utilizadores em simultâneo, no intuito de destacar a mais propícia no âmbito global. A figura 34 ilustra os resultados obtidos, onde pode-se constatar uma pontuação favorável à ferramenta **Power BI**.

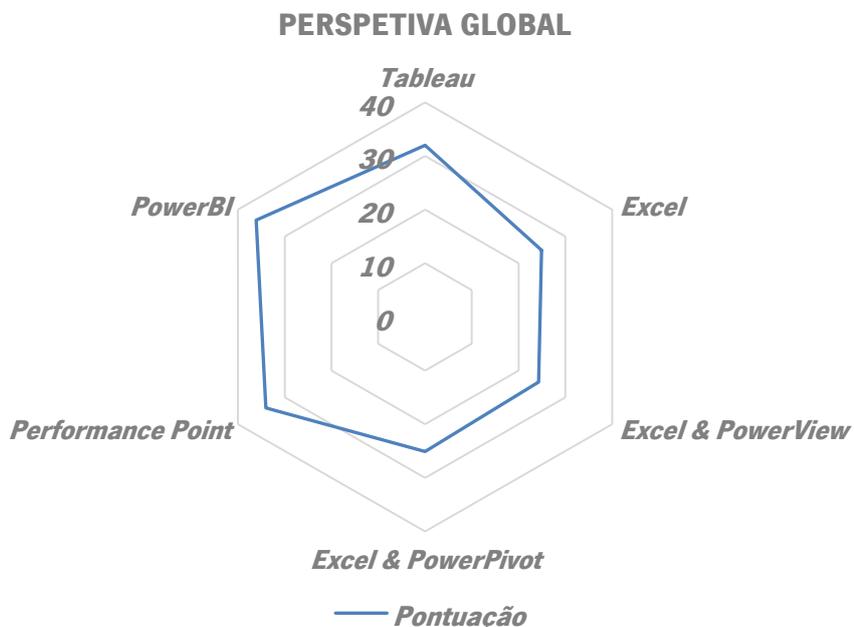


Figura 34 - Resultados da Tipologia – Perspetiva Global (Elaborado pelo autor)

5.4 Discussão dos Resultados

Com a tipologia elaborada é proposta uma nova forma de catalogação das ferramentas de EED, que considera os tipos de utilizadores casuais e os seus requisitos, aplicando uma escala de avaliação para determinar e identificar a melhor ferramenta.

Para o desafio enfrentado e após a conceção da tipologia capaz de determinar a ferramenta mais indicada aos utilizadores casuais, é necessário implementá-la no contexto real. Para isto, é necessário determinar os pesos de cada classe/tipo de utilizador casual, as escalas para cada forma de interação, e as ferramentas a analisar. Este processo poderá ser uma mais valia nas organizações, facilitando a adaptação e adoção de novas ferramentas no ambiente do negócio, atendendo os requisitos/características de cada tipo de utilizador, durante essa adaptação e adoção.

A figura 35 ilustra como o processo poderá ser aplicado.

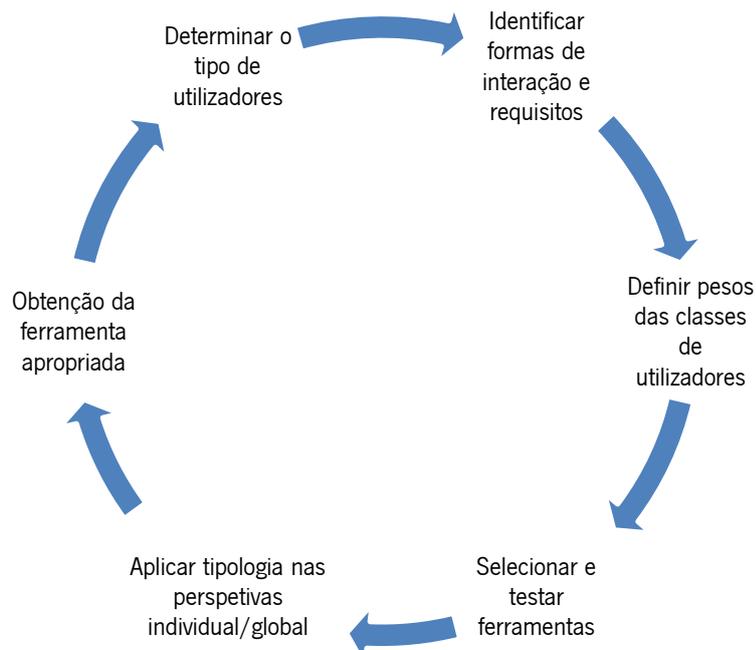


Figura 35 - Processo de Aplicação da Tipologia

6 CONCLUSÕES, LIMITAÇÕES E TRABALHO FUTURO

Neste capítulo serão abordados os aspetos finais desta dissertação, a síntese do trabalho executado, as principais conclusões retiradas, as limitações que surgiram durante o desenvolvimento deste projeto e os aspetos a serem abordados futuramente.

6.1 Síntese do Trabalho

O progresso desta dissertação dividiu-se em quatro fases. A **primeira** coincide com a revisão de literatura e consistiu na pesquisa de informação relacionada com o tema *Business Intelligence, Data Extraction e Data Exploration*, identificação e classificação dos utilizadores.

A **segunda** componente coincide com a pesquisa das ferramentas no mercado, utilizadas na Extração e Exploração de Dados (EED) nos diversos contextos de sistemas de BI, definição dos critérios para filtrar as ferramentas a serem comparadas e concretizando uma pesquisa de informação relativa as formas de iteração com as ferramentas, de acordo com os requisitos dos utilizadores.

Na **terceira** componente foi finalizado a construção do ambiente de experimentação das ferramentas, onde foram testadas as ferramentas de EED de um *Data Warehouse*. Na última fase, foi construída a matriz/tipologia de catalogação das ferramentas, assim como, a avaliação das ferramentas em linha com a tipologia definida.

6.2 Análise dos Objetivos

Para dar resposta aos objetivos definidos no Capítulo 2 deste documento e tendo como ponto de partida o conhecimento retirado do enquadramento conceptual realizado relativamente à Extração e Exploração de Dados, às ferramentas de EED no mercado e os tipos de utilizadores casuais associados a sistemas de *Business Intelligence*. A compreensão dos principais paradigmas e termos associados ao tema e o trabalho experimental elaborado, permitiu compreender parte da oferta existente no mercado,

no que diz respeito às ferramentas e EED, assim como, obter uma tipologia de catalogação e avaliação das ferramentas de EED perante os requisitos dos utilizadores casuais.

Esta solução é capaz de identificar os requisitos dos utilizadores e avaliar em que medida as ferramentas existentes respondem a demanda dos utilizadores casuais, fornecendo assim um processo para determinar a ferramenta no mercado, mais propícia aos mesmos. Além de promover e aumentar a motivação associada à realização de todo o processo incluído neste tema de dissertação, este trabalho permitiu concluir que investir na compreensão e identificação dos requisitos dos utilizadores, perante as tecnologias associadas aos sistemas de BI, é crucial na escolha das ferramentas de EED.

6.3 Análise das Metodologias

O principal objetivo do desenvolvimento desta dissertação é encontrar uma solução para a identificação de quais e como, as ferramentas de Extração e Exploração de Dados EED, podem dar resposta aos requisitos dos utilizadores casuais. Para a realização desse objetivo, recorreu-se à metodologia *Design Science Research*, para encontrar uma solução plausível para o problema. Após a análise do problema e de acordo com a metodologia DSR, a solução encontrada foi a criação de uma tipologia de catalogação, que possibilita identificar como e quais as ferramentas de EED, que melhor se adequam aos requisitos dos utilizadores. Relativamente ao cumprimento da última fase do DSR – a comunicação, artigos estão a ser preparados com o objetivo de comunicar os resultados atingidos com a realização e concretização desta dissertação.

6.4 Resultados Atingidos

O trabalho realizado durante esta dissertação, englobou a **identificação de vários tipos de utilizadores**, mais concretamente os casuais; os requisitos desses utilizadores; a sugestão de uma nova definição para as atividades de **Extração e Exploração de Dados (EED)**; a definição das **formas de interação para avaliar as ferramentas existentes** no mercado e, **a tipologia de catalogação para determinar a ferramenta adequada**. Isto, auxiliará doravante os profissionais e as organizações que utilizam as ferramentas de EED, conectadas com sistemas de *Business Intelligence*, a melhorar o processo de decisão, na medida em que os seus requisitos são alicerces na

adoção de tecnologias/ferramentas, pois, necessitam em situações *ad-hoc*, de analisar, explorar e decidir com base nos dados.

6.5 Limitações e Trabalho Futuro

6.5.1 Limitações

Para o desenvolvimento deste trabalho foram enfrentadas algumas limitações que impediram a realização de novas experiências e aprendizagens mais completas para a tipologia, devido a janela temporal para a conclusão do trabalho, assim como, o condicionamento das licenças das ferramentas disponíveis no mercado.

No que diz respeito à experimentação realizada, esta foi limitada a um contexto de negócio específico (vendas), o que condicionou o tipo testes realizados e os requisitos avaliados nas ferramentas.

A falta de documentação e informações sobre formas de avaliação das ferramentas de EED e a quantidade de ferramentas selecionadas foram as condicionantes que influenciaram diretamente no resultado da avaliação da tipologia.

6.5.2 Trabalho Futuro

Para trabalho futuro é proposto as necessidades identificadas que por motivos diversos não foram contempladas nesta dissertação. Contudo, devido à relevância destas, devem ser referidas com a intenção de abrir novas perspectivas para a realização de trabalhos que poderão dar continuidade ao que foi conseguido nesta dissertação.

A **primeira proposta** é, a identificação de requisitos dos utilizadores casuais num contexto empresarial, no sentido de obter informações reais das necessidades destes, assim como, a delimitação da pesquisa das ferramentas de Extração e Exploração de Dados (EED) no mercado, de acordo com a estratégia de investimento (ferramentas *Open-Source* ou Comercial).

Uma **segunda proposta** é a implementação da tipologia num software de recomendação de ferramentas com base nos parâmetros e escalas de avaliação da tipologia.

Uma **terceira proposta** seria a determinação de um conjunto de testes padrão que poderão fornecer melhor resposta na avaliação das ferramentas do mercado segundo a tipologia.

REFERÊNCIAS

- Ackoff, R. L. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis*, 16(1), 3–9. Retrieved from <https://faculty.ung.edu/kmelton/Documents/DataWisdom.pdf>
- Aruldoss, M., Lakshmi Travis, M., & Prasanna Venkatesan, V. (2014). A survey on recent research in business intelligence. *Journal of Enterprise Information Management*, 27(6), 831–866. <https://doi.org/10.1108/JEIM-06-2013-0029>
- Baars, H., & Kemper, H.-G. (2008). Management Support with Structured and Unstructured Data—An Integrated Business Intelligence Framework. *Information Systems Management*, 25(2), 132–148. <https://doi.org/10.1080/10580530801941058>
- Bellinger, G., Castro, D., & Mills, A. (2004). Data, Information, Knowledge, and Wisdom, 5–7.
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88. <https://doi.org/10.1145/1978542.1978562>
- Chen, H., Chiang, R. H. L., Lindner, C. H., Storey, V. C., & Robinson, J. M. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *Management Information Systems Quarterly*, 36(4), 1165–1188. Retrieved from <http://aisel.aisnet.org/misq/vol36/iss4/16>
- Davenport, T. H., & Prusak, L. (2000). Working knowledge – how organisations manage what they know. *Harvard Business School Press, Boston Massachusetts*, 21(8), 395–403.
- Duncan, O. et. al. (2018). Analysis Services Tutorial Scenario. Retrieved August 11, 2018, from <https://docs.microsoft.com/en-us/sql/analysis-services/analysis-services-tutorial-scenario>
- Dyché, J. (2007). Categorizing Business Intelligence Users, 4. Retrieved from <https://searchbusinessanalytics.techtarget.com/news/2240036691/Categorizing-business-intelligence-users>
- Eckerson, W. W. (2010). *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. Business (2nd ed.). Hoboken, New Jersey: John Wiley & Sons. <https://doi.org/10.2514/6.2008-3494>
- Eckerson, W. W. (2013). Classifying Business Users. Retrieved February 10, 2018, from http://www.b-eye-network.com/blogs/eckerson/archives/2013/09/classifying_bus.php
- El-Sappagh, S. H. A., Hendawi, A. M. A., & El Bastawissy, A. H. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University - Computer and Information Sciences*, 23(2), 91–104. <https://doi.org/10.1016/j.jksuci.2011.05.005>
- Feldman, D., & Himmelstein, J. (2013). *Developing Business Intelligence Apps for SharePoint*. Retrieved from <http://books.google.cv/books?id=xDXfeopC-kMC>
- Ferreira, J., Miranda, M., Abelha, A., & Machado, J. (2010). O Processo ETL em Sistemas Data Warehouse. *INForum 2010 - II Simpósio de Informática*, (9), 757–765. Retrieved from <http://www.di.uminho.pt>
- Golfarelli, M., & Rizzi, S. (2009). Data Warehouse Design, Modern Principles and Methodologies. *Data Warehouse*.
- Grossmann, W., & Rinderle-Ma, S. (2015). *Fundamentals of Business Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-46531-8>
- Hart, M., & Saxton, A. (2018). What is a Power BI dashboard? Retrieved January 30, 2018, from <https://docs.microsoft.com/en-us/power-bi/service-dashboards>
- Hevner, A., & Chatterjee, S. (2010). *Design Science Research in Information Systems*. (S. Ramesh & S. Voß, Eds.), *Design Research in Information Systems* (Vol. 22). Springer. <https://doi.org/10.1007/978-1-4419-5653-8>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems

- Research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Inmon, B. (2005). *Building the Data Warehouse* (4 ed.). Wiley. <https://doi.org/10.1017/CBO9781107415324.004>
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit - The Definitive Guide to Dimensional Modeling* (3 ed.). Wiley. <https://doi.org/10.1145/945721.945741>
- Lauer, J., Cameron, S., Nelson, J., & Rocca, V. (2012). How to Choose the Right Reporting Tools for Your Instrument Control System. Microsoft. Retrieved from [https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2012/jj129615\(v=msdn.10\)](https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2012/jj129615(v=msdn.10))
- Loshin, D. (2012). *Business Intelligence: The Savvy Manager's Guide*. Elsevier Inc.
- Luhn, H. P. (1958). A Business Intelligence System. *IBM Journal of Research and Development*, 2(4), 314–319. <https://doi.org/10.1147/rd.24.0314>
- Macaulay, T. (2016). A timeline of Business Intelligence: How decision-makers have been making sense of data. Retrieved January 29, 2018, from <http://www.cio.co.uk/it-applications/business-intelligence-timeline-3429740/>
- Mendes, P. D. T. (2018). *Desenvolvimento de soluções à medida em Sharepoint*. Universidade de Lisboa. Retrieved from https://repositorio.ul.pt/bitstream/10451/32647/1/ulfc124135_tm_Pedro_Mendes.pdf
- Negash, S. (2004). Business Intelligence. *Communications of the Association for Information Systems*, Vol. 13(February), 177–195. <https://doi.org/10.1002/9781118915240.ch7>
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2008). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Sá, J. V. de O. (2009). *Metodologia de Sistemas de Data Warehouse*. Universidade do Minho. Retrieved from http://repositorium.sdum.uminho.pt/bitstream/1822/10663/4/Tese de doutoramento_Jorge Vaz de Oliveira e Sá_2009.pdf
- Sauter, V. L. V. L. (2011). *Decision Support Systems for Business Intelligence: Second Edition*. *Decision Support Systems for Business Intelligence* (2 ed.). John Wiley & Sons. <https://doi.org/10.1002/9780470634431>
- Trujillo, J., & Mat, A. (2012). Business Intelligence 2.0 : A General Overview. In *Business Intelligence* (Vol. 96, pp. 98–116).
- Turban, E., Sharda, R., Delen, D., King, D., & Aronson, J. E. (2017). *Business Intelligence: A Managerial Approach* (2nd ed.). Prentice Hall. Retrieved from <https://books.google.com/books?id=lvZORAAACAAJ&pgis=1>
- Wang, J. (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (2 ed.). IGI Global. <https://doi.org/10.4018/978-1-60566-010-3>
- Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), xiii–xxiii. <https://doi.org/10.1.1.104.6570>

ANEXOS

Anexo I-Modelo Dimensional

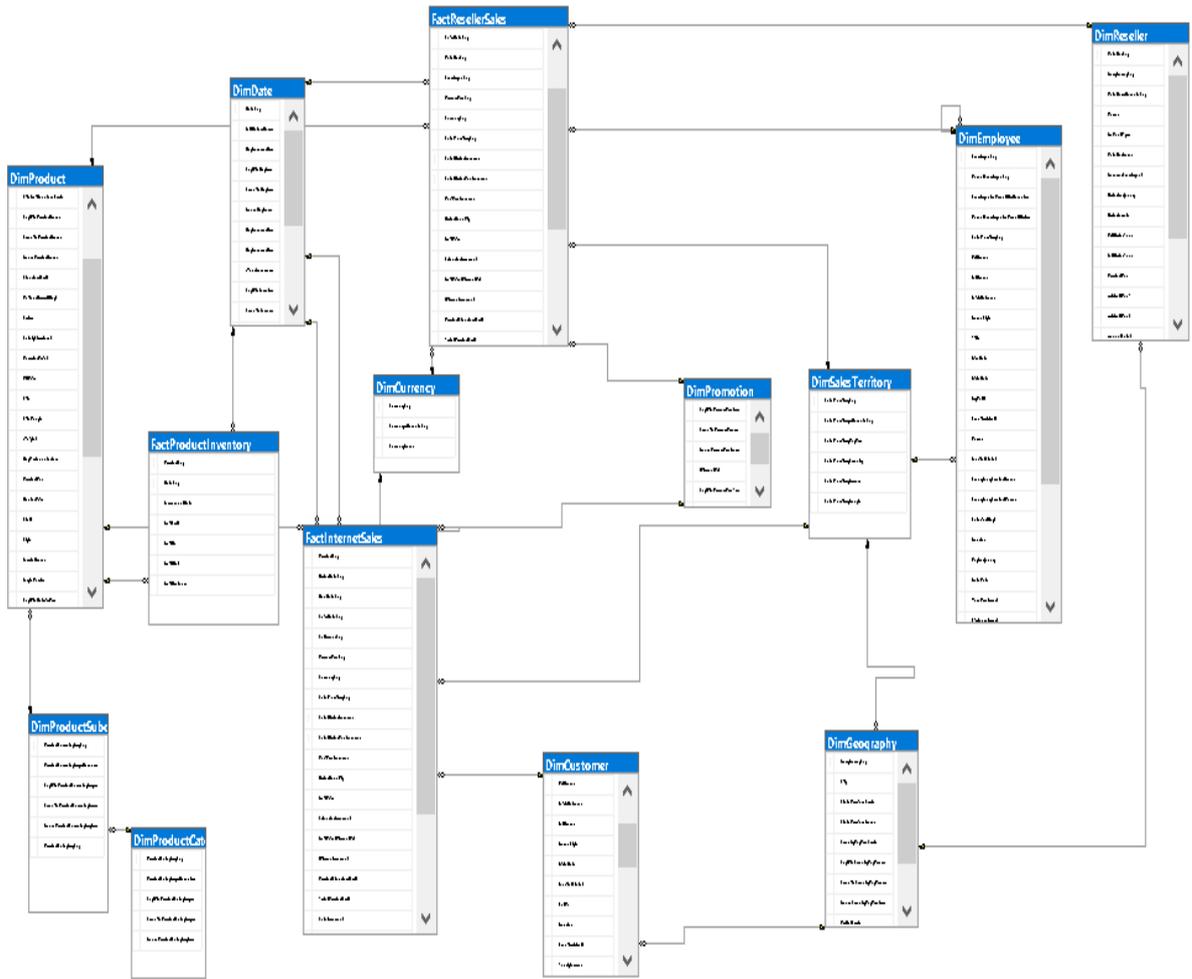


Figura 36 - Modelo dimensional Adventure Works

Anexo II-Tabelas dos Factos

Tabela 13 - Descrição do Facto Internet Sales

Facto <i>Internet Sales</i>		
Atributo	Tipo	Resumo
<i>ProductKey</i>	<i>int</i>	Chave estrangeira da dimensão Product
<i>OrderDateKey</i>	<i>int</i>	Chave estrangeira da dimensão date, que neste caso, corresponde ao dia em que o pedido foi realizado. A dimensão Date atua como <i>role playing dimension</i> , sendo referenciada várias vezes.
<i>DueDateKey</i>	<i>int</i>	Chave estrangeira da dimensão Date, que neste caso, corresponde ao dia em que o caduca o pedido de compra realizado. A dimensão Date atua como <i>role playing dimension</i> , sendo referenciada várias vezes.
<i>ShipDateKey</i>	<i>int</i>	Chave estrangeira da dimensão Date, que neste caso, corresponde ao dia em que o produto foi enviado. A dimensão Date atua como <i>role playing dimension</i> , sendo referenciada várias vezes.
<i>CustomerKey</i>	<i>int</i>	Chave estrangeira da dimensão Customer, que referencia o cliente, que efetua a compra.
<i>PromotionKey</i>	<i>int</i>	Chave estrangeira da dimensão Promotion, que referencia o tipo de promoção realizado.
<i>CurrencyKey</i>	<i>int</i>	Chave estrangeira da dimensão Currency, que referencia o tipo de moeda utilizado na transação da venda.
<i>SalesTerritoryKey</i>	<i>int</i>	Chave estrangeira da dimensão SalesTerritory, que representa o território aonde foi realizado a venda.
<i>SalesOrderNumber</i>	<i>nvarchar(20)</i>	Esta chave estrangeira é de uma dimensão degenerada, que indica o número do pedido de venda
<i>SalesOrderLineNumber</i>	<i>tinyint</i>	Esta chave estrangeira é de uma dimensão degenerada, que indica o número do pedido de venda
<i>RevisionNumber</i>	<i>tinyint</i>	Este atributo indica se houve necessidade de revisão do pedido de compra ou não.
<i>OrderQuantity</i>	<i>smallint</i>	Medida que regista as quantidades pedidas na ordem de compra.
<i>UnitPrice</i>	<i>money</i>	Este atributo indica o preço de cada produto do pedido.
<i>ExtendedAmount</i>	<i>money</i>	Este atributo indica o preço de todos os produtos do pedido.
<i>UnitPriceDiscountPct</i>	<i>float</i>	Este atributo indica o preço de desconto de cada produto do pedido
<i>DiscountAmount</i>	<i>float</i>	Este atributo indica a quantia total de desconto.
<i>ProductStandardCost</i>	<i>money</i>	Medida que indica o custo fixo do produto

Atributo	Tipo	Resumo
<i>TotalProductCost</i>	<i>money</i>	Medida que regista os custos totais do produto
<i>SalesAmount</i>	<i>money</i>	Medida que regista a quantia monetária arrecada com as vendas
<i>TaxAmt</i>	<i>money</i>	Medida que regista a quantia monetária cobrada em termos de impostos com as vendas
<i>Freight</i>	<i>money</i>	Medida que regista a quantia monetária gasta em termos de transporte
<i>CarrierTrackingNumber</i>	<i>nvarchar(25)</i>	Regista o número de <i>tracking</i> do transporte
<i>CustomerPONumber</i>	<i>nvarchar(25)</i>	Número do endereço do cliente
<i>OrderDate</i>	<i>datetime</i>	Regista a data do pedido
<i>DueDate</i>	<i>datetime</i>	Regista a data de expiração do pedido
<i>ShipDate</i>	<i>datetime</i>	Regista a data de expedição

Tabela 14 - Descrição do Facto ResellerSales

Facto Reseller Sales		
Atributo	Tipo	Resumo
<i>ProductKey</i>	<i>int</i>	Chave estrangeira da dimensão Product
<i>OrderDateKey</i>	<i>int</i>	Chave estrangeira da dimensão Date, que neste caso, corresponde ao dia em que o pedido foi realizado. A dimensão Date atua como <i>role playing dimension</i> , sendo referenciada várias vezes.
<i>DueDateKey</i>	<i>int</i>	Chave estrangeira da dimensão Date, que neste caso, corresponde ao dia em que o caduca o pedido de compra realizado. A dimensão Date atua como <i>role playing dimension</i> , sendo referenciada várias vezes.
<i>ShipDateKey</i>	<i>int</i>	Chave estrangeira da dimensão Date, que neste caso, corresponde ao dia em que o produto foi enviado. A dimensão Date atua como <i>role playing dimension</i> , sendo referenciada várias vezes.
<i>ResellerKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>Reseller</i> , referencia o revendedor que efetua a compra.
<i>EmployeeKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>Employee</i> , referencia o funcionário que regista a venda.
<i>PromotionKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>Promotion</i> , referencia o tipo de promoção
<i>CurrencyKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>Currency</i> , que referencia o tipo de moeda utilizada na transação da venda
<i>SalesTerritoryKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>SalesTerritory</i> , que representa o território aonde foi realizado a venda.
<i>SalesOrderNumber</i>	<i>nvarchar(20)</i>	Esta chave estrangeira é de uma dimensão degenerada, que indica o número do pedido de venda
<i>SalesOrderLineNumber</i>	<i>tinyint</i>	Esta chave estrangeira é de uma dimensão degenerada, que indica o número do pedido de venda
<i>RevisionNumber</i>	<i>tinyint</i>	Este atributo indica se houve necessidade de revisão do pedido de compra ou não.
<i>OrderQuantity</i>	<i>smallint</i>	Medida que regista as quantidades pedidas na ordem de compra.
<i>UnitPrice</i>	<i>money</i>	Este atributo indica o preço de cada produto do pedido.
<i>ExtendedAmount</i>	<i>money</i>	Este atributo indica o preço de todos os produtos do pedido.
<i>UnitPriceDiscountPct</i>	<i>float</i>	Este atributo indica o preço de desconto de cada produto do pedido
<i>DiscountAmount</i>	<i>float</i>	Este atributo indica a quantia total de desconto.

<i>ProductStandardCost</i>	<i>money</i>	Medida que indica o custo do produto
<i>TotalProductCost</i>	<i>money</i>	Medida que regista os custos totais do produto
<i>SalesAmount</i>	<i>money</i>	Medida que regista a quantia monetária arrecada com as vendas
Atributo	Tipo	Resumo
<i>TaxAmt</i>	<i>money</i>	Medida que regista a quantia monetária cobrada em termos de impostos com as vendas
<i>Freight</i>	<i>money</i>	Medida que regista a quantia monetária gasta em termos de transporte
<i>CarrierTrackingNumber</i>	<i>nvarchar(25)</i>	Regista o número de tracking do transporte
<i>CustomerPONumber</i>	<i>nvarchar(25)</i>	Número do endereço do cliente
<i>OrderDate</i>	<i>datetime</i>	Regista a data do pedido
<i>DueDate</i>	<i>datetime</i>	Regista a data de expiração do pedido
<i>ShipDate</i>	<i>datetime</i>	Regista a data de expedição

Tabela 15 - Descrição do Facto *ProductInventory*

Facto <i>Product Inventory</i>		
Atributo	Tipo	Resumo
<i>ProductKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>Product</i> , que referencia o produto em inventário
<i>DateKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>Date</i> , que neste caso, corresponde ao dia em que o produto deu entrada no inventário
<i>MovementDate</i>	<i>date</i>	Chave estrangeira da dimensão <i>Date</i> , que desempenha um papel de <i>role playing dimension</i> e neste caso, corresponde ao dia em que o produto deu saída do inventário
<i>UnitCost</i>	<i>money</i>	Medida que regista o preço unitário do produto
<i>UnitsIn</i>	<i>int</i>	Medida que regista os produtos que entraram no inventário
<i>UnitsOut</i>	<i>int</i>	Medida que regista os produtos que saíram do inventário
<i>UnitsBalance</i>	<i>int</i>	Medida que regista os produtos em inventário

Anexo III-Tabelas das Dimensões

Tabela 16- Descrição da Dimensão Sales Territory

Dimensão Sales Territory		
Atributo	Tipo	Resumo
<i>SalesTerritoryKey</i>	<i>int</i>	Chave primária
<i>SalesTerritoryAlternateKey</i>	<i>int</i>	Chave suplente
<i>SalesTerritoryRegion</i>	<i>nvarchar(50)</i>	Região territorial (Norte, sul, oeste, este)
<i>SalesTerritoryCountry</i>	<i>nvarchar(50)</i>	País
<i>SalesTerritoryGroup</i>	<i>nvarchar(50)</i>	Continente
<i>SalesTerritoryImage</i>	<i>varbinary(MAX)</i>	Imagem do território

Tabela 17 - Descrição da Dimensão Geography

Dimensão Geography		
Atributo	Tipo	Resumo
<i>GeographyKey</i>	<i>int</i>	Chave primária
<i>City</i>	<i>nvarchar(30)</i>	Cidade
<i>StateProvinceCode</i>	<i>nvarchar(3)</i>	Código da província do estado
<i>StateProvinceName</i>	<i>nvarchar(50)</i>	Nome da província do estado
<i>CountryRegionCode</i>	<i>nvarchar(3)</i>	Código da região
<i>EnglishCountryRegionName</i>	<i>nvarchar(50)</i>	Nome da região em inglês
<i>SpanishCountryRegionName</i>	<i>nvarchar(50)</i>	Nome da região em espanhol
<i>FrenchCountryRegionName</i>	<i>nvarchar(50)</i>	Nome da região em francês
<i>PostalCode</i>	<i>nvarchar(15)</i>	Código postal
<i>SalesTerritoryKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>Sales Territory</i>
<i>IpAddressLocator</i>	<i>nvarchar(15)</i>	Localização através de endereço de ip

Tabela 18 - Descrição da Dimensão Product

Dimensão Product		
Atributo	Tipo	Resumo
<i>ProductKey</i>	<i>int</i>	Chave primária
<i>ProductAlternateKey</i>	<i>nvarchar(25)</i>	Chave suplente
<i>ProductSubcategoryKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>Sub Category</i>
<i>WeightUnitMeasureCode</i>	<i>nchar(3)</i>	Código da unidade de medida do peso
<i>SizeUnitMeasureCode</i>	<i>nchar(3)</i>	Código da unidade de medida do tamanho
<i>EnglishProductName</i>	<i>nvarchar(50)</i>	Nome do produto em inglês
<i>SpanishProductName</i>	<i>nvarchar(50)</i>	Nome do produto em espanhol
<i>FrenchProductName</i>	<i>nvarchar(50)</i>	Nome do produto em francês
<i>StandardCost</i>	<i>money</i>	Custo
<i>FinishedGoodsFlag</i>	<i>bit</i>	Assinala se o produto é uma componente ou produto final
<i>Color</i>	<i>nvarchar(15)</i>	Cor do produto
<i>SafetyStockLevel</i>	<i>smallint</i>	Quantidade de stock de segurança necessária
<i>ReorderPoint</i>	<i>smallint</i>	Quantidade mínima de stock
<i>ListPrice</i>	<i>money</i>	Preço de venda
<i>Size</i>	<i>nvarchar(50)</i>	tamanho
<i>Weight</i>	<i>float</i>	Peso
<i>DaysToManufacture</i>	<i>int</i>	Número de dias necessários para produzir
<i>ProductLine</i>	<i>nchar(2)</i>	Linha de produção do produto
<i>DealerPrice</i>	<i>money</i>	Preço de compra
<i>Class</i>	<i>nchar(2)</i>	Classe de produto
<i>Style</i>	<i>nchar(2)</i>	Tipo
<i>ModelName</i>	<i>nvarchar(50)</i>	Modelo
<i>LargePhoto</i>	<i>varbinary(MAX)</i>	Imagem do produto
<i>EnglishDescription</i>	<i>nvarchar(400)</i>	Descrição em inglês
<i>FrenchDescription</i>	<i>nvarchar(400)</i>	Descrição em francês
<i>ChineseDescription</i>	<i>nvarchar(400)</i>	Descrição em chinês
<i>ArabicDescription</i>	<i>nvarchar(400)</i>	Descrição em árabe
<i>HebrewDescription</i>	<i>nvarchar(400)</i>	Descrição em hebraico

Atributo	Tipo	Resumo
<i>ThaiDescription</i>	<i>nvarchar(400)</i>	Descrição em tailandês
<i>GermanDescription</i>	<i>nvarchar(400)</i>	Descrição em alemão
<i>JapaneseDescription</i>	<i>nvarchar(400)</i>	Descrição em japonês
<i>TurkishDescription</i>	<i>nvarchar(400)</i>	Descrição em língua turca
<i>StartDate</i>	<i>datetime</i>	Dia de disponibilidade do produto
<i>EndDate</i>	<i>datetime</i>	Dia de indisponibilidade do produto
<i>Status</i>	<i>nvarchar(7)</i>	Estado

Tabela 19 - Descrição da Dimensão Product Category

Dimensão Product Category		
Atributo	Tipo	Resumo
<i>ProductCategoryKey</i>	<i>int</i>	Chave primária
<i>ProductCategoryAlternateKey</i>	<i>int</i>	Chave suplente
<i>EnglishProductCategoryName</i>	<i>nvarchar(50)</i>	Categoria do produto em inglês
<i>SpanishProductCategoryName</i>	<i>nvarchar(50)</i>	Categoria do produto em espanhol
<i>FrenchProductCategoryName</i>	<i>nvarchar(50)</i>	Categoria do produto em francês

Tabela 20 - Descrição da Dimensão Product SubCategory

Dimensão Product SubCategory		
Atributo	Tipo	Resumo
<i>ProductSubcategoryKey</i>	<i>int</i>	Chave primária
<i>ProductSubcategoryAlternateKey</i>	<i>int</i>	Chave suplente
<i>EnglishProductSubcategoryName</i>	<i>nvarchar(50)</i>	SubCategoria do produto em inglês
<i>SpanishProductSubcategoryName</i>	<i>nvarchar(50)</i>	SubCategoria do produto em espanhol
<i>FrenchProductSubcategoryName</i>	<i>nvarchar(50)</i>	SubCategoria do produto em espanhol
<i>ProductCategoryKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>Product Category</i>

Tabela 21 - Descrição da Dimensão Customer

Dimensão Customer		
Atributo	Tipo	Resumo
<i>CustomerKey</i>	<i>int</i>	Chave primária
<i>GeographyKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>Geography</i>
<i>CustomerAlternateKey</i>	<i>nvarchar(15)</i>	Chave suplente
<i>Title</i>	<i>nvarchar(8)</i>	Título de tratamento (Mr ou Miss)
<i>FirstName</i>	<i>nvarchar(50)</i>	Primeiro nome
<i>MiddleName</i>	<i>nvarchar(50)</i>	Segundo nome
<i>LastName</i>	<i>nvarchar(50)</i>	Apelido
<i>NameStyle</i>	<i>bit</i>	Tipo de nome
<i>BirthDate</i>	<i>date</i>	Data nascimento
<i>MaritalStatus</i>	<i>nchar(1)</i>	Estado civil
<i>Suffix</i>	<i>nvarchar(10)</i>	Sufixo
<i>Gender</i>	<i>nvarchar(1)</i>	Género
<i>EmailAddress</i>	<i>nvarchar(50)</i>	Endereço eletrónico
<i>YearlyIncome</i>	<i>money</i>	Salário anual
<i>TotalChildren</i>	<i>tinyint</i>	Número de filhos
<i>NumberChildrenAtHome</i>	<i>tinyint</i>	Número de filhos que são agregados familiares
<i>EnglishEducation</i>	<i>nvarchar(40)</i>	Nome da formação em inglês
<i>SpanishEducation</i>	<i>nvarchar(40)</i>	Nome da formação em espanhol
<i>FrenchEducation</i>	<i>nvarchar(40)</i>	Nome da formação em francês
<i>EnglishOccupation</i>	<i>nvarchar(100)</i>	Nome do emprego/cargo em inglês
<i>SpanishOccupation</i>	<i>nvarchar(100)</i>	Nome do emprego/cargo em espanhol
<i>FrenchOccupation</i>	<i>nvarchar(100)</i>	Nome do emprego/cargo em francês
<i>HouseOwnerFlag</i>	<i>nchar(1)</i>	Posse ou não de um imóvel
<i>NumberCarsOwned</i>	<i>tinyint</i>	Número de automóveis particulares
<i>AddressLine1</i>	<i>nvarchar(120)</i>	Endereço
<i>AddressLine2</i>	<i>nvarchar(120)</i>	Endereço
<i>Phone</i>	<i>nvarchar(20)</i>	Telefone
<i>DateFirstPurchase</i>	<i>date</i>	Data da primeira compra
<i>CommuteDistance</i>	<i>nvarchar(15)</i>	Distância da morada

Tabela 22 - Descrição da Dimensão Employee

Dimensão Employee		
Atributo	Tipo	Resumo
<i>EmployeeKey</i>	<i>int</i>	Chave primária
<i>ParentEmployeeKey</i>	<i>int</i>	Chave estrangeira da mesma dimensão, que referencia o responsável pelo funcionário
<i>EmployeeNationalIDAlternateKey</i>	<i>nvarchar(15)</i>	Chave suplente
<i>ParentEmployeeNationalIDAlternateKey</i>	<i>nvarchar(15)</i>	Chave suplente
<i>SalesTerritoryKey</i>	<i>int</i>	Chave estrangeira da dimensão Sales Territory. Indica o território de venda a que funcionário pertence
<i>FirstName</i>	<i>nvarchar(50)</i>	Primeiro nome
<i>LastName</i>	<i>nvarchar(50)</i>	Apelido
<i>MiddleName</i>	<i>nvarchar(50)</i>	Segundo nome
<i>NameStyle</i>	<i>bit</i>	Tipo de nome
<i>Title</i>	<i>nvarchar(50)</i>	Título de tratamento (<i>Mr</i> ou <i>Miss</i>)
<i>HireDate</i>	<i>date</i>	Data da contratação
<i>BirthDate</i>	<i>date</i>	Data de nascimento
<i>LoginID</i>	<i>nvarchar(256)</i>	Identificação do login para a rede interna
<i>EmailAddress</i>	<i>nvarchar(50)</i>	Endereço eletrônico
<i>Phone</i>	<i>nvarchar(25)</i>	Telefone
<i>MaritalStatus</i>	<i>nchar(1)</i>	Estado civil
<i>EmergencyContactName</i>	<i>nvarchar(50)</i>	Conctato de emergência
<i>EmergencyContactPhone</i>	<i>nvarchar(25)</i>	Conctado de emergência-telefone
<i>SalariedFlag</i>	<i>bit</i>	Assinalação se é assalariado ou não
<i>Gender</i>	<i>nchar(1)</i>	Género
<i>PayFrequency</i>	<i>tinyint</i>	Frequência de pagamento
<i>BaseRate</i>	<i>money</i>	Base salarial
<i>VacationHours</i>	<i>smallint</i>	Horas de férias
<i>SickLeaveHours</i>	<i>smallint</i>	Horas de dispensa médica
<i>CurrentFlag</i>	<i>bit</i>	Assinalação do estado de presença

Atributo	Tipo	Resumo
<i>SalesPersonFlag</i>	<i>bit</i>	Assinalação de cargo de vendas
<i>DepartmentName</i>	<i>nvarchar(50)</i>	Departamento
<i>StartDate</i>	<i>date</i>	Data de início de funções
<i>EndDate</i>	<i>date</i>	Data de termino de funções
<i>Status</i>	<i>nvarchar(50)</i>	Estado de contrato
<i>EmployeePhoto</i>	<i>varbinary(MAX)</i>	Fotografia

Tabela 23 - Descrição da Dimensão Reseller

Dimensão Reseller		
Atributo	Tipo	Resumo
<i>ResellerKey</i>	<i>int</i>	Chave primária
<i>GeographyKey</i>	<i>int</i>	Chave estrangeira da dimensão <i>Geography</i>
<i>ResellerAlternateKey</i>	<i>nvarchar(15)</i>	Chave suplente
<i>Phone</i>	<i>nvarchar(25)</i>	Telefone
<i>BusinessType</i>	<i>varchar(20)</i>	Tipo de negócio
<i>ResellerName</i>	<i>nvarchar(50)</i>	Nome
<i>NumberEmployees</i>	<i>int</i>	Número de funcionários
<i>OrderFrequency</i>	<i>char(1)</i>	Frequência de pedidos
<i>OrderMonth</i>	<i>tinyint</i>	Meses em de realização dos pedidos
<i>FirstOrderYear</i>	<i>int</i>	Primeiro ano de pedidos
<i>LastOrderYear</i>	<i>int</i>	Último ano de pedidos
<i>ProductLine</i>	<i>nvarchar(50)</i>	Linha de produtos
<i>AddressLine1</i>	<i>nvarchar(60)</i>	Endereço
<i>AddressLine2</i>	<i>nvarchar(60)</i>	Endereço
<i>AnnualSales</i>	<i>money</i>	Vendas anuais
<i>BankName</i>	<i>nvarchar(50)</i>	Nome do banco
<i>MinPaymentType</i>	<i>tinyint</i>	Tipo de quantia de pagamento
<i>MinPaymentAmount</i>	<i>money</i>	Quantia mínima de pagamento
<i>AnnualRevenue</i>	<i>money</i>	Receitas anuais
<i>YearOpened</i>	<i>int</i>	Ano de abertura

Tabela 24- Descrição da Dimensão Date

Dimensão Date		
Atributo	Tipo	Resumo
<i>DateKey</i>	<i>int</i>	Chave primária da dimensão
<i>FullDateAlternateKey</i>	<i>date</i>	Chave alternativa no formato de data (ex. 2018-01-20)
<i>DayNumberOfWeek</i>	<i>tinyint</i>	Número do dia da semana
<i>EnglishDayNameOfWeek</i>	<i>nvarchar(10)</i>	Nome do dia da semana em inglês
<i>SpanishDayNameOfWeek</i>	<i>nvarchar(10)</i>	Nome do dia da semana em espanhol
<i>FrenchDayNameOfWeek</i>	<i>nvarchar(10)</i>	Nome do dia da semana em francês
<i>DayNumberOfMonth</i>	<i>tinyint</i>	Número do dia no mês
<i>DayNumberOfYear</i>	<i>smallint</i>	Número do dia no ano
<i>WeekNumberOfYear</i>	<i>tinyint</i>	Número da semana no ano
<i>EnglishMonthName</i>	<i>nvarchar(10)</i>	Nome do mês em inglês
<i>SpanishMonthName</i>	<i>nvarchar(10)</i>	Nome do mês em espanhol
<i>FrenchMonthName</i>	<i>nvarchar(10)</i>	Nome do mês em francês
<i>MonthNumberOfYear</i>	<i>tinyint</i>	Número do mês no ano
<i>CalendarQuarter</i>	<i>tinyint</i>	Calendário trimestral
<i>CalendarYear</i>	<i>smallint</i>	Calendário anual
<i>CalendarSemester</i>	<i>tinyint</i>	Calendário semestral
<i>FiscalQuarter</i>	<i>tinyint</i>	Trimestre fiscal
<i>FiscalYear</i>	<i>smallint</i>	Ano fiscal
<i>FiscalSemester</i>	<i>tinyint</i>	Semestre fiscal

Tabela 25 - Descrição da Dimensão Currency

Dimensão Currency		
Atributo	Tipo	Resumo
<i>CurrencyKey</i>	<i>int</i>	Chave primária da dimensão
<i>CurrencyAlternateKey</i>	<i>nchar(3)</i>	Chave alternativa que especifica a Nomenclatura do tipo de moeda (ex.: AFA, EUR, USD)
<i>CurrencyName</i>	<i>nvarchar(50)</i>	Nome da moeda

Tabela 26- Descrição da Dimensão Promotion

Dimensão Promotion		
Atributo	Tipo	Resumo
<i>PromotionKey</i>	<i>int</i>	Chave primária
<i>PromotionAlternateKey</i>	<i>int</i>	Chave suplente
<i>EnglishPromotionName</i>	<i>nvarchar(255)</i>	Nome da promoção em inglês
<i>SpanishPromotionName</i>	<i>nvarchar(255)</i>	Nome da promoção em espanhol
<i>FrenchPromotionName</i>	<i>nvarchar(255)</i>	Nome da promoção em francês
<i>DiscountPct</i>	<i>float</i>	Porcentagem de desconto
<i>EnglishPromotionType</i>	<i>nvarchar(50)</i>	Tipo de promoção em inglês
<i>SpanishPromotionType</i>	<i>nvarchar(50)</i>	Tipo de promoção em espanhol
<i>FrenchPromotionType</i>	<i>nvarchar(50)</i>	Tipo de promoção em francês
<i>EnglishPromotionCategory</i>	<i>nvarchar(50)</i>	Categoria de promoção em inglês
<i>SpanishPromotionCategory</i>	<i>nvarchar(50)</i>	Categoria de promoção em espanhol
<i>FrenchPromotionCategory</i>	<i>nvarchar(50)</i>	Categoria de promoção em francês
<i>StartDate</i>	<i>datetime</i>	Data de início da promoção
<i>EndDate</i>	<i>datetime</i>	Data de término da promoção
<i>MinQty</i>	<i>int</i>	Quantidade mínima
<i>MaxQty</i>	<i>int</i>	Quantidade máxima