

Modelação Evolutiva de *Big Data Warehouses* em Automação Industrial para o Desenvolvimento Sustentável de Produto

Evolutionary Modelling of Big Data Warehouses in Industrial Automation for Sustainable Product Development

Mafalda Oliveira, Centro de Computação Gráfica, Centro de Investigação ALGORITMI, Universidade do Minho, Portugal, mafalda.oliveira@ccg.pt

Pedro Guimarães, Centro de Computação Gráfica, Centro de Investigação ALGORITMI, Universidade do Minho, Portugal, pedroguimaraes@ccg.cpt

Maribel Yasmina Santos, Centro de Investigação ALGORITMI, Universidade do Minho, Portugal, maribel@dsi.uminho.pt

Resumo

A indústria de automação vê-se atualmente forçada a lidar com grandes quantidades de dados. Porém, pode tornar-se difícil gerir dados com grandes volumes, heterogéneos, e provenientes de diferentes fontes, pelo que é importante dispor de metodologias e sistemas capazes de armazenar e gerir estes dados. Os *Big Data Warehouses* são sistemas de dados que vêm facilitar a gestão de grandes volumes de dados. No entanto, diversos desafios têm de ser endereçados, como a modelação destes repositórios. Este trabalho propõe uma metodologia para a modelação evolutiva de *Big Data Warehouses* que seja capaz de integrar novos processos de negócio e novos dados, fazendo evoluir um modelo já existente. Como caso de demonstração, a metodologia é aplicada no domínio da automação industrial para a proposta de um modelo de dados que suporte o armazenamento e processamento de indicadores analíticos para o desenvolvimento sustentável de produto.

Palavras-chave: *Big Data; Big Data Warehouses; Modelação Evolutiva; Modelação de Dados*

Abstract

The automation industry is currently forced to deal with large amounts of data. However, it can be difficult to manage large volumes of heterogeneous data from different sources, so it is important to have methodologies and systems capable of storing and managing these data. Big Data Warehouses are data systems that facilitate the management of big data, but, as a consequence, other challenges arise, such as data modelling issues. This research proposes a methodology for data models that can evolve with time to integrate new business processes and new data. As demonstration case, the methodology is applied to the automation industry to propose a data model that supports the storage and processing of analytical indicators for sustainable product development.

Keywords: *Big Data; Big Data Warehouses; Evolutionary Modelling; Data Modelling*

1. INTRODUÇÃO

A automação industrial é definida como um conjunto de técnicas que recorrem ao uso da tecnologia, com o objetivo de tornar a realização de tarefas automática. Este conceito vem substituir o esforço humano por elementos eletromecânicos computáveis (Silveira & Lima, 2003).

A evolução global neste ramo da automação industrial impõe às empresas um ritmo de adaptação elevado. De forma a conseguir acompanhar esta evolução e estar ao nível da concorrência, é importante que as organizações encontrem formas de trabalhar capazes de integrar tecnologias de informação que suportam e otimizam os processos, como é o caso dos *Data Warehouses* (DW). Contudo, as características que surgem subjacentes aos dados, como a velocidade, a variedade, a veracidade, ou o volume, desafiam a capacidade dos DW de lidar com a integração de novos dados e processos de negócio. Para ultrapassar esta dificuldade, surgem novas abordagens, conceitos e tecnologias que estendem e otimizam os DW ao longo do seu ciclo de vida.

A capacidade de recolher, armazenar, processar e analisar grandes quantidades de dados provenientes de diferentes fontes, com diferentes taxas de transmissão, é possível graças aos avanços tecnológicos que se têm vindo a reconhecer nos sistemas *Big Data* (Santos et al., 2019). Dentro desta área, tem também crescido a importância do conceito de *Big Data Modelling* (BDM). Deve-se garantir uma correta abordagem de modelação dos dados, para conseguir retirar o devido valor dos mesmos (Patel, 2019).

Um passo fundamental na fase de modelação de grandes quantidades de dados é a formalização dos conceitos. Esta formalização deve ser desenvolvida como sendo um esquema que descreve os dados (metamodelo) e dá informação de como os dados se relacionam. Como o mundo organizacional evolui e, conseqüentemente, evoluem também os processos organizacionais, este esquema deve ser capaz de evoluir e de suportar os novos requisitos que advêm da evolução dos processos. Portanto, é importante que este esquema de dados seja abstrato o suficiente para facilitar esta evolução (Hartmann & Ma, 2016).

Neste artigo será proposta uma metodologia para a modelação de um *Big Data Warehouse* (BDW) baseada no conceito de objetos analíticos, constituindo uma evolução dos conceitos habitualmente usados em contextos de DW, como por exemplo tabelas de factos e de dimensão. Esta metodologia é aplicada à área de indústrias de automação para auxiliar o desenvolvimento sustentável do produto, que seja capaz de integrar novos processos de negócio e novos dados. A sustentabilidade associada ao desenvolvimento do produto está orientada para a área económica e ambiental, na medida em que promove uma redução de custos e emissões de gases. Na abordagem metodológica será proposto um modelo conceptual de domínio no âmbito das indústrias de automação, ou seja, serão identificados conceitos fundamentais na área e definidas relações entre os mesmos, o que permitirá uma descrição concreta daquilo que é esta área de atuação (Weysow et al., 2022). Este modelo conceptual poderá

depois ser instanciado em diferentes modelos lógicos que dependem da implementação física num determinado sistema de dados, como sejam Bases de Dados relacionais, DW, *Big Data Warehouses* ou até mesmo *Data Lakes*.

Sendo o presente trabalho inserido no contexto de Big Data, é necessário lidar com a elevada quantidade de dados que deve ser armazenada e processada (Jukić et al., 2015) e com a velocidade com que estes dados têm de ser recolhidos e analisados (Sivarajah et al., 2017), entre outros aspetos. Tendo isto em conta, um BDW apresenta um modelo lógico que consegue dar resposta a estes desafios, conciliando o armazenamento e o processamento de dados estruturados, semiestruturados ou não estruturados, provenientes de várias fontes. Estes sistemas de dados são capazes de processar grandes volumes de dados, ao contrário das ferramentas tradicionais de captura, armazenamento, gestão e análise de dados (Santoso & Yulia, 2017).

Para auxiliar o desenvolvimento deste trabalho foi seguida a metodologia proposta por (Peffer et al., 2007), a *Design Science Research Methodology for Information Systems*. Esta metodologia, entre outros, orientou o processo de exploração de conceitos e trabalhos realizados na área, assim como a proposta da abordagem metodológica para a modelação evolutiva de um *Big Data Warehouse* para a indústria de automação.

Este artigo está organizado da seguinte forma. A secção 2 aborda trabalhos relacionados com o tema em estudo, modelação de dados e características evolutivas dos mesmos. A secção 3 apresenta e explica a metodologia desenvolvida para a modelação evolutiva de dados. A secção 4 resume o trabalho realizado e apresenta propostas de trabalho futuro.

2. TRABALHO RELACIONADO

Os autores (Galvão et al., 2020) apresentam uma proposta de modelação de BDW propondo um método baseado em regras e padrões de conceção, que visa reunir a informação de um determinado domínio de aplicação mapeado num modelo conceptual relacional. Este trabalho foca-se na categorização de cada elemento do modelo como sendo um tipo de objeto específico, e disponibiliza informação que orienta a conceção do modelo de dados. Porém, tal abordagem não responde às necessidades evolutivas dos modelos de dados, não estando preparado para evoluir o modelo quando estiverem disponíveis novos processos de negócio, fontes de dados ou requisitos analíticos.

O trabalho dos autores (Galvão et al., 2022) estende o trabalho anterior propondo uma abordagem estruturada para a conceção e implementação de BDW, focada na modelação de objetos autónomos, tendo em conta questões de desempenho, que integram os dados relevantes para responder a questões analíticas específicas. A estes objetos dá-se o nome de objetos analíticos que podem incluir atributos analíticos e descritivos. Estes objetos podem ser integrados com objetos analíticos complementares, para partilhar dados entre vários objetos analíticos, objetos especiais, para normalizar atributos

comuns como a data, a hora, ou atributos espaciais, e objetos materializados, para implementar fisicamente vistas que melhoram o desempenho.

Os autores (Hartmann & Ma, 2016) propõem um processo para conduzir a evolução de um esquema de base de dados, alargando a modularização da base de dados, de forma a trabalhar de um modo evolutivo. O processo evolutivo de modularização da base de dados é executado durante a modelação conceptual, melhorando, assim, a capacidade de abstração do esquema de dados gerado.

Os autores (Tria et al., 2018) propõe uma metodologia para o *design* de um BDW que adopta abordagens ágeis e automáticas, com o intuito de reduzir o tempo necessário para integrar novos processos de negócio e dados. A metodologia é composta por cinco passos: análise de requisitos, onde se faz um levantamento e representação de requisitos; integração das fontes, onde se produz um esquema integrado das fontes de dados, utilizando uma abordagem ontológica; desenho conceptual, que representa conceitos multidimensionais utilizando um esquema baseado em árvores; desenho lógico, que se baseia nos documentos multidimensionais utilizando a lógica do modelo chave-valor; desenho físico, que visa implementar os metadados necessários expostos pelo DW.

Todos os trabalhos relacionados descritos apresentam contributos importantes na área da modelação dos dados, porém, nenhum apresenta uma metodologia de modelação de dados com características evolutivas robustas e concretas. A característica evolutiva proposta no presente trabalho está descrita na secção 3, onde se apresentam as possíveis evoluções que o modelo de dados pode sofrer, de forma a tornar possível a integração de diferentes processos da mesma área. Como caso de demonstração, é exemplificada a automação industrial, mas a abordagem pode ser utilizada em qualquer domínio de aplicação.

3. MODELAÇÃO EVOLUTIVA DE *BIG DATA WAREHOUSES*

A metodologia de modelação de dados aqui proposta teve por base o trabalho realizado pelos autores (Galvão et al., 2022), mas estende este trabalho de forma a incorporar as características evolutivas, sendo capaz de integrar novos processos de negócio e dados. É apresentado um conjunto de passos, sistematizado na Figura 1, que deve ser seguido de forma sequencial para assegurar uma correta aplicação da metodologia.

No que diz respeito às características evolutivas da presente abordagem, sempre que forem identificados novos processos de negócio ou dados, existe a possibilidade de integrá-los no modelo de dados base, desenvolvido com base na metodologia proposta. A evolução do modelo de dados de um BDW pode ser conseguida através das seguintes possibilidades:

Adicionar Atributos Analíticos e/ou Descritivos

Uma das possíveis evoluções que o modelo de dados pode sofrer está relacionada com a adição de atributos analíticos e/ou descritivos ao mesmo. Se esta necessidade se verificar é importante ter em atenção um conjunto de aspetos. Se existir a necessidade de adicionar atributos analíticos ao modelo de dados, é necessário manter a coerência dos dados históricos já armazenados no BDW e a análise dos mesmos. Para tal, os registos já existentes, e para os quais não estarão disponíveis os novos atributos analíticos agora recolhidos, devem integrar o valor *null* ou equivalente, de forma a não influenciar qualquer função de agregação (COUNT, SUM, AVG, ...) realizada sobre os dados. No que diz respeito a novos atributos descritivos, registos anteriores devem ser preenchidos com a etiqueta “Não disponível”, “Desconhecido” ou similar.

Adicionar Famílias Analíticas e/ou Descritivas

Outra possível evolução que o modelo de dados pode sofrer é a adição de famílias analíticas e/ou descritivas aos objetos que constituem o modelo. Neste caso, e como estamos a referir conjuntos de atributos analíticos ou descritivos, deve ser aplicado o mesmo princípio utilizado para adicionar atributos. A diferença prende-se com o facto de as famílias serem constituídas por um conjunto de atributos, pelo que, quando se pretende adicionar uma família analítica deve ter-se em atenção todos os atributos que a constituem. Se estes atributos apenas registam valores a partir de um certo período de tempo, todos os registos até então devem ser preenchidos com valores nulos ou similares. Se se estiver a adicionar uma família descritiva na qual os atributos fornecem informação a partir de um determinado período no tempo, os registos até à data devem ser preenchidos com o valor “Não disponível”, “Desconhecido” ou similar.

Adicionar Novos Objetos

É ainda possível adicionar novos objetos ao modelo de dados, nomeadamente Objetos Analíticos, Objetos Analíticos Complementares, Objetos Especiais e/ou Objetos Materializados. Nesta situação, é importante ter em atenção uma série de fatores. Inicialmente, é importante perceber que tipo de objeto se pretende adicionar. Se o objeto for relativo a tempo ou espaço é automaticamente considerado um objeto especial. Senão, deve ser seguida a metodologia proposta de forma a perceber que tipo de objeto deve ser adicionado. Seguindo a metodologia, se o objeto resultar de um assunto analítico será considerado um Objeto Analítico. Se o objeto resultar do mapeamento entre uma entidade e vários assuntos analíticos, este vai dar origem a um Objeto Analítico Complementar. Se o objeto for utilizado para dar resposta a necessidades específicas do utilizador, incluindo - por exemplo - uma agregação dos registos de Objetos Analíticos e/ou de Objetos Analíticos Complementares, é considerado um objeto materializado.

Independentemente do tipo de objeto que se pretende adicionar, os novos relacionamentos estabelecidos entre os objetos devem garantir que os dados históricos mantêm a sua coerência, pelo que deverá ser equacionada a necessidade de criação de registos ou valores que sinalizem estas

alterações. Tal permitirá manter a coerência entre todos os objetos do modelo. Nesta situação de adição de objetos é fundamental ter em atenção a granularidade do modelo, e de cada objeto em particular, e assegurar que os novos objetos respeitam a granularidade dos restantes objetos do modelo de dados.

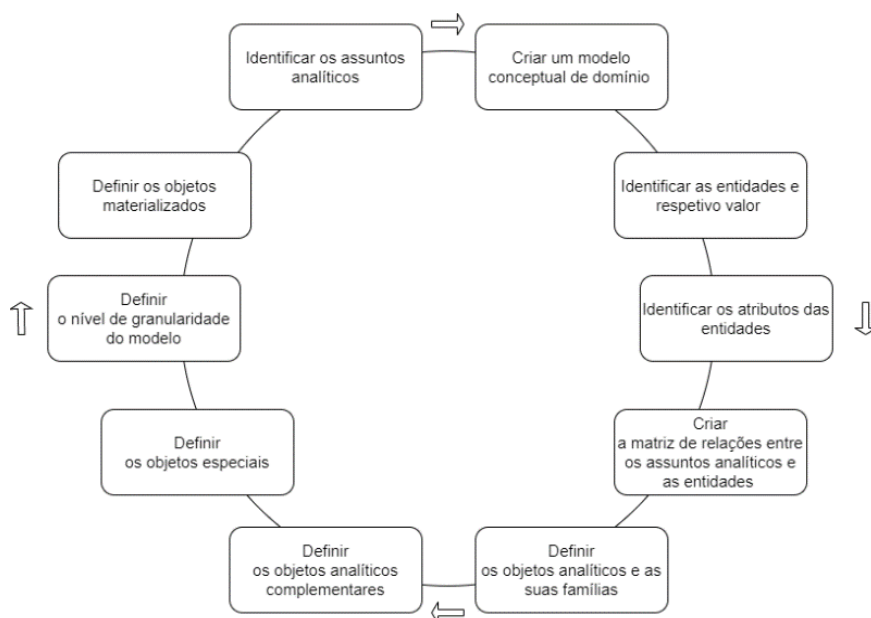


Figura 1 - Passos da Metodologia de Modelação de Dados

Passo 1 – Identificar os assuntos analíticos

Este primeiro passo tem como objetivo a identificação dos temas a analisar, aquilo que se pretende estudar, tendo em conta os objetivos analíticos e questões analíticas a que se quer dar resposta. É fundamental para este passo a perceção do negócio, o entendimento sobre aquilo que se está a estudar, de forma a facilitar a identificação dos assuntos relevantes. Essencialmente, estamos a referir os processos de negócio a analisar.

O resultado deste primeiro passo é uma lista de assuntos analíticos de interesse no contexto em análise. Tendo em conta que este trabalho se insere no ramo da automação industrial voltada para o desenvolvimento sustentável do produto, é importante analisar aquilo que está relacionado com este processo. Após contacto com algumas indústrias de automação, foi possível perceber o contexto de trabalho e os fatores que influenciam o desenvolvimento do produto. Neste domínio de aplicação é importante estudar dados relativos às máquinas responsáveis pelo processo produtivo. Estes dados podem ser referentes a consumos da máquina ou a possíveis defeitos que tenham sido registados.

Neste sentido, os assuntos analíticos identificados são: gestão de sensores e gestão de defeitos. Estes assuntos analíticos são os principais conceitos que podem influenciar a sustentabilidade associada ao processo de desenvolvimento de produtos.

Num contexto evolutivo, onde há necessidade de inserir novos assuntos analíticos como consequência, por exemplo, de um novo processo de negócio emergente, é fundamental ter este passo 1 em consideração. Assim, e de forma a evoluir o modelo com o aparecimento de um novo processo de negócio, é importante identificar os assuntos analíticos decorrentes desse processo. Neste caso de demonstração, identifica-se um novo processo de negócio relacionado com a Gestão de Desperdícios. Assim sendo, e de forma a seguir a metodologia de modelação proposta, foi identificado um novo assunto analítico para dar resposta ao novo processo de negócio: Gestão de Desperdícios.

Passo 2 – Definir o modelo conceptual do domínio

Nesta fase, deve ser desenvolvido um modelo conceptual de domínio, sob a forma de diagrama de classes, diagrama de entidades e relacionamentos, ou outra abordagem similar. É fundamental que neste modelo estejam representadas as principais entidades que caracterizam os processos de negócio e os relacionamentos entre as mesmas.

As entidades são facilmente identificadas depois da definição dos assuntos analíticos, visto que acabam por ser uma representação dos conceitos importantes relacionados com os mesmos.

O modelo conceptual de domínio inicialmente proposto para o contexto de análise em estudo, ou seja, para as indústrias de automação preocupadas com a sustentabilidade aplicada ao desenvolvimento do produto, encontra-se representado na Figura 2.

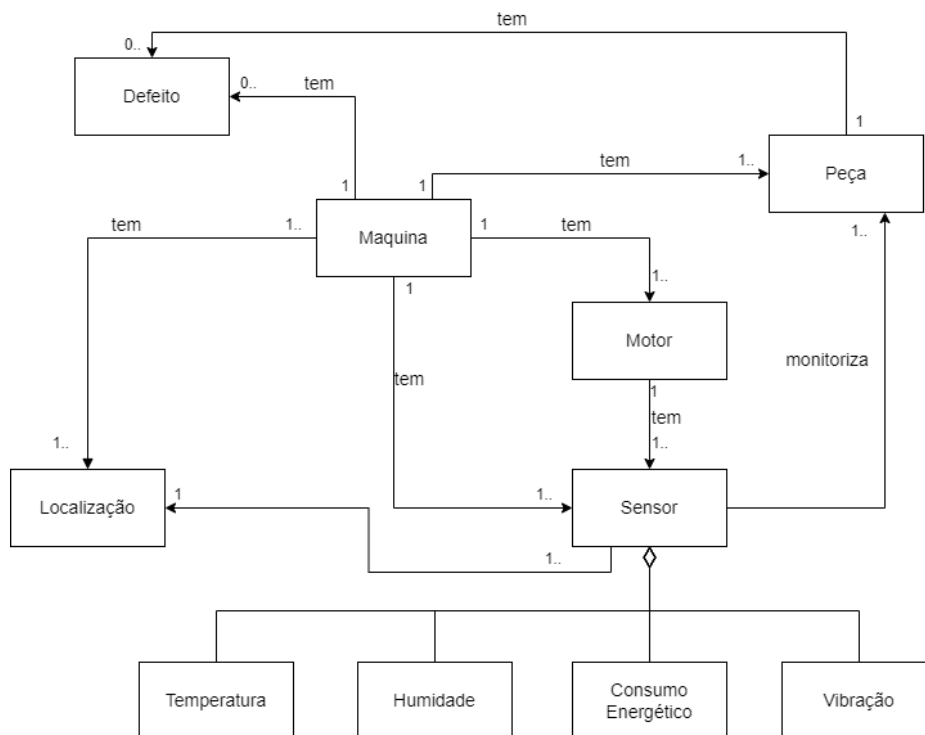


Figura 2 - Modelo Conceptual de Domínio

Os conceitos gerais associados a este modelo de dados foram representados utilizando classes e relacionamentos entre as mesmas. Neste ramo da indústria de automação, as máquinas são constituídas por motores e sensores, onde os próprios motores podem ter também sensores instalados. As máquinas podem ter localizações associadas, da mesma forma que, numa localização, podem encontrar-se várias máquinas. As máquinas e as peças podem apresentar defeitos e, por norma neste contexto, os sensores avaliam fatores como: temperatura, humidade, consumo energético e vibração, como indicadores relevantes para o desenvolvimento de um produto mais sustentável. Ainda associado aos sensores, estes podem ter uma localização associada e monitorizar peças.

Com o aparecimento do novo processo de negócio ligado à gestão de desperdícios e, de forma a evoluir o modelo conceptual de domínio base, devem ser acrescentadas as entidades relacionadas com este processo de negócio e representar as suas relações com as restantes entidades no modelo. Assim, é possível obter um modelo conceptual de domínio que representa uma nova realidade, representado na Figura 3.

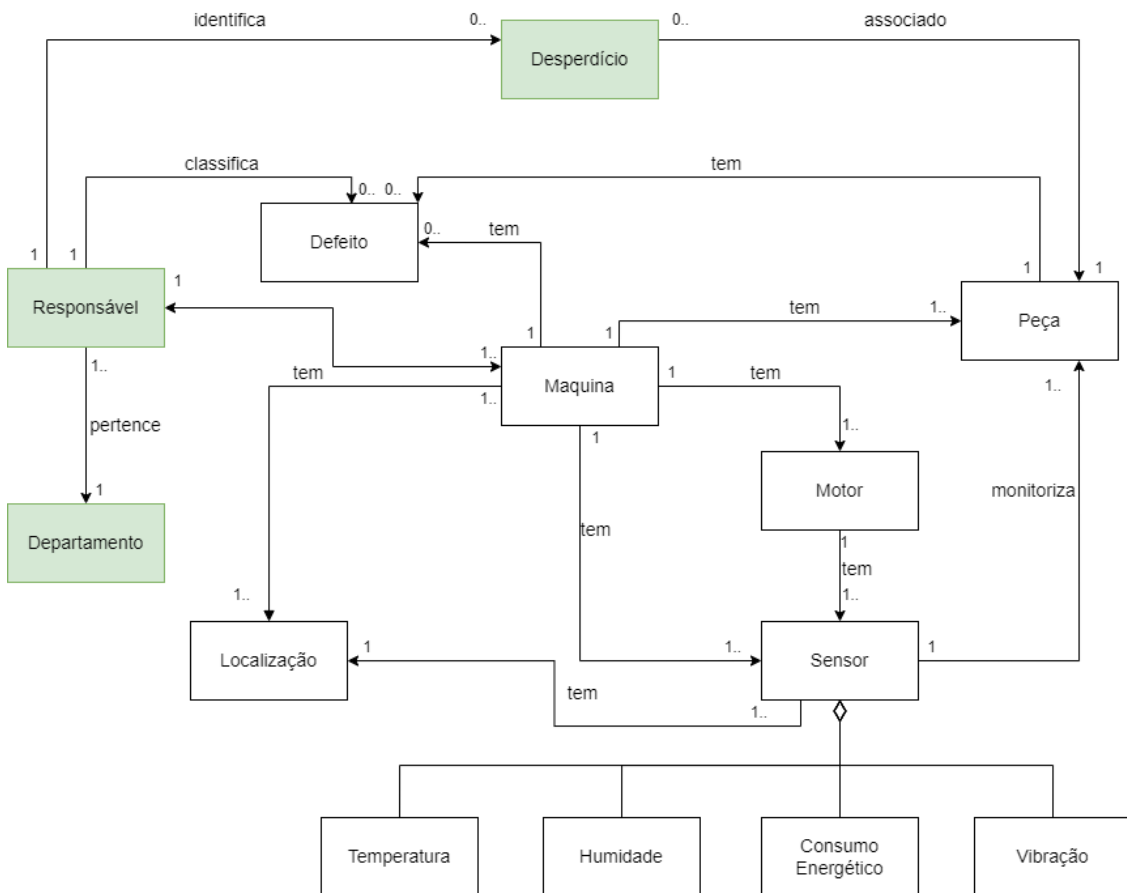


Figura 3 - Novo Modelo Conceptual de Domínio

O modelo conceptual de domínio atualizado integra o novo assunto analítico identificado no passo 1, Gestão de Desperdícios, e relaciona as entidades que o caracterizam com as entidades que já tinham sido identificadas para os restantes assuntos analíticos referentes aos restantes processos de negócio identificados anteriormente. Para representar as relações entre o novo assunto analítico e os já existentes houve a necessidade de adicionar ao modelo duas entidades: Responsável e Departamento. Desta forma, é possível perceber que cada máquina é da responsabilidade de um funcionário que pertence a um determinado departamento e que cada responsável tem o papel de identificar os desperdícios de peças, sempre que existirem, e de classificar os defeitos.

Passo 3 - Identificar o valor de cada entidade

As entidades identificadas neste trabalho foram: Máquina, Motor, Sensor (Temperatura, Humidade, Consumo Energético e Vibração), Defeito, Peça e Localização. Para cada entidade é necessário perceber qual o seu valor, analítico ou descritivo. Para identificar entidades com valor analítico recorre-se à premissa de que as entidades com valor analítico têm duas características: (1) fornecem os principais indicadores empresariais ou analíticos necessários para o apoio à decisão; (2) têm uma elevada cardinalidade, uma vez que a taxa de crescimento das linhas é consideravelmente superior à verificada nas entidades com valor descritivo. As entidades com valor descritivo são utilizadas para fornecer contexto aos indicadores analíticos e permitir diferentes níveis de detalhe na análise.

Neste caso, as entidades Máquina, Motor, Localização e Peça proporcionam um valor descritivo, enquanto que as entidades Sensor e Defeito proporcionam valor analítico.

Quando evoluído, o modelo conceptual de domínio incorpora novas entidades. Destas entidades, é possível atribuir à entidade Desperdício valor analítico e às entidades Responsável e Departamento valor descritivo.

Passo 4 - Identificar os atributos das entidades

Para cada entidade com valor analítico ou descritivo, é importante identificar a informação que a caracteriza, nomeadamente os seus atributos. Os atributos de cada umas das entidades identificadas, posteriormente agrupados em famílias analíticas ou descritivas, estão representadas no modelo de dados final, que se encontra representado na Figura 6.

Passo 5 - Criar a matriz de relações entre os assuntos analíticos e as entidades

Com base no conhecimento do domínio, a matriz de relações precisa de ser definida ou atualizada, mapeando cada tema identificado, ou seja, cada assunto analítico identificado no primeiro passo com as entidades identificadas no segundo passo.

	Assunto Analítico 1	Assunto Analítico 2
Entidade 1	X	
Entidade 2	X	X
Entidade 3		X
Entidade 4		X

Figura 4 - Matriz de Relações

Depois de mapeados os relacionamentos, o passo seguinte, isto é, o passo 6, fica facilitado. Os assuntos analíticos identificados podem dar origem a Objetos Analíticos e as entidades podem dar origem às famílias (analíticas ou descritivas) dos objetos. No caso do exemplo representado na Figura 4, o Assunto Analítico 2 pode dar origem a um Objeto Analítico, cujas famílias vão corresponder às Entidades 2, 3 e 4. A Entidade 2, sendo partilhada pelos Assuntos Analíticos 1 e 2, pode ser considerada como um Objeto Analítico Complementar, tal como explicado no passo 7. Se for, passa a ser um objeto autónomo. Caso contrário, passa a existir no Assunto Analítico 1 e no Assunto Analítico 2. Dependendo da classificação atribuída a cada entidade (passo 3), valor analítico ou descritivo, as entidades dão origem a famílias descritivas ou analíticas.

No contexto do presente trabalho, a matriz de relações encontra-se representada na Tabela 1.

ENTIDADE	GESTÃO DE SENSORES	GESTÃO DE DEFEITOS
Máquina	X	X
Motor	X	
Sensor de Temperatura	X	
Sensor de Consumo Energético	X	
Sensor de Humidade	X	
Sensor de Vibração	X	
Peça	X	X
Localização	X	
Defeito		X

Tabela 1 - Matriz de Relações em contextos de automação industrial

Para integrar o novo processo de negócio, Gestão de Desperdícios, neste passo devem ser integrados os novos assuntos analíticos e entidades que lhe deram origem. Posto isto, na Tabela 2 em baixo apresentada foram adicionadas as novas entidades e o novo assunto analítico identificado. Assim, é perceptível a relação e integração dos mesmos com os assuntos analíticos e entidades que já existiam.

ENTIDADE	GESTÃO DE SENSORES	GESTÃO DE DEFEITOS	GESTÃO DE DESPERDÍCIO
Máquina	X	X	
Motor	X		
Sensor de Temperatura	X		
Sensor de Consumo Energético	X		
Sensor de Humidade	X		
Sensor de Vibração	X		
Peça	X	X	X
Localização	X		
Defeito		X	
Responsável		X	X
Departamento			X
Desperdício			X

Tabela 2 - Matriz de Relações com a evolução para contextos de automação industrial

Com o intuito de criar um modelo de dados capaz de ser utilizado por várias organizações na área da indústria de automação, as organizações que o pretendem utilizar devem selecionar da matriz de relações apenas os assuntos e as entidades com relevância analítica nos seus contextos.

Passo 6 - Definir os objetos analíticos e as suas famílias

Os Objetos Analíticos são considerados objetos de interesse para fins analíticos. São estruturas altamente desnormalizadas e autónomas que são capazes de responder a consultas sem a constante necessidade de juntar tabelas. Como já foi referido, através da matriz de relações obtida no passo 5, é possível identificar quais os objetos analíticos existentes no modelo. Estes objetos podem integrar um conjunto de assuntos analíticos, dois ou mais, ou resultar de um único assunto, dependendo do contexto ao qual se está a aplicar a metodologia, cabendo ao engenheiro de dados decidir sobre os diferentes assuntos analíticos e a sua organização nos diversos objetos analíticos. A metodologia proposta auxilia nesta tarefa, mas não substitui o engenheiro de dados na proposta do modelo final. Neste trabalho, os objetos analíticos que resultam da matriz de relações representada na tabela 1 são: Sensor e Defeito.

Depois de evoluir o modelo conceptual de domínio e a matriz de relações, é identificado um novo objeto analítico: Desperdício. Este objeto analítico vai ser constituído por duas famílias descritivas: Funcionário e Departamento e por uma família analítica: Desperdício.

Passo 7 - Definir os objetos analíticos complementares

Um Objeto Analítico Complementar é um objeto que, como o próprio nome indica, complementa outros objetos analíticos, fornecendo uma estrutura autónoma com valor analítico que é utilizada para complementar as diferentes perspetivas analíticas fornecidas pelos objetos analíticos. De forma a identificar e criar objetos analíticos complementares devem ser seguidas um conjunto de regras que confirmem se é possível criar um objeto analítico complementar (Costa, 2019):

1. A família descritiva é incluída frequentemente noutros objetos analíticos;
2. A família descritiva tem baixa cardinalidade, uma vez que é importante que o objeto analítico complementar não ocupe muito espaço em memória de forma a melhorar o desempenho do modelo de dados;
3. A frequência da ingestão dos dados do objeto analítico complementar resultante é equivalente à frequência de ingestão dos dados dos objetos analíticos com os quais se relaciona, sendo que esta frequência pode variar de projeto para projeto, podendo ser semanal, mensal, anual ou até mesmo em tempo real;
4. O objeto analítico complementar é capaz de fornecer valor analítico por si mesmo;
5. Os registos do objeto analítico complementar devem ser imutáveis.

É importante referir que um Objeto Analítico Complementar deve emergir de entidades que proporcionem valor descritivo, mas que pode incluir características analíticas (ponto 4).

Se este conjunto de pontos se verificar, é então possível transformar o objeto identificado num objeto analítico complementar e, assim, melhorar o desempenho do modelo.

Através da matriz de relações é possível identificar as entidades que são frequentemente incluídas em vários objetos analíticos, como é o caso da Máquina e Peça. Como as entidades Máquina e Peça cumprem os requisitos 1, 2, 3, 4 e 5, enunciados acima, podem ser considerados objetos analíticos complementares.

No caso da evolução, o objeto analítico Desperdício não cumpre os requisitos descritos em cima, pelo que não se verifica a necessidade de definir novos objetos analíticos complementares.

Passo 8 - Definir os objetos especiais

Os objetos especiais são constituídos por atributos temporais e/ou espaciais, como é o caso do calendário, atributos descritivos temporais ou espaciais relevantes no domínio da aplicação. É necessário um cuidado especial nos objetos temporais e espaciais. Estes objetos são utilizados para normalizar este tipo de informação no BDW. Não se recomenda a utilização de atributos com elevada granularidade (ou seja, segundos, latitude ou longitude), uma vez que estes aumentarão muito o número de registos nestas tabelas (Galvão et al., 2022).

No trabalho desenvolvido, a entidade localização pode ser considerada um objeto especial.

No que toca à nova entidade, Desperdício, esta não cumpre os requisitos para ser considerado objeto especial.

Passo 9 - Definir o nível de granularidade do modelo

O nível de granularidade representa o nível de detalhe que os registos a armazenar num Objeto Analítico devem verificar. O nível de granularidade é definido recorrendo a um ou mais atributos descritivos que identificam de forma única um registo. Estes atributos são identificados através da sigla GK, *Granularity Key*. De notar que os mesmos podem, ou não, ser fisicamente implementados num sistema de dados como uma chave primária (Galvão et al., 2022).

Como resultado dos passos descritos é possível identificar um modelo de dados, representado na Figura 5, capaz de integrar toda a informação relevante para obter um desenvolvimento de produtos mais sustentável na área da automação industrial.

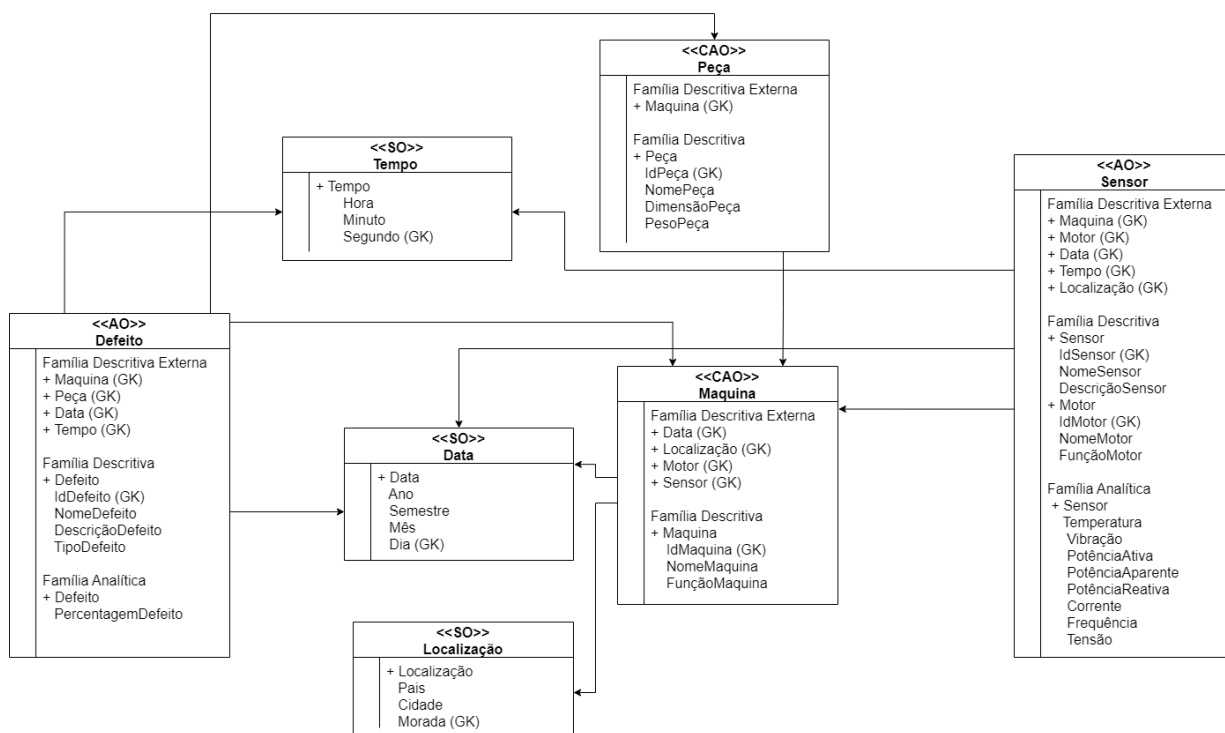


Figura 5 - Modelo de Dados

No modelo de dados obtido são perceptíveis os objetos, as relações entre os mesmos e os indicadores analíticos presentes no processo de desenvolvimento do produto sustentável. Neste contexto da automação industrial, os objetos analíticos identificados têm associados um conjunto de indicadores que auxiliam no processo de tomada de decisão. No caso do Objeto Analítico Defeito é possível retirar informações, como por exemplo, sobre a percentagem de defeito, isto é, se a peça está completamente inoperacional (100%), parcialmente operacional (50%), o tipo de defeito associado, a descrição do defeito, entre outros. O Objeto Analítico Sensor armazena informação sobre

a melhor realização de consultas frequentes, realizando uma pré-agregação dos dados e as junções pré-computacionais demoradas entre objetos de grande dimensão.

Os objetos materializados são úteis e eficientes quando há a necessidade de integrar diversos indicadores de objetos analíticos. Os objetos materializados são capazes de armazenar os resultados de um processo de consultas, aumentando o desempenho do sistema *Big Data Warehousing*, uma vez que vários consumidores de dados podem consumir este objeto materializado muito mais rapidamente do que os objetos analíticos com elevado nível de detalhe.

Este passo deve ser desenvolvido quando uma organização necessita de recolher informações específicas de diferentes objetos analíticos. Veja-se, a título exemplificativo, uma empresa de automação industrial que recolha dados relativos à temperatura, à humidade e à percentagem de defeito. Para este caso, a empresa pode criar um objeto materializado, que apenas vai armazenar essa informação, organizada em diferentes famílias analíticas.

De notar que, se após várias evoluções do modelo de dados se verificar que um objeto deixa de ser refrescado com dados recentes, esse objeto poderá posteriormente ser classificado como sendo um “objeto descontinuado”. No entanto, estes aspetos devem ser aprofundados e devem ser detalhados os princípios que ditam a descontinuação de um objeto ou a sua remoção, nos casos em que o histórico não precise ser mantido.

4. CONCLUSÃO E TRABALHO FUTURO

Atualmente é possível melhorar os processos de negócio das organizações com análises corretas aos dados, no entanto, é importante saber trabalhar em contextos de grandes quantidades de dados. É fundamental uma correta abordagem de modelação de dados em contextos de *Big Data* para conseguir tirar um melhor proveito aos dados. Neste trabalho foi proposta uma metodologia para a modelação de dados, em contextos de grandes volumes de dados, capaz de evoluir no sentido de possibilitar e facilitar a integração de novos processos de negócio e dados. Nesta metodologia foi seguido um conjunto de dez passos para identificar o modelo de dados da área da automação industrial, com o intuito de auxiliar o desenvolvimento sustentável dos produtos, e foram ainda apresentados os possíveis caminhos para evoluir o modelo de dados, sustentados com alguns exemplos na área da indústria da automação. A aplicação da metodologia em contextos de automação industrial impactou diferentes aspetos. Através da aplicação da metodologia foi possível otimizar o processamento de queries, com acessos aos dados mais diretos e organizados. Como consequência deste melhor desempenho e organização, foi possível identificar facilmente impactos nos valores associados a consumos energéticos e ambientais, resultando numa mudança em alguns processos produtivos de indústrias de automação, impactando a sustentabilidade associada a estes

contextos. Esta metodologia resulta numa mais valia em contextos empresariais, na medida em que a gestão e organização dos dados vê-se facilitada.

Como trabalho futuro pode aprofundar-se o conceito de objeto descontinuado e perceber como identificar objetos deste tipo e trabalhar com os mesmos.

AGRADECIMENTOS

Este trabalho foi suportado pela FCT – Fundação para a Ciência e Tecnologia, no âmbito das unidades de I&D, Centro ALGORITMI, Projeto UIDB/00319/2020. Foi elaborado com o apoio do Centro de Computação Gráfica, no contexto do projeto PRODUTECH4S&C, referência POCI-01-0247-FEDER-046102.

REFERÊNCIAS BIBLIOGRÁFICAS

- Costa, C. (2019). *Advancing the Design and Implementation of Big Data Warehousing Systems* (Issue April), PhD Thesis, University of Minho.
- Galvão, J., Leon, A., Costa, C., Santos, M. Y., & López, Ó. P. (2020). Towards Designing Conceptual Data Models for Big Data Warehouses: The Genomics Case. *Lecture Notes in Business Information Processing*, 402, 3–19. https://doi.org/10.1007/978-3-030-63396-7_1
- Galvão, J., Ribeiro, D., Machado, I., Ferreira, F., Gonçalves, J., Faria, R., Moreira, G., Costa, C., Cortez, P., & Santos, M. (2022). *Bosch 's Industry 4.0 Advanced Data Analytics: Historical and Predictive Data Integration for Decision Support*. In International Conference on Research Challenges in Information Science (pp. 580-596). Springer, Cham.
- Hartmann, S., & Ma, H. (2016). Evolutionary Database Design: Enhancing Data Abstraction Through Database Modularization to Achieve Graceful Schema Evolution. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9828 LNCS, V–VI. <https://doi.org/10.1007/978-3-319-44403-1>
- Jukić, N., Sharma, A., Nestorov, S., & Jukić, B. (2015). Augmenting Data Warehouses with Big Data. *Information Systems Management*, 32(3), 200–209. <https://doi.org/10.1080/10580530.2015.1044338>
- Patel, J. (2019). An Effective and Scalable Data Modeling for Enterprise Big Data Platform. *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 2691–2697. <https://doi.org/10.1109/BigData47090.2019.9005614>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Santos, M. Y., Costa, C., Galvão, J., Andrade, C., Pastor, O., & Marcén, A. C. (2019, June). Enhancing Big Data Warehousing for Efficient, Integrated and Advanced Analytics. In International Conference on Advanced Information Systems Engineering (pp. 215-226). Springer, Cham.
- Santos, L. W., & Yulia. (2017). Data Warehouse with Big Data Technology for Higher Education. *Procedia Computer Science*, 124, 93–99. <https://doi.org/10.1016/j.procs.2017.12.134>
- Silveira, L., & Lima, W. Q. (2003). Um breve histórico conceitual da Automação Industrial e Redes para Automação Industrial. *UFRN-PPgEE*, 3.
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Tria, F. Di, Lefons, E., & Tangorra, F. (2018). *A Proposal of Methodology for Designing Big Data Warehouses*. June. <https://doi.org/10.20944/preprints201806.0219.v1>

- Weysow, M., Sahraoui, H., & Syriani, E. (2022). Recommending metamodel concepts during modeling activities with pre-trained language models. *Software and Systems Modeling*, 21(3), 1071–1089. <https://doi.org/10.1007/s10270-022-00975-5>