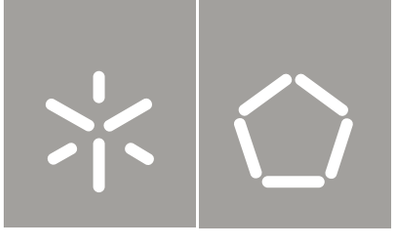
Universidade do Minho
Escola de Engenharia

Sérgio António Ramos da Costa

Sistema de Business Intelligence como suporte à Gestão Estratégica

Sistema de Business Intelligence como suporte à Gestão Estratégica
Sérgio António Ramos da Costa
UMinho | 2012

Outubro de 2012



Universidade do Minho
Escola de Engenharia

Sérgio António Ramos da Costa

Sistema de Business Intelligence como suporte à Gestão Estratégica

Tese de Mestrado
Mestrado em Engenharia e Gestão de Sistemas de Informação

Trabalho efetuado sob a orientação do
Professora Doutora Maribel Yasmina Santos

Outubro de 2012

Agradecimentos

Aproveitando esta oportunidade expresso os meus sinceros agradecimentos:

Primeiramente aos meus pais, pela educação que me proporcionaram, pelo apoio incondicional demonstrado, pelos incentivos constantes na minha formação e por me fazerem procurar em mim o melhor.

À minha orientadora, Professora Doutora Maribel Yasmina Santos, pela sua orientação, apoio e disponibilidade demonstradas ao longo da duração do projeto.

À organização alvo do estudo, aos seus colaboradores no geral e em particular ao Engenheiro André, pelas condições proporcionadas e apoio na realização deste projeto.

À Carolina e ao José João por toda a compreensão e paciência.

Aos meus familiares e amigos por estarem sempre presente e pelo incentivo demonstrado.

E por último, a todos os outros que de alguma forma contribuíram para a concretização do presente projeto de dissertação.

A todos, um sincero, obrigado!

Resumo

Com a conjuntura atual o ambiente organizacional encontra-se em constantes mudanças. Isto implica que as organizações sejam ágeis e tomem frequentemente decisões, por vezes complexas, de forma menos intuitiva e mais fundamentada na informação, sejam estas decisões estratégicas, táticas ou operacionais. Como tal, a informação organizacional e por consequente o conhecimento que dela advém tem-se tornado na função central do negócio. Os gestores das organizações estão cada vez mais cientes de que quem dispõe de informação sistematizada, em quantidade, de boa qualidade, confiável e no momento certo tem todas as condições para alcançar vantagem competitiva. Neste contexto, a utilização de sistemas de *Business Intelligence* pelas organizações tem-se mostrado uma escolha eficaz sempre que se pretenda disponibilizar informação e suportar a tomada de decisão.

No presente projeto de dissertação é dada ênfase ao desenvolvimento e uso de um sistema de *Business Intelligence* para suportar a gestão estratégica de uma organização, que se dedica à comercialização de equipamentos eletrónicos.

No âmbito do trabalho realizado: (i) foi identificada uma arquitetura a ser adotada no sistema a implementar; (ii) foi especificado um modelo de dados para o *Data Warehouse* e definido o processo de ETL; (iii) foram identificados um conjunto de indicadores de desempenho a considerar; (iv) foi implementado um sistema de processamento analítico de dados; (v) foram identificados modelos de *Data Mining* para associação de vendas; e, (vi) foram especificadas e concretizadas as aplicações de *front-end* de acesso à informação.

Os resultados obtidos consistem num sistema funcional que se adapta às necessidades especificadas pela organização para a introdução dos conceitos associados aos sistemas de *Business Intelligence* no suporte à gestão estratégica.

Palavras-Chave: *Business Intelligence*, *Data Warehousing*, *Data Mining*, *Front-End*, Gestão Estratégica, Processo de Tomada de Decisão e *Key Performance Indicators*.

Abstract

The present economic situation truly impacts how businesses do their work. Today's organizations need to be agile, to take better and deliver faster decisions whether strategic, tactical or operational (most of them with great complexity due its quick data grow and lack of consistency). As such, the organizational information and consequently the knowledge that it derives has become the core function of the business. The managers of organizations are increasingly aware of this so they need reliable, centralized and precise data analytics delivered at the right time of the decision making process. In this context, business intelligence systems became an effective choice whenever it provides information or supporting the management activity.

In this project the development process of a Business Intelligence system is shown. Also, its added value to the decision making process is demonstrated making use of a real organization that works in the electronic field. Within the scope of this work: (i) an architecture to support the System was developed; (ii) a data model for the Data Warehouse was specified and the ETL process was designed and implemented; (iii) a set of KPI (key performance indicators) were defined; (iv) an analytical processing system was implemented and data mining models focused on sales activity were identified; and, (v) finally, the front-end applications needed to analyze the data we also made available to the decision makers. The obtained results show a fully functional system adapted to the requirements of the organization that aimed the implementation of a Business Intelligence system to support the decision making process at the strategic level.

Keywords: *Business Intelligence*, *Data Warehousing*, *Data Mining*, *Front-End*, *Strategic Management*, *Decision Making Process*, *Key Performance Indicators.*

Conteúdo

Lista de Figuras

Lista de Tabelas

Notação e Terminologia

Notação

A notação utilizada ao longo do 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.

Lista de Acrónimos

CCTV	*Closed Circuit Tele Vision*
CE	*Community Edition*
CRISP-DM	*CRoss-Industry Standard Process for Data Mining*
CRM	*Customer Relationship Management*
DSA	*Data Staging Area*
EE	*Enterprise Edition*
EIS	*Executive Information System*
ERP	*Enterprise Resource Planning*
ETL	*Extract Transform and Load*
GIS	*Geographical Information Systems*
HOLAP	*Hybrid On-Line Analytical Processing*
KPI	*Key Performance Indicators*
MC	Margem de Contribuição

MCB Margem de Contribuição Bruta

MCR Margem de Contribuição Residual

MDX *MultiDimensional eXpressions*

MOLAP *Multidimensional On-Line Analytical Processing*

OLAP *On-Line Analytical Processing*

OLTP *On-Line Transaction Processing*

RCD *Rapidly Changing Dimension*

RCAAP Repositório Científico de Acesso Aberto de Portugal

ROLAP *Relational On-Line Analytical Processing*

SAD Sistemas de Apoio à Decisão

SCD *Slowly Changing Dimension*

SEMMA *Sample, Explore, Modify, Model, Assess*

SGBD Sistemas de Gestão de Base de Dados

Capítulo 1

Introdução

O capítulo que dá início a esta dissertação começa por descrever e referir o enquadramento e a motivação para a sua realização. De seguida, é descrita a finalidade da dissertação e o contributo que pretende dar, enunciando os seus principais objetivos, bem como as metodologias de investigação utilizadas. Conclui-se o capítulo com a descrição da estrutura do documento e dos restantes capítulos que o compõem.

1.1 Enquadramento e Motivação

No contexto atual, de um mundo de negócios cada vez mais competitivo, as organizações encontram-se sobre constantes pressões. Como consequência, responder às condições de mudança, ser inovadoras no atendimento às necessidades dos seus clientes e no modo de atuarem perante o mercado, torna-se um fator crítico de sucesso. Isto implica que as organizações sejam ágeis e que tomem regularmente rápidas decisões, sejam elas estratégicas, táticas ou operacionais (Vercellis, 2009). No entanto, para tomar boas decisões, necessitam de grande quantidade de dados, de informação e de conhecimento.

Assim, a informação organizacional é tida como um bem de elevada importância. Cada vez mais, as organizações sentem necessidade de informação e de conhecimento, pelo que, se a informação for transformada em conhecimento torna-se num recurso fundamental e na função central de negócio. Quem dispõe de informação sistematizada, em quantidade adequada, de boa qualidade, confiável e no momento certo têm as condições necessárias para alcançar vantagem competitiva. Contrariamente, a falta de informação nas organizações conduz a erros e à perda de oportunidades de negócio.

As exigências do ambiente organizacional e o aumento da concorrência, influenciam de tal forma as organizações que estas têm vindo, cada vez mais, a investirem

em meios que as tornem mais competitivas no mercado. O desenvolvimento de sistemas que permitem efetuar análises para a tomada decisão são, cada vez mais, identificados como essenciais para a melhoria da quantidade e da qualidade da informação disponível para a tomada de decisão nas organizações (Turban, Sharda, & Delen, 2010).

Estrategicamente as organizações sentem a necessidade de olhar para instrumentos que facilitem a aquisição, o processamento e a análise de grandes quantidades de dados (dispersos pela organização), e que sirvam como uma base sólida para descobrir novo conhecimento (Olszak & Ziemba, 2007).

Esta necessidade influenciou o aparecimento de conceitos como o ***Business Intelligence*** e os sistemas que o suportam.

Os sistemas de *Business Intelligence* são atualmente de grande importância para as organizações, apoiando os gestores no processo de tomada de decisão, aumentando a qualidade da mesma, através da disponibilização de informação útil.

Neste âmbito, surge o projeto descrito nesta dissertação e que tem como finalidade o estudo, a conceção e a implementação de um sistema de *Business Intelligence* que suporte a gestão estratégica de uma organização que atua na área da Engenharia Eletrónica Industrial. Ao atuar num mundo global, a organização tem sentido algumas dificuldades no processo de tomada de decisão, e tem identificado a falta de mecanismos que possibilitem o acesso e a exploração de informação. Assim, a organização procura uma solução que lhe permita alcançar melhor qualidade na tomada de decisão estratégica utilizando para isso a informação que se encontra disponível nos seus **Sistemas de Informação**.

Por razões de confidencialidade não é divulgado o nome da organização nem quaisquer resultados aos dados analisados.

1.2 Finalidade e Objetivos

No âmbito deste projeto de dissertação procurou-se identificar uma questão de investigação que fosse a si subjacente e que permitisse, de certa forma, ser uma base para o estudo da solução: "Quais os requisitos que o sistema de *Business Intelligence* deve ter para suportar a tomada de decisão?".

A resposta à questão de investigação é obtida, naturalmente, recorrendo a duas vias: (1) uma de carácter mais teórico, onde foi efetuada uma revisão e análise da literatura, mais relevante, sobre a área em estudo e (2) uma outra de carácter aplicacional, que consiste na conceção e implementação de um sistema de *Business Intelligence* na organização alvo do estudo.

Para dar cumprimento à finalidade deste projeto de dissertação, foram definidos um conjunto de objetivos a alcançar para a sua concretização, nomeadamente:

- Estudar a solução, em termos tecnológicos, que melhor se enquadra à organização;
- Definir uma arquitetura para o sistema de *Business Intelligence*, especificando os seus componentes e tecnologias;
- Definir e implementar um modelo multidimensional que sirva de suporte ao armazenamento dos dados e dos principais indicadores do negócio;
- Definir e implementar um processo ETL[1], que permita a extração e transformação dos dados, das diversas fontes de dados, e os canalize para o modelo multidimensional implementado;
- Definir e implementar um sistema de processamento analítico de dados (cubos OLAP [2]), e identificação de modelos e análise dos dados recorrendo a técnicas avançadas de *Data Mining*;
- Definir e implementar uma aplicação *front-end* que permita o acesso, exploração e visualização da informação.

Com a elaboração deste projeto espera-se a concretização de um sistema de *Business Intelligence* que proporcione aos gestores da organização uma tomada de decisão mais eficiente e eficaz, através da disponibilização de vários mecanismos de recolha, exploração e análise de informação relevante sobre as diversas atividades organizacionais.

Os principais resultados assentam na definição da arquitetura do sistema de *Business Intelligence* e a sua posterior implementação.

1.3 Metodologias Adotadas

Nesta subsecção faz-se referência às metodologias de investigação e de desenvolvimento adotadas no decorrer do projeto de dissertação, bem como a abordagem à revisão de literatura efetuada.

[1]Extração, Transformação e Carregamento, do inglês *Extract Transform and Load.*
[2]Sistemas de Processamento Analítico, do inglês *On-Line Analytical Processing.*

1.3.1 Metodologia de Investigação

Em termos metodológicos as abordagens a seguir são, naturalmente, numa primeira fase a revisão de literatura (detalhada na secção 1.3.3), com o objetivo de enquadrar e clarificar os vários conceitos que se associam ao problema em investigação, e em paralelo analisar o estado da arte nesta área.

Numa segunda fase, para a implementação do sistema de *Business Intelligence* considera-se a metodologia *Design Science/Constructive Research*.

Muito do trabalho científico desenvolvido na área do *Business Intelligence* é construtivo. Esta metodologia é normalmente aplicada em contextos reais, onde o objetivo passa por resolver problemas do mundo real, num determinado domínio a fim de criar conhecimento especifico, o que implica a construção de um artefacto (prático, teórico ou ambos) (Crnkovic, 2010). De uma forma geral, esta metodologia apresenta resultados que podem ter relevância prática e teórica, apresentando-se como parte importante na investigação do âmbito da Engenharia, das Ciências da Computação e dos **Sistemas de Informação** (Vaishnavi & Kuechler, 2011). Normalmente os *outputs* desta metodologia são construções de modelos, teorias, algoritmos, interface, metodologias, entre outros artefactos (Vaishnavi & Kuechler, 2011).

Considerando o objetivo principal desta dissertação, como sendo o desenvolvimento de um sistema de *Business Intelligence* que permita à organização uma tomada de decisão mais eficiente e eficaz, ou seja, um artefacto que procura resolver uma necessidade, e considerando também que o conhecimento será obtido através do desenvolvimento deste artefacto, acredita-se que esta metodologia é a mais apropriada.

Com a solução final deste projeto de dissertação pretende-se a "resolução" de um problema real, assim como uma contribuição para a área em estudo.

1.3.2 Metodologias de Desenvolvimento

Em termos práticos, para o desenvolvimento do projeto, consideram-se duas metodologias: (1) a metodologia para o desenvolvimento de projetos deste tipo baseada na perspetiva de Ralph Kimball (Kimball & Ross, 2002) e (2) a metodologia CRISP-DM de apoio ao processo de *Data Mining* (Chapman, et al., 2000).

Em 1998, Ralph Kimball apresenta uma metodologia para o desenvolvimento de projetos de sistemas de *Data Warehousing*. Esta metodologia, apresentada na Figura 1.1, é designada por metodologia de ciclo de vida dimensional do negócio.

A metodologia apresentada é composta por onze atividades principais e uma

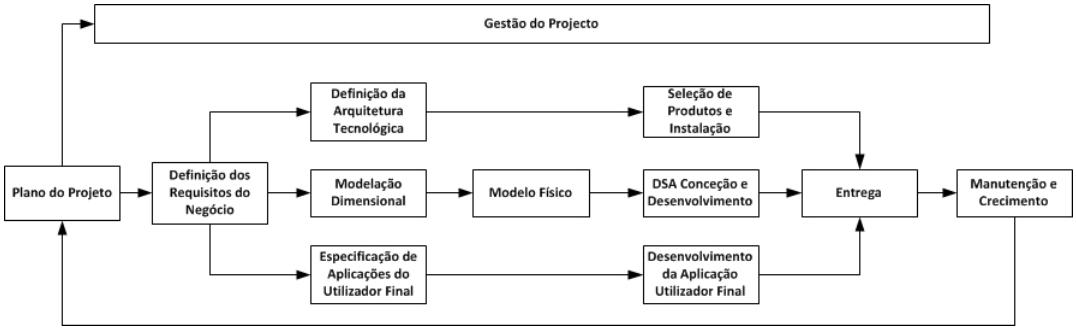


Figura 1.1: Metodologia de ciclo de vida dimensional de negócio

atividade de suporte. Na primeira atividade da metodologia é realizado o Plano do Projeto, onde é identificado e avaliado o nível da organização e a sua preparação para suportar um sistema deste tipo. Desta atividade resulta uma visão ampla de todo o projeto e os critérios de sucesso para sua implementação. Segue-se a atividade de Definição de Requisitos que apresenta como principal objetivo a identificação e desenvolvimento dos requisitos do sistema através do diálogo com os responsáveis e com os utilizadores finais.

No seguimento, e após estarem concluídas as atividades anteriores, parte-se para a realização de mais três atividades que decorrem em paralelo: as atividades de Definição da Arquitetura Tecnológica, Modelação Dimensional e Especificações do Utilizador Final. Na atividade de Definição da Arquitetura Tecnológica definem-se e configuram-se as características de todo o ambiente estrutural e aplicacional. Na atividade de Modelação Dimensional procede-se à modelação dos esquemas do modelo de dados e dos sistemas de armazenamento e carregamento dos dados. Na atividade seguinte, o Modelo Físico e a atividade de DSA[3] Conceção e Desenvolvimento é definido todo o processo ETL. A atividade Especificações Finais do Utilizador tem como resultado a interação com o utilizador garantindo que os requisitos pretendidos estão a ser considerados na modelação do sistema. A atividade de Desenvolvimento das Aplicações do Utilizador é realizada em paralelo e apresenta como resultado a definição das aplicações *front-end*, que vão ser utilizadas pelos gestores para aceder ao *Data Warehouse*.

Posteriormente, procede-se à realização da atividade de Entrega, garantindo que o *Data Warehouse* está construído, completamente carregado de informação, assim como as aplicações de acesso disponíveis. Porém, surge aqui a necessidade de uma tarefa adicional, e não menos importante, que é a formação e suporte para os gestores. Como resultado desta tarefa poderá surgir um manual de suporte a nível aplicacional e de utilização.

[3] *Data Staging Area,* do português Área de Estágio dos Dados.

Por último, a atividade de Manutenção e Crescimento que é a responsável por monitorizar, gerir e definir as prioridades do sistema de *Data Warehousing*, em termos de alterações e propostas de novos desenvolvimentos que possam acontecer, o que faz com que se proceda novamente a todo este ciclo de vida.

A metodologia CRISP-DM de apoio ao processo de *Data Mining,* visto ser uma componente especifica de uma etapa do projeto de dissertação, encontra-se referenciada no decorrer da revisão de literatura efetuada.

1.3.3 Abordagem à Revisão de Literatura

Nos últimos anos vários estudos têm vindo a ser realizados visando a análise do trabalho desenvolvido na área de *Business Intelligence*. Destaca-se um estudo sobre a análise de literatura em *Business Intelligence* que incide na análise e síntese de 167 artigos publicados, entre 1997 e 2006, nas dez principais revistas de Sistemas de Informação, sobre diversos temas estritamente relacionados com a prática de *Business Intelligence* (Jourdan, Raine, & E. Marshall, 2008). Neste estudo foram identificadas um conjunto de revistas, consideradas relevantes, nesta área (e.g., *Journal of the Association for Computing Machinery, Journal of Management Information Systems, Management Science, Information Systems Research, European Journal of Information Systems*).

Num outro estudo levado a cabo por Fitriana, Eriyatno, e Djatna (2011) verificou-se que 46,67%, num universo de estudo composto por 60 artigos e periódicos publicados em revistas como o *International Journal Like Proquest*, *Ebscohost*, *Emerald*, *Science Direct* e IEEE *Conference*, centram-se essencialmente na teoria, no método e em software de sistemas de *Business Intelligence*, mais concretamente na sua definição, metodologia, arquitetura, estudos de caso e *softwares* utilizados. Os restantes 54,33% são artigos que relacionam *Business Intelligence* com um outro tema (e.g., CRM[4], *Data Mining*).

Deste modo, com base nos estudos identificados, desenvolveu-se uma estratégia de pesquisa bibliográfica tendo em conta os objetivos e o tempo disponível para a sua realização.

Primeiramente foram recolhidos, junto do orientador do projeto, os livros e artigos mais relevantes para a área em estudo. Foram selecionadas as fontes de informação iniciais como: artigos, dissertações, teses e livros. Seguidamente, realizaram-se várias pesquisas de artigos científicos recorrendo a base de dados online de pesquisa especializada na procura de publicações científicas, como é o caso do *Google Scho-*

[4] *Customer Relationship Management* do Português Gestão de Relacionamento com o Cliente.

lar[5], do *Scopus*[6] e *Scielo*, da *B-On*[7] e do Repositório Científico de Acesso Aberto de Portugal (RCAAP[8]). As palavras chave utilizadas foram: *Strategic Management*, *Decision Support Systems*, *Business Intelligence*, *Data Warehouse*, *Microsoft SQL Server Business Intelligence*, *Key Performance Indicators* (KPI) e *Data Mining*.

O objetivo principal desta revisão de literatura passa essencialmente por enquadrar os vários conceitos associados aos sistemas de *Business Intelligence* e as suas principais tecnologias, assim como, questões relacionadas com a gestão estratégica e o processo de tomada de decisão.

Com base na revisão percebe-se que a área de *Business Intelligence* é muito ampla e abrangente, e encontra-se em constante evolução.

1.4 Estruturação da Dissertação

Na elaboração do presente documento, que relata todo o trabalho desenvolvido no âmbito do projeto de dissertação, procurou-se adaptar uma escrita simples e fundamentada. Ainda neste âmbito, procurou-se sempre que possível, enunciar as referências mais relevantes na área de *Business Intelligence*, de forma a que, o trabalho apresentasse rigor e relevância nos conceitos apresentados.

É importante salientar que o documento foi escrito tendo como base o novo acordo ortográfico que entrou em vigor em Janeiro de 2009 em Portugal e que a estrutura adotada respeita o guia de dissertação disponibilizado pela coordenação do Mestrado em Engenharia e Gestão de Sistemas de Informação.

Assim, este documento encontra-se estruturado em seis capítulos:

Capítulo 1: O presente capítulo tem como principal objetivo contextualizar e enquadrar o projeto na área em estudo, descrevendo a sua principal finalidade e os objetivos para a sua realização. As metodologias de investigação utilizadas também são aqui apresentadas e justificadas.

Capítulo 2: O capítulo dois evidencia a revisão de literatura efetuada no âmbito do projeto, fazendo referência aos conceitos mais relevantes para o seu desenvolvimento.

Capítulo 3: O capítulo três faz referência às tecnologias utilizadas para o desenvolvimento de projetos de *Business Intelligence* e em particular à tecnologia que foi utilizada neste projeto.

[5] *Http://scholar.google.pt.*
[6] *Http://www.scopus.com.*
[7] *Http://www.b-on.pt.*
[8] *Http://www.rcaap.pt.*

Capítulo 4: No capítulo quatro os requisitos, as necessidades de informação e o modelo de controlo de gestão da organização foram retratados. Neste capítulo efetua-se a análise e caracterização dos dados operacionais e apresenta-se a arquitetura a ser adotada para a implementação do sistema de *Business Intelligence* na organização.

Capítulo 5: O capítulo cinco retrata todo o trabalho prático desenvolvido, detalhando o modelo de dados do *Data warehouse* concebido, faz referência aos processos de ETL implementados para a extração, transformação e carregamento dos dados para o *Data Warehouse*. Posteriormente apresenta-se o sistema de processamento analítico (ou cubos OLAP) implementados e que são a base para o ambiente de análise de negócio. O capítulo termina fazendo referência aos modelos de *Data Mining* identificados para a associação de vendas e às aplicações de *front-end* utilizadas para aceder à informação.

Capítulo 6: Por último, o capítulo seis, apresenta as conclusões e resultados do trabalho realizado, evidenciando as dificuldades sentidas no seu desenvolvimento e fazendo uma referência ao trabalho futuro proposto.

Para além dos seis capítulos enunciados, este documento faz referência a um conjunto de anexos que complementam o trabalho desenvolvido e à bibliografia utilizada para a sua concretização.

Capítulo 2

Enquadramento Conceptual

Neste capítulo evidencia-se a revisão de literatura efetuada. Fazem parte desta os principais conceitos e metodologias que fundamentam o trabalho desenvolvido. O capítulo é iniciado pela caracterização dos conceitos relacionados com a gestão estratégica e o processo de tomada de decisão. Posteriormente é apresentado o estado da arte associado ao *Business Intelligence*, introduzindo o tema, seguido das tecnologias e técnicas que o suportam. Finalmente, são analisados, através de alguns exemplos, aplicações de sistemas de *Business Intelligence* nas organizações.

2.1 Gestão Estratégica

2.1.1 Da Gestão à Gestão Estratégica

Num mundo cada vez mais competitivo a gestão têm-se revelado uma atividade fundamental para as organizações, para que possam competir com melhor qualidade no mercado. A gestão, como atividade, afeta os resultados de uma organização e pode ser entendida como um processo de coordenação e integração de atividades organizacionais, através do planeamento, organização, direção e controlo, que visa assegurar a consecução dos objetivos definidos, através dos recursos disponíveis, de uma forma eficiente e eficaz (Santos, 2008). Contudo, para prosseguir com uma boa gestão da organização, é essencial recorrer a uma estratégia, planeando o que se pretende fazer e os objetivos a atingir, mantendo sempre no pensamento a organização e o seu desenvolvimento futuro. A estratégia assenta na formulação de um plano definido e orientado, levando uma organização a atuar de forma a superar os seus concorrentes de mercado. A formulação de uma estratégia resulta, normalmente, numa série de indicadores chave de desempenho (KPI), com o objetivo de mensurar o seu grau de progresso e sucesso, de uma forma rápida e objetiva (Laursen & Thorlund, 2010).

Os KPI podem ser financeiros e não financeiros e são normalmente utilizados em várias áreas de uma organização. O seu conteúdo depende da unidade de negócio que se pretende controlar e mensurar. Uma relação entre os KPI e os objetivos estratégicos das organizações, mostrando-se atingíveis e consistentes, é importante de forma a permitir gerir possíveis desvios no negócio e no desempenho operacional da organização (Osadník & Landryová, 2011). No seu estudo, Osadník e Landryová (2011), fizeram um levantamento dos KPI mais frequentes e relevantes para um conjunto de pequenas e médias organizações, entre os quais se destacam: (1) Vendas por mês – que representa a vendas globais num período de tempo (mês); (2) Lucro por produto – montante ganho com a venda do produto; (3) Vendas geradas – quantidade de vendas num determinado período de tempo; (4) Fluxo de caixa – controlo de como os fundos da organização se movimentam internamente; e, (5) Entregas dentro do prazo – relacionam-se com as entregas aos clientes e se estas foram efetuadas atempadamente. As áreas mais predominantes são a gestão de clientes, o prazo de execução de um processo e os custos de produção/operacionalização.

Normalmente associada a uma estratégia está um conjunto de decisões que, de certa forma, proporcionam aos clientes e fornecedores mais valor do que aquele que é prestado pelos seus concorrentes. Assim, as pequenas e médias organizações necessitam cada vez mais de definir estratégias de negócio favoráveis, tendo em vista o seu crescimento.

Neste contexto, surge a gestão estratégica que consiste no planeamento e na tomada de decisão estratégica e, na execução de ações que permitam atingir os objetivos que uma organização se propôs alcançar, de forma a criar vantagem competitiva. Alguns fatores de análise podem ser a si associados, como é o caso das oportunidades e ameaças, dos pontos fortes e fracos da organização, do próprio mercado e dos concorrentes. A gestão estratégica é, em síntese, um processo onde as principais tarefas são a formulação dos objetivos, a análise do ambiente (organizacional) externo, a formulação e a implementação da estratégia e o respetivo controlo (Gonçalves, 1986, citado por Santos, 2008).

É importante referir que as mudanças exigem dos gestores das organizações capacidade de formulação e implementação de estratégias que permitam superar os desafios do mercado, para que os objetivos previamente definidos sejam atingidos, quer eles sejam de curto ou médio e longo prazo. Os gestores devem assim tomar decisões estratégicas para que as suas organizações possam competir no mercado em que se encontram inseridas.

Por sua vez, as decisões devem ser transformadas em ações, a fim de se tornarem eficazes.

2.1.2 O Processo de Tomada de Decisão

Assim como na vida pessoal, as boas decisões são a chave para o sucesso de uma organização e, estas, de certa forma, devem fazer boas escolhas.

Julga-se importante compreender como um processo de tomada de decisão é estruturado, percebendo as necessidades que levam os gestores das organizações a tomar decisões, e perceber como as decisões são tomadas, com base nessas mesmas necessidades e no contexto onde estas se envolvem.

As decisões tomadas numa organização (pública ou privada) são escolhas de várias alternativas. Normalmente estas decisões dizem respeito a um plano estratégico, o que implica escolhas de investimento, iniciativas de marketing, previsões de vendas e conceção de planos de produção, permitindo assim gerir eficientemente e eficazmente os recursos e as tecnologias disponíveis empregues (Vercellis, 2009).

O processo de tomada de decisão, segundo Vercellis (2009), é o processo pelo qual um indivíduo tenta preencher uma falha ou oportunidade entre as condições operacionais atuais de um determinado sistema (como ele é) e as condições mais vantajosas a serem alcançadas no futuro (o que vai ser). Certamente que para atingir o estado esperado é necessário definir um conjunto de opções válidas para o alcançar, e depois selecionar uma decisão baseada na comparação efetuada entre as vantagens e desvantagens de cada opção. A maioria das decisões são tomadas em resposta a um problema.

As decisões podem ser classificadas de acordo com a sua natureza e o seu âmbito (Vercellis, 2009). De acordo com a sua natureza as decisões podem ser: (1) estruturadas, caso sejam baseadas numa estrutura bem definida e recorrente no processo de tomada de decisão. Normalmente, as decisões estruturadas são mais adequadas para a automação operacional, e podem ser definidas, por exemplo, por um algoritmo mais ou menos explícito ao gestor que toma a decisão; (2) não-estruturadas, quando existe pelo menos um elemento do sistema (*inputs*, *outputs* e meios de transformação) que não pode ser detalhado e pré-definido. Este tipo é por vezes verificado quando ocorre um processo de tomada de decisão pela primeira vez, ou que acontece muito raramente; e (3) semiestruturada, quando alguma das fases de decisão são estruturadas e outras não. As decisões enfrentadas pelos gestores das organizações são maioritariamente semiestruturadas.

A natureza de um processo de decisão depende das características de uma organização, dentro do qual o sistema está inserido, das atitudes dos gestores que tomam as decisões, da disponibilidade de metodologias adequadas para a resolução de problemas e da disponibilidade de tecnologias eficazes que suportem e apoiem a tomada de decisão (Vercellis, 2009).

De acordo com o seu âmbito, as decisões podem ser (Vercellis, 2009):

- Decisões estratégicas: As decisões são estratégicas quando afetam a organização como um todo, nomeadamente o seu trabalho ou, pelo menos, uma parte substancial da mesma por um período de tempo longo. Estas decisões influenciam os objetivos e políticas gerais de uma organização, tendo amplas implicações para a sua direção (Panian, 2007). As decisões estratégicas são tomadas a um nível superior da organização, geralmente pela gestão de topo;

- Decisões táticas: As decisões são táticas quando afetam apenas parte da organização e limitam-se a um único departamento. Estas decisões encontram-se dentro de um contexto que é determinado por decisões estratégicas. As decisões táticas são tomadas a um nível médio da organização, geralmente por gestores de departamento, estando focadas na gestão de processos (Panian, 2007);

- Decisões operacionais: As decisões são operacionais quando se referem a atividades específicas desenvolvidas no interior de uma organização e têm um impacto reduzido no futuro da mesma. Assim como as anteriores, as decisões operacionais encontram-se dentro de um contexto que é determinado pelas decisões estratégicas e táticas, e são geralmente tomadas a um nível inferior da organização.

O mesmo autor afirma que embora a natureza e o âmbito da decisão não estejam inteiramente correlacionados, a maioria das decisões tomadas no mundo real enquadram-se dentro da elipse apresentada na Figura 2.1, adaptada de Vercellis (2009).

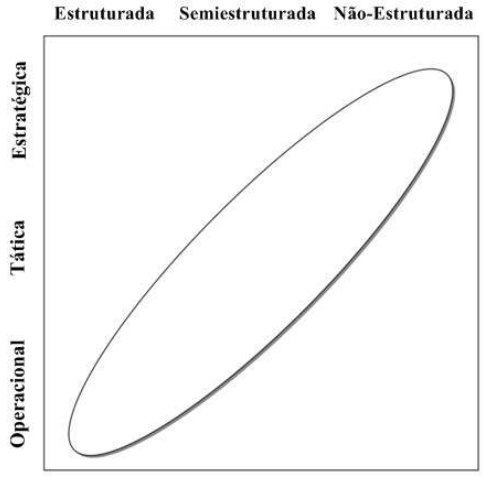


Figura 2.1: Taxonomia de tomada de decisões

Esta observação é importante quando se torna necessário definir as características de um sistema para facilitar o processo de tomada de decisão de natureza e âmbito específicos (Vercellis, 2009). No seu trabalho, Panian (2007) afirma que para proporcionar valor a este tipo de sistemas devem ser compreendidos os tipos (âmbitos) das decisões nas organizações.

As decisões estratégicas são maioritariamente não-estruturadas, as decisões operacionais estruturadas e as decisões mais táticas são semiestruturadas. É assim possível afirmar que as decisões tomadas a um nível estratégico têm grande impacto sobre o resto da organização.

Os gestores a fim de tomarem as melhores decisões necessitam de informação sistematizada e com qualidade, sobre as diversas atividades organizacionais, sejam elas estratégicas, táticas ou operacionais, não se focando especificamente numa parte da organização mas sim nesta como um todo. Todavia, para tomar decisões corretas e no momento oportuno, são necessários sistemas que superem a preocupação da descentralização do processo de decisão organizacional e o tornem num processo mais eficiente. Para tal, este trabalho utiliza os sistemas de *Business Intelligence* como plataforma para suportar a tomada de decisão nas organizações.

2.2 *Business Intelligence*

Com a evolução das **Tecnologias e Sistemas de Informação** e das mudanças organizacionais surge, na década de 80, o conceito de *Executive Information System*[1] (EIS) (Vercellis, 2009). Os EIS vieram melhorar a qualidade da gestão estratégica nas organizações através de novas tecnologias e diversas técnicas para extração, transformação, processamento e apresentação de dados, suportando o processo de tomada de decisão. Contudo, mais recentemente, na década de 90, o paradigma mudou e através da *Gartner Research Group*[2] emergiu o conceito de *Business Intelligence*.

O *Business Intelligence* pode ser definido como sendo um *umbrella term* que combina arquiteturas, ferramentas, bases de dados, ferramentas analíticas, aplicações e metodologias (Turban, Sharda, & Delen, 2010). Por sua vez, Habul e Pilav-Velic (2010) definem o *Business Intelligence* como um processo de recolha, tratamento, análise e utilização de informação estratégica para as organizações.

Com o crescimento acentuado, nas duas últimas décadas, do número de produtos e serviços de *Business Intelligence* oferecidos, bem como na adoção destes por parte

[1] Do Português Sistemas de Informação para Executivos.
[2] Consultora de pesquisas de mercado na área das Tecnologias da Informação.

das organizações (Chaudhuri, Dayal, & Narasayya, 2011), a área de *Business Intelligence* tem vindo cada vez mais a ser identificada como essencial para a melhoria da quantidade e da qualidade da informação disponível para a tomada de decisão nas organizações (Turban, Sharda, & Delen, 2010). Com maior amplitude para a gestão estratégica, estes sistemas surgem como um importante apoio para responder às necessidades dos gestores na gestão de informação, quer seja ao nível estratégico, tático ou operacional de uma organização. Através destes, os gestores têm mais facilidade em tomar decisões, de uma forma menos intuitiva e mais fundamentada na informação.

Assim, e segundo Santos e Ramos (2009), poder-se-á dizer que os sistemas de *Business Intelligenc*e combinam a recolha de dados operacionais, permitem o seu posterior armazenamento em repositórios adequados, que por sua vez vão permitir a gestão de conhecimento através de diferentes ferramentas de análise, exploração e apresentação da informação, dita essencial, para a tomada de decisão.

De um modo geral, um sistema de *Business Intelligence* serve para (Santos & Ramos, 2009):

- Analisar dados passados ou atuais;
- Prever fenómenos e tendências;
- Analisar e comparar dados do passado com novos dados de forma a perceber o que mudou;
- Permitir o acesso *ad-hoc* a dados para responder a questões que não se encontram predefinidas;
- Analisar a organização de modo a obter um conhecimento mais profundo das suas atividades.

A implementação de um sistema deste tipo pode contribuir de forma eficiente e precisa no desenvolvimento de uma organização, criando o conhecimento necessário para projetos futuros, de modo a minimizar falhas do passado e maximizar o seu desempenho.

Após ser explicado o que se entende por *Business Intelligence* e sistemas de *Business Intelligence*, julga-se relevante adotar uma arquitetura de suporte a estes sistemas. Através da revisão de literatura efetuada foi possível identificar várias arquiteturas de sistemas de *Business Intelligence* diferentes entre si, de acordo com cada autor e o contexto onde estas estavam a ser aplicadas (Han & Kamber, 2011; Chaudhuri, Dayal, & Narasayya, 2011; Al-Debei, 2011; Dayal & Chaudhuri, 1997).

Na Figura 2.2 é apresentada uma arquitetura geral de um sistema de *Business Intelligence* incorporando os conceitos apresentados pelos diversos autores, em particular pela arquitetura apresentada por Chaudhuri, Dayal, e Narasayya (2011), mostrando de forma simples cada elemento que integra este tipo de sistemas.

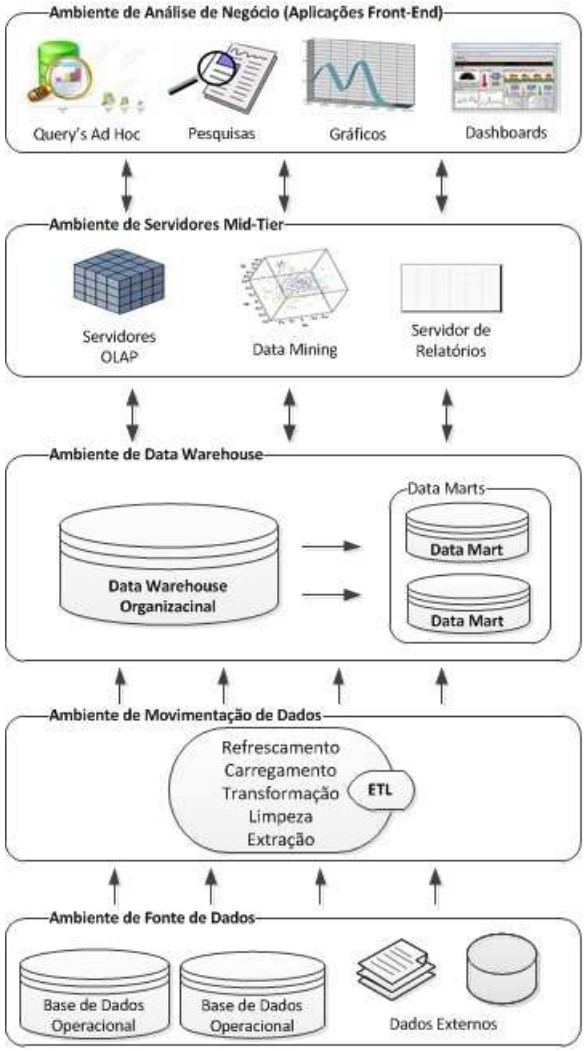


Figura 2.2: Arquitetura de apoio ao processo de *Business Intelligence*

A definição de uma arquitetura em camadas facilita a identificação das fases de implementação e tecnologias necessárias para concretizar sistemas de *Business Intelligence*. Esta arquitetura é constituída por cinco camadas, correspondendo cada uma destas a um determinado ambiente:

- No ambiente de fontes de dados encontram-se todas as origens dos dados que vão suportar o sistema. As fontes de dados poderão ser internas e/ou externas à organização, como *Enterprise Resource Planning* (ERP), folhas de cálculo, ficheiros de texto, bases de dados isoladas, entre outras;

- Na camada imediatamente a seguir encontra-se o ambiente de movimentação de dados. Aqui realiza-se todo o processo de ETL, utilizando ferramentas apropriadas para o tratamento dos dados, que provêm das diversas fontes, de forma a integrar os mesmos, transformando-os e carregando-os para os repositórios de dados apropriados;

- Na camada seguinte encontra-se o ambiente de *Data Warehouse*. Este integra o *Data Warehouse* e os diversos *Data Marts* da organização, que são repositórios de informação organizacional. Estes são normalmente carregados e refrescados no processo de ETL a partir das várias fontes de dados existentes na organização;

- A penúltima camada representa o ambiente de servidores *mid-tier*. Aqui será possível trabalhar os dados, acedendo ao *Data Warehouse* ou *Data Marts* organizacionais, com recurso a várias técnicas, como OLAP e *Data Mining*, de forma a poder-se gerar informação relevante para a tomada de decisão e disponibilizar a mesma aos gestores;

- A última camada representa o ambiente de análise de negócio. Existem várias aplicações de *front-end* que permitem o acesso e a manipulação da informação através das quais os gestores executam tarefas de *Business Intelligence*, como é o caso dos portais web de pesquisa e aplicações de gestão de desempenho, para acompanhar o desempenho do negócio utilizando ferramentas como *dashboards* e consultas *ad-hoc*.

Desta forma é possível dizer que existe um vasto conjunto de tecnologias associadas ao sistemas de *Business Intelligence* (Vercellis, 2009) e a cada ambiente em particular.

Tendo por base o objetivo do projeto de dissertação, são explorados os sistemas de *Data Warehousing*, sistemas OLAP, *Data Mining* e as aplicações de *front-end*.

2.2.1 Sistemas de *Data Warehousing*

Nesta secção são abordados os conceitos, as metodologias e as técnicas que integram um sistema de *Data Warehousing*.

2.2.1.1 *Data Warehouse*

Inmon, que introduziu o termo *Data Warehouse* pela primeira vez em 1991, define-o como sendo um conjunto de dados integrados, orientados e organizados a um

determinado assunto, não voláteis e estruturados temporalmente de forma a suportar o processo de tomada de decisão (Inmon, 2005).

Como é possível reconhecer na literatura (Sá, 2009; Al-Debei, 2011; Santos & Ramos, 2009), e segundo a perspetiva de Inmon, um *Data Warehouse* é caracterizado por ser (Inmon, 2005):

- **Organizado e orientado por assunto/tema**: Os dados são organizados por assuntos/temas chave de uma organização (e.g., processo de vendas, processo de compras). Estes são normalmente apresentados de forma compartimentada, de acordo com as necessidades específicas dos utilizadores finais do sistema. Visam a análise de dados para o suporte à tomada de decisão, oferecendo uma visão simples de um determinado assunto, excluindo os dados que não são relevantes no processo de tomada de decisão;

- **Integrado**: O *Data Warehouse* terá que ser uma fonte de dados única e ao mesmo tempo abrangente sobre e para o negócio. Os vários dados que se encontram nas diversas fontes de dados são selecionados, integrados e posteriormente armazenados num *Data Warehouse*, garantindo e fornecendo uma visão única e coerente do negócio, de forma a responder às questões que são colocadas pelos utilizadores;

- **Catalogado temporalmente**: O *Data Warehouse* apresenta o histórico dos dados e a informação atual sobre o negócio. O seu objetivo principal é fornecer informação válida sobre a perspetiva histórica. Para isso, devem conter uma dimensão temporal que permita o registo da informação histórica;

- **Não volátil**: O *Data Warehouse* apresenta dados estáveis, isto é, podem ser adicionados novos dados mas não devem ser removidos os dados lá armazenados. Consequentemente, no ambiente de *Data Warehouse* existem três tipos de operações: o carregamento inicial dos dados, o posterior refrescamento dos dados e o acesso aos dados para processamento de consultas.

Segundo Han e Kamber (2011), um *Data Warehouse* é considerado um repositório de dados consistentes, através do qual se constitui um modelo de dados de suporte à decisão, armazenando informação relevante para a tomada de decisão estratégica da organização. São uma cópia de registos informacionais de uma transação, estruturados para que sobre eles se possam efetuar interrogações e análises (Kimball, Reeves, & Thornthwaite, 1998).

Os *Data Warehouses* encontram-se separados das bases de dados operacionais da organização. Um dos motivos pelo qual isso acontece centra-se na finalidade de cada um dos sistemas (Dayal & Chaudhuri, 1997).

Os sistemas de bases de dados operacionais, conhecidos por *On-Line Transaction Processing* (OLTP), são sistemas concebidos para registar todas as operações (e.g., encomendas, faturas, vendas) do dia-a-dia de uma organização, através das operações de inserção, modificação e eliminação de informação na base de dados num determinado período de tempo. Estas tarefas, segundo Dayal e Chaudhuri (1997) são estruturadas e repetitivas, e consistem em transações atómicas e isoladas.

Contrariamente, os sistemas de *Data Warehouse* são direcionados para suportar a decisão das organizações, e portanto são considerados sistemas analíticos, conhecidos por OLAP (Han & Kamber, 2011), uma vez que apresentam informação consolidada, histórica e resumida, sendo este um fator importante (Dayal & Chaudhuri, 1997).

Para perceber a diferença entre estes dois tipos de sistemas, a Tabela 2.1, adaptada de (Polsku, 2009) e (Vercellis, 2009), sistematiza as suas principais características.

Uma das características, do *Data Warehouse*, está relacionada com o facto de este integrar informação referente a um determinado assunto, ou vários, da organização, caracterizando-a como um todo e não parte dela. Neste sentido, se um *Data Warehouse* engloba poucos assuntos e responde a um número reduzido de unidades de uma organização, por exemplo departamental (e.g., vendas, marketing, finanças, recursos humanos), estamos perante um *Data Mart*.

Os *Data Marts* são repositórios de dados multidimensionais, mais pequenos do que os *Data Warehouses,* que reúnem um conjunto de tabelas dimensionais de suporte a um determinando processo de negócio (Kimball & Ross, 2002).

2.2.1.2 Abordagens de Implementação

Uma das maiores preocupações em sistemas de *Data Warehousing* é como planear a sua construção. Duas abordagens são normalmente aceites na implementação destes sistemas, uma do tipo *top-down* e outra do tipo *bottom-up*.

A primeira abordagem, sugerida por Inmon (2005), subdivide-se em duas etapas. A primeira etapa consiste na definição do esquema global do *Data Warehouse* e a segunda baseia-se na implementação de *Data Marts* de acordo com as necessidades e características das várias unidades de negócio da organização. Esta implica mais tempo de desenvolvimento e maior risco de não ser concluída dentro do prazo estabelecido (Vercellis, 2009).

A segunda abordagem, sugerida por Kimball e Ross (2002), baseia-se na experimentação e desenvolvimento de protótipos (Santos & Ramos, 2009). É uma abordagem mais simples e incremental. O objetivo passa por construir esquemas, individuais, de cada *Data Mart*, tendo em consideração as necessidades de cada

Tabela 2.1: Sistema OLTP vs. sistema OLAP

Características	Sistemas OLTP	Sistemas OLAP
Fonte de Dados	- Dados operacionais; - OLTP são as fontes originais dos dados;	- Consolidação de dados; - Dados OLAP surgem de vários sistemas OLTP;
Propósito dos Dados	- Executar e controlar tarefas fundamentais do negócio;	- Ajudar no planeamento, resolução de problemas e suporte à decisão;
Tipo de Dados	- Revelam o momento dos acontecimentos nos processos de negócio; - São atómicos, normalizados, atualizados e isolados;	- Evidenciam várias perspetivas (multidimensionais) das atividades organizacionais; - São históricos, sumarizados, multidimensionais e integrados;
Inserções e Atualizações	- Curtas e rápidas; - Atualizações efetuadas pelos utilizadores;	- Periódicas e de longa duração;
Acessos/ Consultas	- Leitura/escrita; - Consultas/transações padronizadas e simples (envolvem poucas tabelas e retornam poucos registos);	- Leitura na maior parte das vezes; Consultas complexas que envolvem agregações (envolvem várias tabelas e retornam muitos registos);
Velocidade de Processamento	- Normalmente muito rápido para inserir, apagar e alterar, menos rápido para consultar;	- Depende da quantidade de dados envolvidos; - Está otimizado para processamento de questões;
Requisitos de Espaço	- Pode ser relativamente pequena caso os dados históricos não sejam guardados;	- Requer mais espaço devido à existência de agregação dos dados e história. - Requer mais índices do que os Sistemas OLTP;
Conceção da Base de Dados	- Altamente normalizados e com muitas tabelas; - Orientado às aplicações; - Otimizados para atualizações;	- Normalmente desnormalizados e com menos tabelas; - Orientados aos assuntos; - Otimizados para o processamento de questões;
Utilização	- Repetitiva (diariamente).	- *Ad-hoc.*

unidade de negócio. Estes esquemas devem ser modelados tendo em vista a futura unificação dos mesmos, de modo a obter-se um *Data Warehouse* organizacional (Sá, 2009). Geralmente esta abordagem é mais rápida, fornece resultados mais tangíveis, mas carece claramente de uma visão conjunta de todo o sistema a ser desenvolvido (Vercellis, 2009).

O tipo de abordagem depende muito das necessidades organizacionais. A abordagem *top-down* traduz-se num modelo complexo já que engloba toda a informação da organização num *Data Warehouse*. A abordagem *bottom-up* traduz-se em *Data Marts* construídos para no futuro serem integrados num *Data Warehouse*. Ambas são implementadas em diferentes arquiteturas como poder-se-á verificar na próxima subsecção.

2.2.1.3 Arquiteturas

As organizações devem identificar qual a arquitetura, de *Data Warehouse*, que melhor satisfaz as suas necessidades (Santos & Ramos, 2009). A seleção da arquitetura é um dos principais fatores que condiciona e influencia o sucesso do armazenamento dos dados (Ariyachandraa & Watson, 2010). Porém, existem vários fatores que influenciam e afetam a escolha de um tipo de arquitetura e a respetiva implementação de um *Data Warehouse*, destacando-se a interdependência da informação entre as unidades da organização, a necessidade da informação pela gestão da organização e a urgência da necessidade de um *Data Warehouse*, a visão estratégica do *Data Warehouse* antes da sua implementação e questões técnicas (Ariyachandraa & Watson, 2006).

A organização pode assim optar pela implementação de um *Data Warehouse* organizacional, pela implementação de *Data Marts* independentes ou pela implementação de *Data Marts* dependentes (Santos & Ramos, 2009). Estas são as arquiteturas mais comuns e as tradicionalmente apresentadas na literatura (Ariyachandra & Watson, 2006; Ariyachandraa & Watson, 2010; Ariyachandra & Watson, 2011; Santos & Ramos, 2009).

A arquitetura *Data Warehouse* organizacional, apresentada na Figura 2.3, integra todo o conjunto de dados associado às várias unidades negócio, sendo transversal a toda a organização.

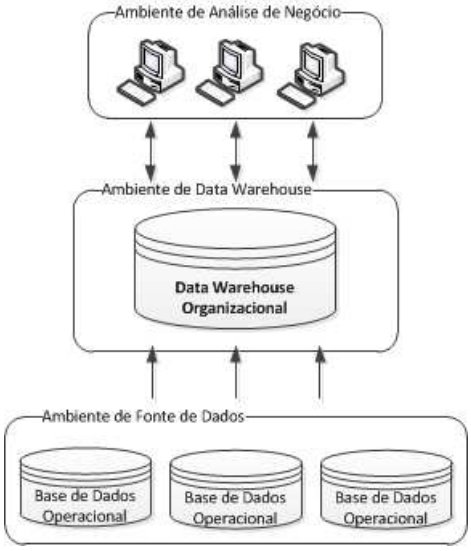


Figura 2.3: Arquitetura *Data Warehouse* organizacional

Para abranger a organização na sua totalidade, um *Data Warehouse* organizacional deve ser o espelho do negócio, implicando desta forma uma modelação detalhada do mesmo.

A arquitetura *Data Marts* independentes, apresentada na Figura 2.4, não traduz o negócio da organização como um todo, mas sim uma unidade de negócio específica (e.g., processos de negócio ou departamentos organizacionais). Neste tipo de

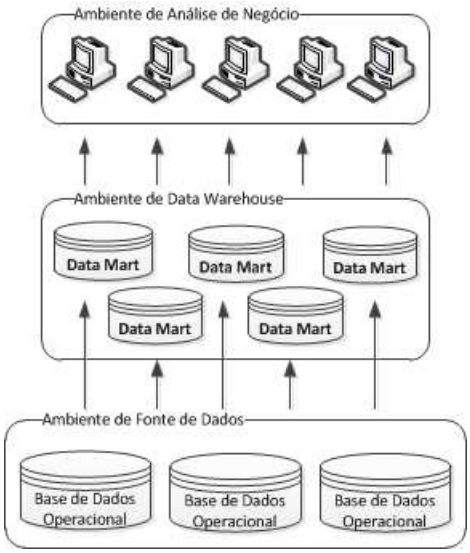


Figura 2.4: Arquitetura *Data Marts* independentes

arquitetura apenas se encontram integrados dados que vão ser relevantes para um determinado grupo de utilizadores. Embora os *Data Marts* satisfaçam necessidades para as quais foram criados, eles não oferecem uma visão integrada dos dados, pois podem apresentar dados inconsistentes e usar diferentes dimensões e medidas, o que dificulta a análise dos dados (Ariyachandra & Watson, 2006). Se existem vários *Data Marts* e se a sua modelação não for realizada tendo em vista a unificação dos mesmos no futuro poder-se-á, de facto, ter problemas ao nível da consistência dos dados. Este tipo de arquitetura é normalmente adotado pelas organizações devido aos custos elevados que as implementações de arquiteturas de *Data Warehouse* organizacionais apresentam e ao facto dos riscos serem consideravelmente reduzidos.

A arquitetura *Data Marts* dependentes, apresentada na Figura 2.5, é o inverso da arquitetura apresentada anteriormente. Nesta, os dados são carregados diretamente do *Data Warehouse* organizacional para os respetivos *Data Marts*. A análise inicial dos requisitos dos dados a um nível organizacional é a base para este tipo de arquitetura (Ariyachandra & Watson, 2006). Utilizando uma visão dos dados transversais de toda a organização, os *Data Marts* são desenvolvidos de uma forma iterativa. Esta arquitetura implica a implementação inicial de um *Data Warehouse* organizacional. Os *Data Marts* dependentes podem ser desenvolvidos por área, por departamento, ou para fins específicos como é o caso de *Data Mining* (Ariyachandra & Watson, 2006).

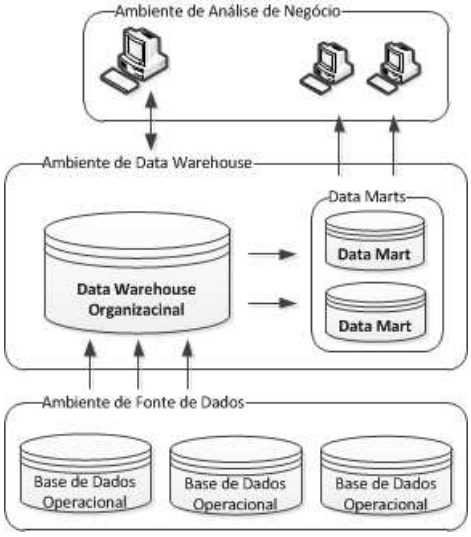


Figura 2.5: Arquitetura *Data Marts* dependentes

Contudo, e sendo estas as arquiteturas mais comuns, identificaram-se outras arquiteturas de *Data Warehouse*. Arriyachandra e Watson (2006) apresentam no seu estudo mais duas outras arquiteturas: *Centralized Data Warehouse* e *Federated Architecture*. Rainardi (2008) apresenta a *Federated Data Warehouse*, que apesar de a referir com uma designação diferente, revela as mesmas características.

A *Centralized Data Warehouse* é muito semelhante com a arquitetura de *Data Marts* independentes. Ao invés da implementação física, como acontece na anterior, esta é tipicamente uma arquitetura lógica, pois o repositório de dados contém dados atómicos e resumidos, e apresentam sobre si uma vista dimensional, mas no entanto as consultas e todas as aplicações de acesso são realizadas aos dados relacionais e a vistas dimensionais (Ariyachandra & Watson, 2006).

A *Federated Architecture* consiste em vários repositórios de dados com uma camada de recuperação sobre eles (Rainardi, 2008). Os dados são logicamente ou fisicamente integrados utilizando chaves partilhadas, consultas distribuídas, metadados globais ou integração de informação empresarial (Ariyachandra & Watson, 2006; Ariyachandraa & Watson, 2010; Rainardi, 2008). O conceito é semelhante com o das arquiteturas anteriormente apresentadas, contudo o *Federated Architecture* recupera os dados dos repositórios existentes (*Data Marts*, *Data Warehouse*, sistemas OLTP) e, através de uma integração lógica ou física, carrega-os para um repositório de dados multidimensional, neste caso unificado. Esta arquitetura apresenta como principal vantagem a unificação dos diversos repositórios de dados existentes e, desta forma, o tempo de implementação é mais curto (Rainardi, 2008). Ariyachandraa e Watson (2010) afirmam que este tipo de arquitetura é realista para as organizações que apresentam Sistemas de Apoio à Decisão (SAD) complexos e que

não serão reconstruídos. No entanto, na prática é difícil construir um repositório de dados de boa qualidade a partir dos padrões encontrados nos diversos repositórios já existentes (Rainardi, 2008), tornando-se assim numa desvantagem para a adoção desta arquitetura.

2.2.1.4 Modelação Multidimensional

Anteriormente foi evidenciado que a estrutura do conteúdo de um sistema OLTP e de um *Data Warehouse* apresenta-se de forma diferente. O primeiro recorre a técnicas de relacionamento entre entidades e técnicas de normalização (Dayal & Chaudhuri, 1997) (se for relacional deverá estar, normalmente, devidamente normalizado), já o segundo recorre a dois tipos de modelos de dados, o relacional e o multidimensional (Wu e Buchmann, 1997 citado por Sá, 2009, (Vercellis, 2009)).

Os modelos de dados que recorrem a técnicas de relacionamento entre entidades são inadequados para modelos de *Data Warehouse*, uma vez que o objetivo principal passa pela eficiência das consultas e do carregamento dos dados. Partindo deste principio é necessário recorrer a uma modelação multidimensional para conceber um modelo de dados para este tipo de sistemas. Segundo Santos e Ramos (2009), este tipo de modelação traduz um modelo de dados fácil de compreender e utilizar e, ao mesmo tempo, proporciona uma melhor otimização do sistema, no que ao processamento de questões diz respeito (Vercellis, 2009), contrariamente ao modelo de dados relacional que assenta na otimização do processamento de atualizações.

É natural derivar o modelo de dados multidimensional de um *Data Warehouse* a partir dos esquemas já existentes que recorrem a técnicas de relacionamento entre entidades (Golfarelli, Maio, & Rizzi, 1998). A modelação multidimensional apresenta como componentes básicos os factos, as dimensões e as hierarquias (Golfarelli, Maio, & Rizzi, 1998), e pode ser implementada sobre um dos seguinte três esquemas (Dayal & Chaudhuri, 1997):

- Esquema em Estrela (*Star Schema*);
- Esquema em Constelação (*Constellation Schema*);
- Esquema em Floco de Neve (*Snowflake Schema*).

O esquema em Estrela é habitualmente o modelo de dados multidimensional mais utilizado para modelar um *Data Warehouse* ou um *Data Mart*. Este esquema, representado na Figura 2.6, é constituído por uma única tabela de factos (vendas) e por várias tabelas de dimensão (produto, data, vendedor, cliente e cidade).

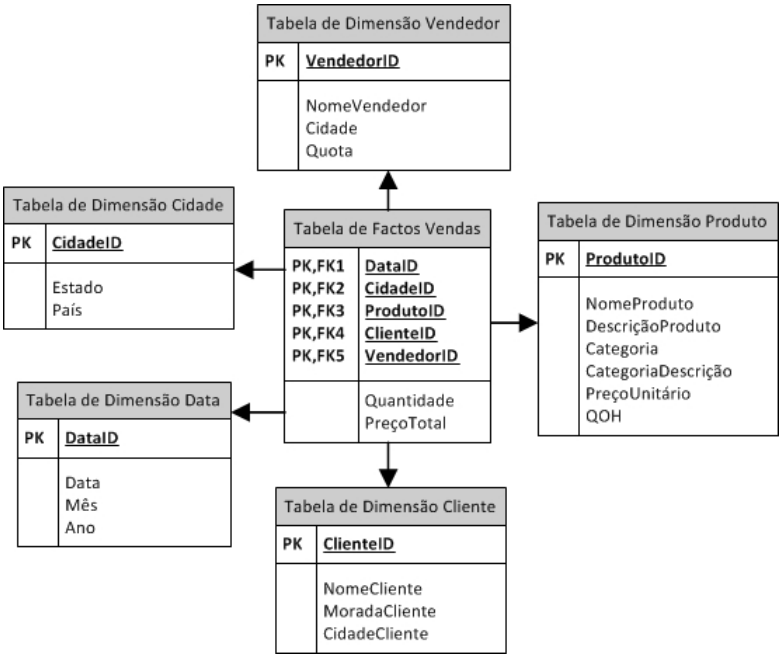


Figura 2.6: Exemplo de um esquema em Estrela

A **tabela de factos** é um *driver* do negócio, isto é, corresponde a um determinado assunto que se pretende analisar, normalmente uma unidade de negócio específica (e.g., vendas, compras, movimentos de stock). Esta tabela armazena enormes quantidades de dados, representando quase a totalidade do espaço ocupado pelo *Data Warehouse*. Na tabela de factos cada registo integra um conjunto de colunas (chaves estrangeiras) que a relacionam com as tabelas de dimensão e colunas que descrevem valores numéricos (e.g., quantidade vendida, preço de venda), que retratam as medidas e/ou indicadores de análise (que podem ou não ser agregados), normalmente conhecidos por factos. Os factos são classificados em três tipos: os aditivos, os semi-aditivos e os não-aditivos (Golfarelli, Maio, & Rizzi, 1998). Os factos aditivos são aqueles que podem ser agregados pelas diversas dimensões relacionadas com a tabela de factos. Os factos semi-aditivos podem ser agregados apenas por uma ou algumas dimensões. Os factos não-aditivos são aqueles que não podem ser agregados por nenhuma das dimensões. Os factos semi-aditivos e não aditivos podem ainda ser agregados usando, por exemplo, operadores como a média, máximo ou mínimo (Golfarelli, Maio, & Rizzi, 1998).

As **tabelas de dimensão** correspondem às variáveis de análise pelas quais a tabela de factos pode ser analisada (e.g., dimensão tempo, a dimensão produto, a dimensão clientes, a dimensão geográfica). Apresentam um número muito inferior de registos em relação aos que se encontram nas tabelas de factos e estes são principalmente textuais. As tabelas de dimensão são geralmente normalizadas e

normalmente estruturadas hierarquicamente (e.g., {ano, mês, dia}, {dia, semana, ano} e {dia, mês, trimestre, ano}) (Vercellis, 2009). Isto determina como os factos podem ser agregados e selecionados (Golfarelli, Maio, & Rizzi, 1998), permitindo analisar os dados de uma forma estruturada.

Os esquemas em Constelação são modelos de dados mais complexos, onde várias tabelas de factos partilham uma ou diversas tabelas de dimensão. A Figura 2.7 apresenta um exemplo de um esquema em Constelação constituído por duas tabelas de factos (vendas e encomendas) e cinco tabelas de dimensão (produto, data, vendedor, tipoencomenda, cliente e cidade). É possível verificar que as tabelas de factos

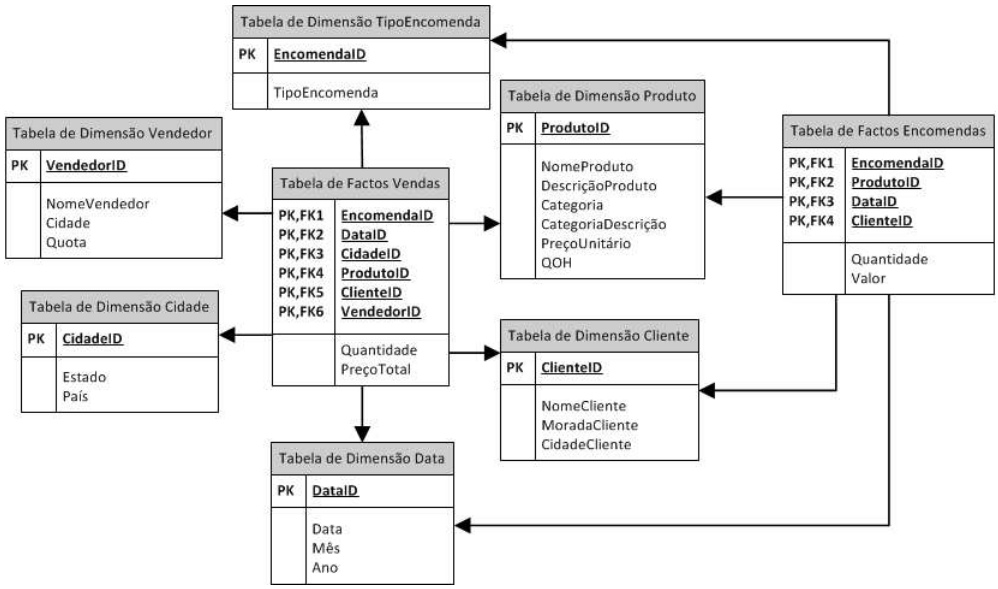


Figura 2.7: Exemplo de um esquema em Constelação

partilham quatro dimensões em comum (data, tipoencomenda, produto e cliente).

Os esquemas em Floco de Neve são esquemas em Estrela que apresentam as tabelas de dimensão completamente normalizadas. A Figura 2.8 ilustra um exemplo de um esquema em Floco de Neve constituído por uma tabela de factos (vendas) e cinco tabelas de dimensão normalizadas (vendedor, cliente, cidade, produto, data).

O esquema em Estrela e o esquema em Floco de Neve são equivalentes em termos do conteúdo de dados e interrogações que suportam, mas no entanto, o esquema em Floco de Neve apresenta uma estrutura mais complexa, que por vezes dificulta a sua compreensão (Santos & Ramos, 2009). O esquema em Floco de Neve apresenta vantagens em manter as tabelas de dimensão normalizadas (Dayal & Chaudhuri, 1997) fornecendo uma separação lógica dos dados e estando devidamente normalizado evita que seja armazenada informação redundante. Paralelamente às vantagens a si associadas, o esquema de Floco de Neve apresenta, em relação aos demais, como

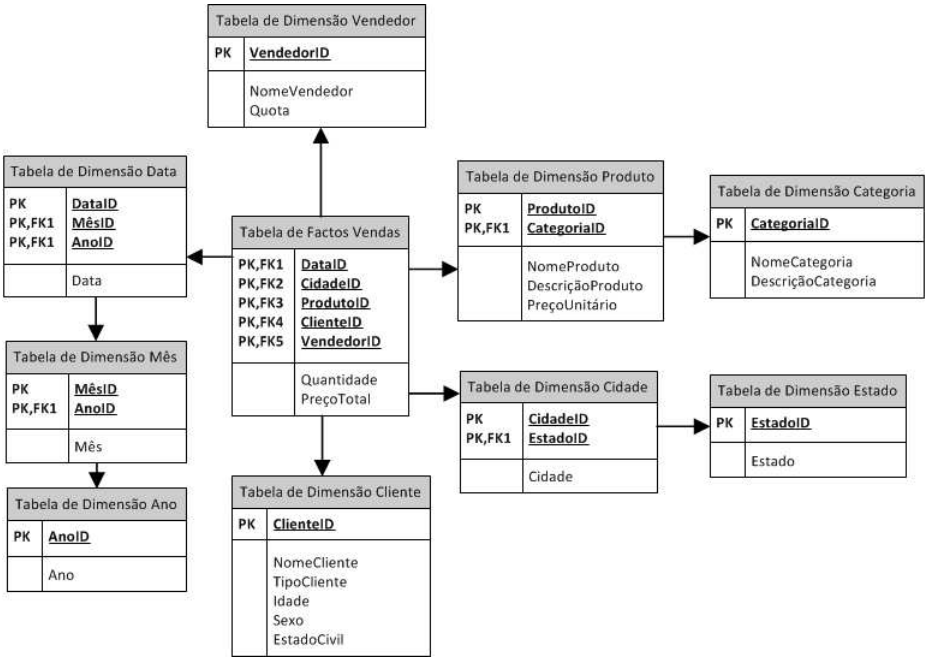


Figura 2.8: Exemplo de um esquema em Floco de Neve

desvantagem a perda de desempenho no processamento de interrogações e consultas aos dados (Levene & Loizou, 2002; Han & Kamber, 2011).

Na modelação multidimensional deve-se também ter em consideração os requisitos do negócio para os dados que se encontram nas tabelas de facto e dimensão, especialmente quando estas são atualizadas.

Ao contrário da tabela de factos que por definição representa uma série temporal de medições com uma dimensão temporal explicita (Nguyen, Tjoa, Nemec, & Windisch, 2006), as tabelas de dimensão requerem um tratamento especial. As tabelas de dimensão não sofrem alterações de uma forma previsível pois evoluem lentamente à medida que o tempo passa (Rainardi, 2008), e normalmente quando sofrem alterações são mudanças físicas (e.g., nova morada de cliente) ou correções de erros nos dados.

Este aspeto refere-se à necessidade de efetuar alterações nos atributos de uma dimensão, pois estes podem sofrer alterações ao longo do tempo e os respetivos registos necessitarem de ser alterados. Para lidar com estas alterações é necessária uma estratégia de dimensão com alteração lenta (*Slowly Changing Dimension* - SCD) (Kimball & Ross, 2002). O objetivo desta estratégia é manter a relação, sem alterações, entre a tabela de factos e as tabelas de dimensão, ou seja, sem atualizar as tabelas de factos quando os dados das dimensões estão a sofrer alterações. As alterações nas diversas dimensões só podem ocorrer quando são realizados refrescamentos no *Data Warehouse* (Santos & Ramos, 2009).

A estratégia de SCD pode basear-se num dos seguintes tipos (Kimball & Ross, 2002):

- Tipo 1 - Escrever por cima;
- Tipo 2 - Inserir um novo registo na dimensão;
- Tipo 3 - Criar atributos adicionais nas dimensões.

Escrever por cima é a técnica de atualização mais simples e rápida, porque simplesmente substitui o valor mais antigo que se encontra na dimensão por um outro valor atual sem manter qualquer registo anterior. Desta forma não é possível manter o histórico do registo que é atualizado. Este tipo de SCD é frequentemente utilizado quando, na conceção do modelo de dados, se decide que determinado atributo não deve ser armazenado em termos históricos, ou seja, não apresenta qualquer interesse em guardar o seu valor anterior (Santos & Ramos, 2009; Nguyen, Tjoa, Nemec, & Windisch, 2006). A Figura 2.9 ilustra um exemplo utilizando o SCD do Tipo 1, onde é possível observar a atualização do registo no atributo "Morada".

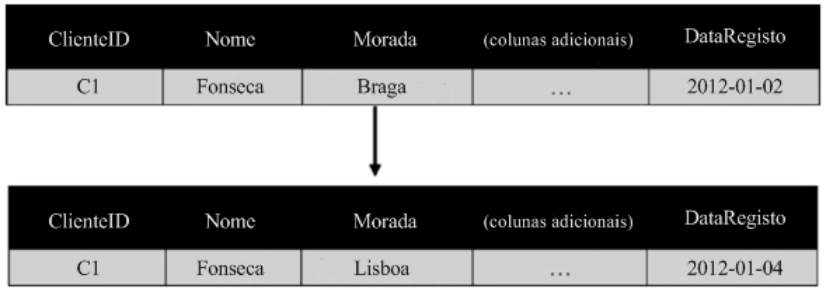


Figura 2.9: Exemplo da utilização do SCD do Tipo 1

Inserir um novo registo na dimensão é a opção de atualização mais utilizada (Nguyen, Tjoa, Nemec, & Windisch, 2006) porque permite criar um novo registo na dimensão, mantendo o histórico das alterações realizadas. No entanto, é a técnica mais complexa de implementar na medida em que é necessário associar as diversas atualizações ao registo em causa. O novo registo terá uma chave diferente da do registo original. Uma das técnicas mais utilizadas, neste caso, é a utilização de uma chave primária que integra dois componentes (Santos & Ramos, 2009), onde o primeiro classifica o registo e o segundo identifica as diversas alterações realizadas ao registo em causa. Uma outra técnica pode passar pela criação de um atributo (e.g., data de registo, estado) na dimensão que permita identificar as diversas alterações realizadas no registo em causa (Nguyen, Tjoa, Nemec, & Windisch, 2006). A Figura 2.10 ilustra um exemplo utilizando o SCD do Tipo 2.

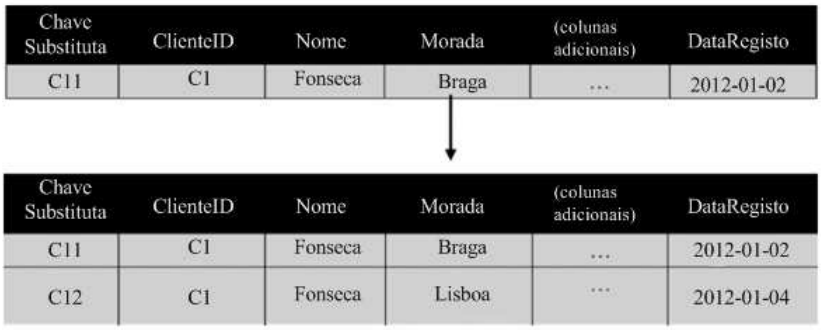


Figura 2.10: Exemplo da utilização do SCD do Tipo 2

Criar atributos adicionais nas dimensões é a terceira opção de atualização. Consiste em criar um campo adicional para os atributos (em causa) que possam sofrer alterações ao longo do tempo e registar esta mesma alteração, permitindo apenas guardar parte da história dos valores dos atributos. O facto fica associado com o valor original e o valor atual de uma coluna da dimensão. Esta opção normalmente não é utilizada porque modifica a estrutura da tabela de dimensão ao adicionar mais colunas (Nguyen, Tjoa, Nemec, & Windisch, 2006). A Figura 2.11 ilustra um exemplo utilizando o SCD do Tipo 3. Nesta é possível observar dois atributos que representam a morada do cliente: "MoradaAntiga" que corresponde à morada inicial e "MoradaAtual" que corresponde à morada atual. Caso o registo da morada seja atualizado o atributo "MoradaAntiga" vai guardar a informação da antiga morada, por sua vez, o atributo "MoradaAtual" guarda a informação da morada atual do cliente.

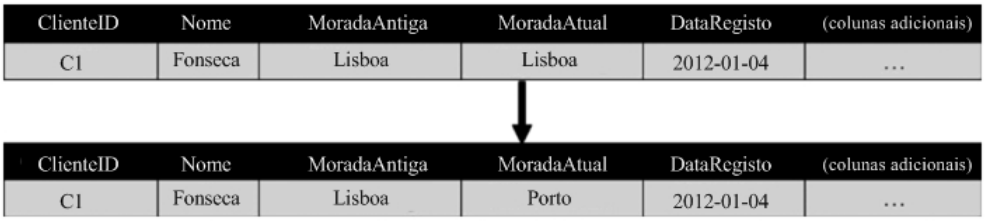


Figura 2.11: Exemplo da utilização do SCD do Tipo 3

Dos três tipos de SCD, o Tipo 2 é o mais utilizado em sistemas de *Data Warehousing*, pois contrariamente ao Tipo 1 e Tipo 3, que ao atualizar os dados existentes, eliminam a informação antiga, o Tipo 2 guarda o histórico dos dados, e desta forma respeita a característica de *Data Warehouse*, apresentado por Immon (2005), que é ser não-volátil. Na Tabela 2.2 pode-se observar, de forma resumida, as vantagens e desvantagens de cada um dos tipos. Todavia, as alterações em atributos das dimensões que ocorrem uma vez por trimestre ou menos, considera-se que estão a evoluir

Tabela 2.2: Desvantagens vs. vantagens dos tipos de SCD

	Tipo 1	**Tipo 2**	**Tipo 3**
Vantagens	Fácil de implementar;	Guarda todo o histórico;	Fácil de implementar;
Desvantagens	Perda dos valores anteriores (histórico).	Complicado de gerir e implementar.	Os valores intermédios são perdidos; Modifica a estrutura da dimensão.

lentamente. Contrariamente, se a dimensão está a sofrer alterações com uma regularidade mais frequente, por exemplo ao mês ou menos, está a evoluir rapidamente. Surge assim o conceito de dimensões com alteração rápida (*Rapidly Changing Dimension* - RCD). Normalmente pode acontecer devido ao tamanho da dimensão, à relação de um atributo com os restantes da dimensão ou devido à frequência da mudança de um atributo em relação aos demais (Rainardi, 2008). Como exemplo, têm-se uma tabela de dimensão produtos com 15 atributos, onde os atributos de 1 a 14 sofrem alterações uma vez por ano e o atributo 15 sofre alterações uma vez por dia. Reinardi (2008) sugere que se remova a coluna 15 da dimensão produto e que se coloque na tabela de factos como um atributo de dimensão degenerada (*Degenerate Dimension*)[3]. Porém, uma outra solução pode passar pelo particionamento da dimensão em uma ou mais dimensões, separando os atributos mais voláteis dos atributos menos voláteis.

2.2.1.5 Processo de ETL

Analisados os conceitos relacionados com as arquiteturas dos *Data Warehouse* e das estruturas dos modelos de dados, descrevem-se de seguida os mecanismos que permitem o povoamento dos repositórios.

Como referido anteriormente, um *Data Warehouse* é um repositório de dados, preparado especialmente para ser utilizado para apoiar a tomada de decisão nas organizações. Para que seja efetivamente possível, os dados devem ser extraídos dos sistemas OLTP, devem ser transformados e só depois carregados para o *Data Warehouse* (Ariyachandraa & Watson, 2010). A este processo dá-se o nome de ETL e inclui um conjunto de ferramentas especializadas de extração, transformação e carregamento, que permitem tratar a complexidade encontrada nos dados, tratando da sua homogeneidade, a sua posterior limpeza e o respetivo carregamento para o *Data Warehouse* (Vassiliadis, Simitsis, & Skiadopoulos, 2002). Por outras palavras, é o processo de recuperação e transformação dos dados dos sistemas OLTP para o

[3]Dimensão degenerada é uma dimensão que deriva da tabela de factos. Por essa razão é mantida na tabela de factos ao invés da criação de uma dimensão. Uma dimensão degenerada não apresenta qualquer chave de dimensão. São utilizadas, por exemplo, quando se pretende manter um identificador do sistema OLTP como o número de fatura ou número de ordem.

seu posterior armazenamento no *Data Warehouse*.

A Figura 2.12 representa o processo de um sistema ETL apresentada por Vassiliadis et al. (2002). Nos atuais sistemas de ETL, o fluxo de dados é praticamente

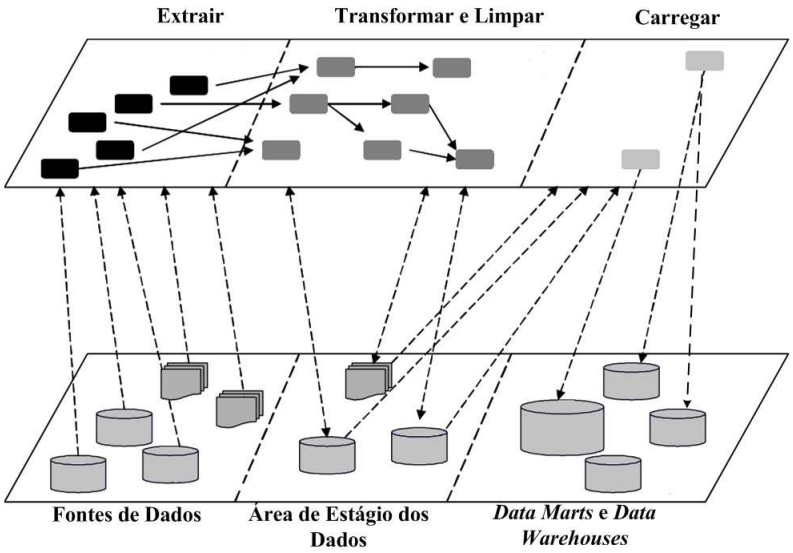


Figura 2.12: Processo de um sistema ETL

unidirecional, a partir das fontes de dados para o *Data Warehouse* (Dayal, Castellanos, Simitsis, & Wilkinson, 2009).

Os dados das diversas fontes, quer sejam internas ou externas à organização, são extraídos com os procedimentos adequados. Normalmente, o volume de dados que está a ser recuperado é grande (centenas de *megabytes* ou dezenas de *gigabytes*), e aliado ao facto dos sistemas OLTP serem desenvolvidos para suportar transações e recuperações de dados pequenas, deve-se ter em atenção a forma de como os dados são extraídos, para não condicionar a sua performance.

A extração dos dados num processo de ETL pode ser classificada em quatro abordagens (Rainardi, 2008):

- Um processo de ETL que extrai os dados do sistema OLTP regularmente. O ETL conecta-se à base de dados e consulta os dados de forma a extraí-los. Esta abordagem é a mais utilizada;
- *Triggers* na base de dados, baseadas em instruções SQL[4], que executam cada vez que ocorre uma atualização numa determinada tabela;
- Um processo agendado no sistema OLTP que extrai e exporta os dados regularmente. Esta abordagem é idêntica à primeira, no entanto esta não recorre a

[4]Linguagem de Consulta Estruturada, do inglês *Structured Query Language*, é uma linguagem declarativa de pesquisa em bases de dados relacionais.

um programa externo para consulta dos dados, mas sim um exportador interno que é executado no servidor do sistema;

- Um leitor de *log* que lê os ficheiros de *log* da base de dados do sistema OLTP de forma a identificar as alterações realizadas. Um leitor de *log* é um programa que lê e interpreta o formato dos dados no ficheiro *log*, extraindo os dados e armazenando-os num outro lugar.

As extrações dos dados podem ser distinguidas numa primeira extração, que corresponde ao primeiro carregamento dos dados para o *Data Warehouse* vazio, e as posteriores extrações incrementais dos dados que vão atualizar o *Data Warehouse* com os novos dados que vão surgindo gradualmente, da atividade diária da organização ao longo do tempo. A esta segunda extração dá-se o nome de refrescamento do *Data Warehouse* (Santos & Ramos, 2009).

Estando os dados devidamente extraídos, algumas transformações e limpezas são necessárias de forma a melhorar a sua qualidade (Han & Kamber, 2011). Assim, os dados são normalmente propagados temporariamente para a área de estágio dos dados (*Data Staging Area* - DSA) onde são transformados e limpos, antes de serem carregados para o *Data Warehouse*.

Algumas das anomalias mais habituais encontradas nos dados, removidos nesta fase do processo de ETL são (Rahm & Do, 2000; Han & Kamber, 2011; Vercellis, 2009):

- Duplicação de dados;
- Diferentes representações de valor;
- Dados com erros;
- Dados em falta.

Os problemas da qualidade dos dados estão relacionados e devem ser tratados de uma maneira uniforme (Rahm & Do, 2000; Han & Kamber, 2011) em relação a uma única ou várias fontes de dados. A Figura 2.13, adaptada de Rahm e Do (2000), apresenta a classificação dos problemas de qualidade dos dados nas respetivas fontes de dados.

O principal objetivo da atividade de transformação e limpeza passa por melhorar a qualidade dos dados que são extraídos das diferentes fontes, corrigindo as inconsistências encontradas, os erros identificados e os valores que se encontram em falta. A limpeza dos dados envolve normalmente várias fases (Rahm & Do, 2000):

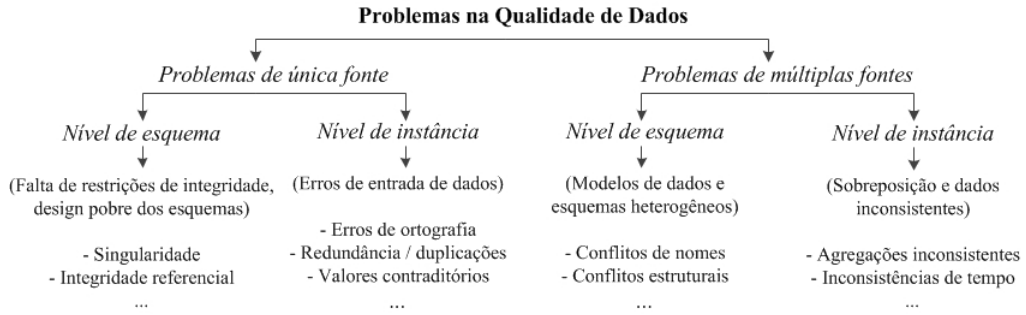


Figura 2.13: Classificação dos problemas na qualidade de dados

1. A análise e compreensão dos dados: Esta tem como principal objetivo identificar quais os tipo de erros e inconsistências que devem ser resolvidos. Para além de uma análise "manual", a utilização de aplicações de análise de dados deve ser levada em consideração, de forma a detetar problemas na sua qualidade;

2. Definição do fluxo de trabalho de transformação e regras de mapeamento: O número de transformações a realizar nos dados depende muito do número de fontes de dados e do grau da sua heterogeneidade. O fluxo de transformações e de limpeza definem o processo de ETL;

3. Verificação: O fluxo de transformação deve ser testado e avaliado, em termos de eficácia e adequação. Poderão ser necessárias múltiplas iterações de análise e verificação, uma vez que alguns problemas e erros nos dados apenas são detetados após realizadas algumas transformações;

4. Transformação: Executar transformações através do processo de ETL para carregar e refrescar os dados para o *Data Warehouse*;

5. Refluxo de limpeza se dados: Depois da identificação e correção das anomalias e dos erros encontrados, deve-se também substituir os dados com erros nas fontes de dados originais a fim de evitar refazer o trabalho de limpeza de dados para futuras extrações, que condicionam o desempenho do sistema.

Depois de concluído todo o processo de transformação, os dados que se encontram na DSA são carregados para o *Data Warehouse*, através do mapeamento entre os sistemas, respeitando as questões de integridade entre ambos.

A implementação de um processo de ETL pode ser conseguida segundo três abordagens: (i) num servidor que se encontra entre os sistemas OLTP e o servidor de *Data Warehouse*, permitindo assim que o ETL seja executado num servidor próprio para o efeito, sem condicionar o desempenho dos restantes; (ii) o processo de ETL é executado no servidor do *Data Warehouse* e (iii) o processo de ETL é executado no

servidor onde se encontram os sistemas OLTP (Rainardi, 2008). Na primeira abordagem, o processo de ETL é executado no seu próprio servidor não sendo necessário utilizar outros recursos, no entanto é mais caro porque necessita de um servidor adicional e licenças de *software*. A segunda abordagem, é mais económica pois não precisa de servidores adicionais, no entanto necessita da não operacionalização do *Data Warehouse*. A terceira abordagem, é implementada quando há necessidade de *real-time Data Warehousing*, porém apenas quando o processador apresenta capacidade para tal. Estas observações podem ser consideradas no momento da implementação de um processo de ETL.

2.2.1.6 Metadados

Associados aos sistemas de *Data Warehousing*, encontram-se também os metadados. Os metadados são dados sobre os dados, e tornam-se vitais num processo (ciclo de vida) de desenvolvimento de um *Data Warehouse* (Inmon, 2005). Normalmente, os metadados apresentam-se como: (1) metadados de negócio e (2) metadados técnicos. Os primeiros têm valor para os gestores, os segundos apresentam valor para os técnicos que desenvolvem o sistema. Porém, os metadados técnicos descrevem um *Data Warehouse* contendo (Inmon, 2005):

- Uma estrutura dos dados tal como é conhecida para os técnicos;
- Uma estrutura de dados tal como é conhecida para os gestores;
- As fontes de dados que o suportam;
- O mapeamento e transformação dos dados à medida que são carregados;
- A relação entre o modelo de dados e o *Data Warehouse*;
- Os registos de todas as atualizações e carregamentos;
- As definições e/ou descrições dos dados;
- As especificações do modelo de dados.

No armazenamento dos dados, os metadados apresentam as definições dos dados (o significado e a origem de cada coluna), das fontes de dados, da estrutura do *Data Warehouse*, dos processos de ETL e qualidade dos dados, informação de auditoria (que processos correm e quando eles correram) e informação sobre a sua utilização (que relatórios e cubos são utilizados, por quem e quando). Com isto, Rainardi (2008) apresenta sete tipos de metadados:

- Metadados de definição e mapeamento dos dados, que descrevem o significado de cada facto e tabelas de dimensão, assim como a origem dos dados e de onde estes vêm;
- Metadados de estrutura de dados, que descrevem a estrutura do modelo de dados;
- Metadados das fontes de dados, que descrevem a estrutura das bases de dados operacionais;
- Metadados do processo de ETL, que descrevem cada fluxo de dados;
- Metadados da qualidade dos dados, que descrevem as regras da qualidade dos dados, os seus níveis de risco, e as suas ações;
- Metadados de auditoria, que contêm registos dos processos e das atividades no repositório de dados;
- e, os Metadados de uso, que apresentam os *logs* de eventos e da utilização de aplicações *front-end*.

Os metadados constituem uma parte importante de um sistema de *Data Warehousing*, ajudando a compreender a estrutura, a definição e o significado dos dados de um *Data Warehouse*, assim como, entender o que realmente aconteceu e o que está a acontecer no repositório de dados.

2.2.2 Sistemas de Processamento Analítico

Os sistemas de processamento analítico (sistemas OLAP) são uma tecnologia de *Business Intelligence* utilizada para explorar um *Data Warehouse* e para apoiar as organizações nas análises efetuadas.

Um *Data Warehouse* é baseado num modelo de dados multidimensional que por sua vez possibilita uma vista sobre um cubo de dados. Os sistemas OLAP permitem recuperar e analisar informação a partir dos dados, geralmente em um *Data Warehouse* ou *Data Mart* através da criação dos cubos multidimensionais. Enquanto um *Data Warehouse* é utilizado para armazenar a informação os sistemas OLAP são utilizados para recuperá-la. Estes sistemas apresentam os dados utilizando factos, dimensões, cubos e hierarquias de conceitos através de diversas técnicas e operações de visualização. A Figura 2.14, adaptada de Han e Kamber (2011), representa um exemplo de um cubo de três dimensões (Produtos, Localização e Tempo) pelas quais as vendas podem ser analisadas.

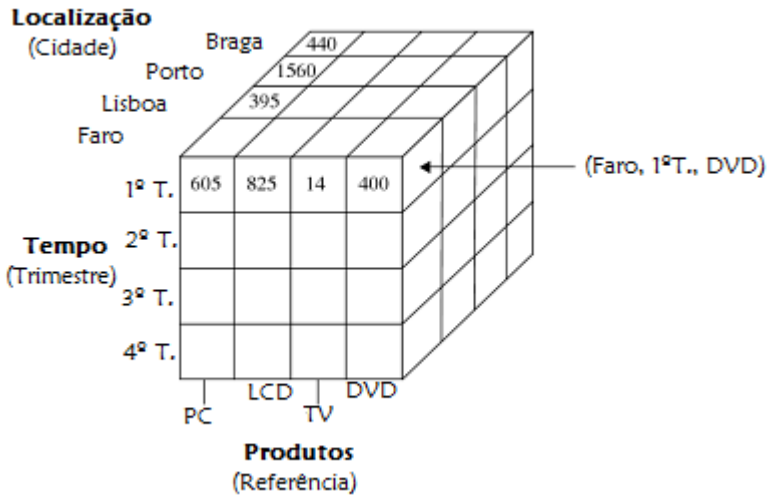


Figura 2.14: Exemplo de um cubo de três dimensões

A organização dos dados, imposta pelos esquemas multidimensionais, permite que estes sejam analisados sob diferentes perspetivas. Contudo, as análises que recorrem a este tipo de sistemas são baseadas em hierarquias de conceitos para consolidar os dados e para criar visualizações ao longo das dimensões de um *Data Warehouse* (Vercellis, 2009). Através destas o utilizador, que manuseia e visualiza o cubo OLAP, tem a possibilidade de executar as seguintes operações (Vercellis, 2009; Han & Kamber, 2011; Santos & Ramos, 2009):

- A operação de *Drill-Down* ou *Roll-Down* que permite a navegação de dados mais agregados e generalizados para dados mais detalhados. Esta operação permite uma visão mais pormenorizada dos dados que estão a ser analisados de duas formas:

  ▷ Descendo no nível da hierarquia de uma determinada dimensão. Por exemplo, para a dimensão "Localização" é possível descer de um nível "País" para o nível "Distrito" e assim sucessivamente;

  ▷ Adicionando uma nova dimensão de análise ao cubo;

- A operação de *Roll-Up* ou *Drill-Up* que representa exatamente o oposto da anterior. Esta permite agregar os dados visualizados no cubo, utilizando uma determinada hierarquia, de duas formas:

  ▷ Subindo no nível da hierarquia de uma determinada dimensão. Por exemplo, para a dimensão "Localização" é possível subir de um nível "Distrito" para o nível "País" e assim sucessivamente;

    - ▷ Eliminando uma dimensão de análise ao cubo;

- A operação de *Slice and Dice* possibilita limitar a visualização da informação com recurso ao corte (*slice*), selecionando e fixando um atributo de uma dimensão, sendo posteriormente analisado sobre as restantes dimensões, e com recurso à redução (*dice*), que permite definir um subcubo selecionando várias dimensões simultaneamente (mais de duas);

- A operação de *Pivot* ou *Rotate* que permite rodar os eixos de visualização dos dados obtendo-se uma visão diferente do cubo de dados.

Estes sistemas podem armazenar os dados sobre uma das seguintes três arquiteturas comummente aceites (Larson, 2009; Han & Kamber, 2011; Rainardi, 2008):

- *Multidimensional* OLAP (MOLAP) – Este tipo de arquiteturas suportam uma vista multidimensional dos dados que já se encontram armazenados sob a forma de cubos multidimensionais, isto é, utilizam bases de dados multidimensionais para armazenar os dados. Este tipo de arquitetura apresenta como principal vantagem o seu excelente desempenho e uma rápida indexação a dados pré-processados;

- *Relational* OLAP (ROLAP) – Esta arquitetura, contrariamente à anterior, serve como intermediário entre as bases de dados relacionais e as ferramentas de análise de dados (aplicações *front-end*). Para tal utilizam um Sistema de Gestão de Bases de Dados (SGBD) relacional para armazenar e gerir os dados que serão analisados. A arquitetura ROLAP pode armazenar grandes quantidades de dados de outras arquiteturas OLAP. A maioria dos sistemas ROLAP usa um esquema em Estrela para representar o modelo de dados multidimensional (Chaudhuri, Dayal, & Narasayya, 2011). Este tipo de arquitetura apresenta como principal vantagem a capacidade de lidar com grande quantidade de dados. No entanto, como desvantagem, apresenta dificuldades na recuperação dos valores agregados e torna-se mais lento do que qualquer outra arquitetura OLAP, pois cada análise efetuada representa uma ou mais consultas SQL à base de dados relacional, e esta quanto maior for mais demorada se torna a consulta;

- *Hybrid* OLAP (HOLAP) – Esta arquitetura combina e integra as duas anteriores, tirando o maior proveito dos seus pontos fortes e minimizando as suas fraquezas, nomeadamente a escalabilidade da arquitetura ROLAP e da capacidade de maior processamento e velocidade da arquitetura MOLAP. Assim,

esta arquitetura permite o armazenamento dos dados num repositório de dados multidimensional. Isso proporciona uma rápida recuperação dos agregados presentes em arquiteturas MOLAP. Porém, não precisa de tempo para copiar os dados, pois estes são atualizados no *Data Warehouse* ou *Data Mart*, não condicionando o seu processamento. Muitas das tecnologias disponibilizadas suportam a arquitetura HOLAP.

Estas arquiteturas, apresentadas na Figura 2.15 adaptada de (Larson, 2009), representam tipos de servidores OLAP que permitem analisar multidimensionalmente os dados sobre um qualquer tipo de repositório de dados (e.g., base de dados relacionais, bases de dados distribuídas, *Data Warehouses*) (Han & Kamber, 2011).

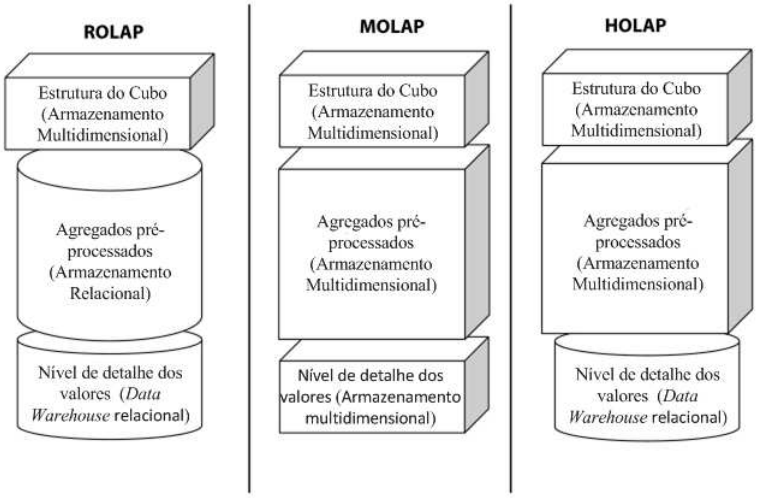


Figura 2.15: Arquiteturas OLAP

Os sistemas OLAP enquadram-se em modelos simples de análise e identificação de padrões nos dados.

2.2.3 *Data Mining*

Com o crescimento exponencial da quantidade dos dados armazenados nas organizações a sua compreensão, por parte dos seres humanos, torna-se cada vez mais trabalhosa e difícil. Por todos estes dados encontra-se informação, potencialmente útil e raramente aproveitada (Berry & Linoff, 2004). A necessidade de utilizar modelos capazes de desempenhar um papel ativo torna-se num fator crítico de sucesso (Vercellis, 2009), pois os gestores aparentemente sentem dificuldades para formular hipóteses (bem fundamentadas) sobre grandes quantidades de dados. Daí a utilidade do *Data Mining* que é um processo que utiliza técnicas estatísticas, matemáticas,

da inteligência artificial e da aprendizagem máquina para extrair e identificar informação útil de grandes conjuntos de dados. Isto é conseguido através da descoberta de padrões matemáticos, que podem ser regras, afinidades, correlações, tendências ou modelos de previsão (Turban, Sharda, & Delen, 2010). Através da utilização de algoritmos específicos e/ou mecanismos de pesquisas avançadas, identifica padrões discerníveis, relações e tendências nos dados (Fayyad, Piatetsky-Shapiro, G., & Uthurasamy, 1996).

Como tecnologia integrante de um sistema de *Business Intelligence,* o *Data Mining* pode ser aplicado sobre diversas áreas: no comércio/retalho, no processo produtivo, na saúde, na banca, na indústria agrícola, etc. Varcellis (2009) reconhece que pode ser aplicado em diversos campos como marketing relacional, vendas, reconhecimento de imagem, diagnósticos médicos, *Web Mining* e *Text Mining*, deteção de fraudes e avaliação de riscos.

2.2.3.1 Taxonomia do *Data Mining*

O *Data Mining* é baseado em duas abordagens (Berry & Linoff, 2004): (1) abordagem direta ou modelo orientado à verificação e (2) abordagem indireta ou modelo orientado à descoberta. A abordagem orientada à verificação é utilizada quando se sabe o que pesquisar, onde o utilizador formula hipóteses, refinando-a sempre que necessário, e o sistema encarrega-se de verificar essas hipóteses (Maimon & Rokach, 2010). Contrariamente, na abordagem orientada à descoberta, os padrões nos dados são identificados de forma automática sem interferência do utilizador.

A Figura 2.16, adaptada de Maimon e Rokach (2010), apresenta de forma simples uma taxonomia de *Data Mining*, em termos de abordagens, técnicas e objetivos.

As tarefas de *Data Mining* enquadram-se em duas categorias ou métodos: Os métodos de descrição e os métodos de previsão.

Os métodos de descrição são orientados à interpretação dos dados, permitindo aumentar o conhecimento acerca dos dados analisados e a relação entre eles (Maimon & Rokach, 2010), descrevendo todos os dados (Han & Kamber, 2011), reconhecendo-lhes padrões regulares e expressá-los através de regras e critérios que podem ser facilmente compreendidos (Vercellis, 2009). Por exemplo, uma organização de retalho pode utilizar técnicas de *Data Mining* para agrupar os clientes que possuem cartões de fidelidade de acordo com o seu perfil de compra, o que possibilita a identificação de novos nichos de mercado (Vercellis, 2009).

Os métodos de previsão são orientados à identificação de modelos de comportamento, capazes de prever os valores de uma ou mais variáveis a partir de valores já conhecidos de outras variáveis (Han & Kamber, 2011) relacionadas com a amos-

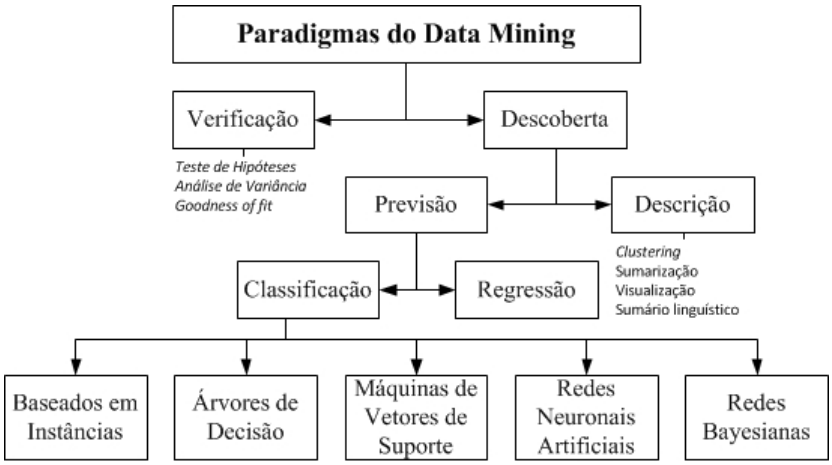


Figura 2.16: Taxonomia de *Data Mining*

tra, utilizando para isso atributos que se encontram na base de dados (Maimon & Rokach, 2010). Alguns dos métodos orientados à previsão permitem perceber os dados. Por exemplo, uma organização de retalho pode utilizar técnicas de *Data Mining* para poder prever as vendas de um determinado produto durante as semanas subsequentes (Vercellis, 2009).

Os modelos de Classificação, são modelos baseados na aprendizagem supervisionada uma vez que permitem descobrir relações entre os atributos de entrada (variáveis independentes) e os atributos de saída (variável dependente) (Maimon & Rokach, 2010). Estes permitem mapear um conjunto de dados numa classe discreta predefinida, identificando a classe a que cada registo pertence, utilizando observações realizadas no passado (conjuntos de treino) a fim de identificar um modelo que pode predizer a classe alvo das observações futuras cujos valores dos atributos são conhecidos (Vercellis, 2009). As classes representam um conjunto de valores possíveis e os atributos o valor de saída para o processo de Classificação. Assim, o desenvolvimento de um modelo de Classificação baseia-se fundamentalmente em três fases: a fase de treino, a fase de teste e a fase de previsão. As técnicas de Classificação mais comuns em *Data Mining* são as Árvores de Decisão, os Algoritmos Genéticos e as Redes Neuronais Artificiais (Chapman, et al., 2000; Maimon & Rokach, 2010).

Os modelos de Regressão são utilizados para prever atributos com valores contínuos. Com base nos atributos disponíveis o objetivo é prever, aproximadamente, o valor da variável alvo para cada observação (Vercellis, 2009). A Regressão permite, por exemplo, prever as vendas de um determinado produto dadas as suas características (e.g., preço, promoção, campanhas).

2.2.3.2 Metodologia CRISP-DM

O processo de *Data Mining* se for enquadrado numa metodologia tornar-se-á mais fácil de compreender, implementar e de desenvolver (Santos & Azevedo, 2005). As principais metodologias e mais conhecidas em *Data Mining* são: SEMMA (*Sample, Explore, Modify, Model, Assess*) e CRISP-DM (*CRoss-Industry Standard Process for Data Mining*) (Santos & Azevedo, 2005; Santos & Ramos, 2009). No âmbito deste projeto de dissertação, considerando os resultados que se esperam obter, opta-se pela metodologia CRISP-DM.

A Figura 2.17, adaptada de Chapman et al. (2000), evidencia todo o ciclo de vida da metodologia CRISP-DM que se desenvolve em seis fases independentes: **Compreensão do Negócio**, **Compreensão dos Dados**, **Preparação dos Dados**, **Modelação**, **Avaliação** e **Implementação**.

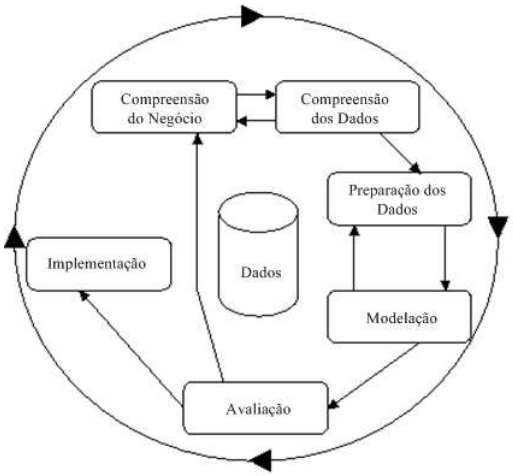


Figura 2.17: Ciclo de vida da metodologia CRISP-DM

Seguidamente é apresentada uma breve descrição de cada uma das seis fases definidas na metodologia, bem como as técnicas que se pretendem considerar neste projeto de dissertação.

Compreensão do Negócio

Na primeira fase os objetivos e os requisitos do projeto (do ponto de vista do negócio) devem ser analisados e compreendidos, transformando depois este conhecimento na definição de um problema de *Data Mining* e num plano para atingir os objetivos (Chapman, et al., 2000).

Compreensão dos Dados

Definidos os objetivos do projeto, inicia-se a fase de compreensão dos dados através da recolha e a exploração dos dados, de forma a identificar eventuais problemas.

Neste projeto a compreensão dos dados será realizada a par da compreensão do negócio da organização e da conceção de um modelo multidimensional.

Preparação dos Dados

A preparação dos dados envolve todas as atividades necessárias (e.g., processos de ETL) para a construção de um conjunto de dados finais, que posteriormente são analisados através de algoritmos de *Data Mining*. As tarefas para a preparação dos dados incluem a seleção de tabelas, registos e atributos, bem como a transformação e limpeza dos dados (Chapman, et al., 2000).

Modelação

Na fase de modelação selecionam-se as técnicas de *Data Mining* a aplicar para a concretização dos objetivos. Os parâmetros das técnicas são normalmente ajustados de forma a melhorar e otimizar os resultados obtidos (Chapman, et al., 2000). Todavia, em *Data Mining* existem várias técnicas que podem ser aplicadas ao mesmo tipo de problema, apesar de algumas técnicas terem requisitos específicos em relação ao formato dos dados. Tais restrições podem requerer o retrocesso à fase anterior (Chapman, et al., 2000).

No âmbito deste projeto de dissertação, e considerando os objetivos das análises a realizar, está previsto a utilização de técnicas de ***Clustering***, bem como **Árvores de Decisão** e **Regras de Associação**. Posteriormente, e em trabalho futuro, poderão ser incluídas outras técnicas.

As **Árvores de Decisão** permitem representar um conjunto de decisões que seguem uma hierarquia de classes ou valores. São baseadas em forma de árvore para classificar os dados num número finito de classes, tendo por base o valor dos dados de entrada. É importante referir que nestas encontram-se integrados nós (atributos a classificar), ramos (valores possíveis para os atributos) e folhas (classes em que cada registo pode ser classificado) (Santos & Ramos, 2009). Estas árvores podem ainda ser representadas por conjuntos de regras do tipo "se... então...".

As Árvores de Decisão são amplamente utilizadas em projetos de *Data Mining* devido à sua simplicidade conceptual, ao seu processamento e, acima de tudo, à capacidade de interpretação das regras que são geradas. No entanto, estas são mais adequadas para analisar dados divididos em intervalos (classes) (Turban, Sharda, & Delen, 2010).

O ***Clustering*** permite agrupar os dados em conjuntos homogéneos. Estes visam segmentar uma população heterogénea num determinado número de subgrupos (classes) compostos de observações que partilhem características similares, ao contrário das observações incluídas em grupos diferentes que apresentam características

distintas. No *Clustering* não existem classes pré-definidas para classificar os dados em estudo (Vercellis, 2009), pois estes são agrupados com base na sua homogeneidade. Diversos algoritmos podem ser utilizados para identificar segmentos nos dados. Segundo Berry e Linoff (2004), as estratégias podem passar pela divisão dos registos a segmentar ou pela agregação de registos em grupos.

As **Regras de Associação** são conhecidas também como grupos de afinidade, e são normalmente utilizadas para identificar associações interessantes e recorrentes em grupos de dados (Vercellis, 2009). Esta técnica é simples e intuitiva, e é frequentemente utilizada para investigar as transações de vendas e análise de mercado (Vercellis, 2009). Nas Regras de Associação os resultados surgem em forma de regras (A $\rightarrow$ B) onde A e B podem ser proposições verdadeiras ou falsas. Se A for verdade B também é verdadeiro, isto é, se A está presente, então B também tem probabilidade de estar presente, onde A representa um conjunto de atributos e B um único atributo que não se encontra presente em A. As regras apresentam duas medidas a si associadas: (1) suporte e (2) confiança. A medida de suporte representa o número de ocorrências de uma regra nos dados (e.g. número de vezes que A está associado a B), ou seja, são os registos que satisfazem a regra. A medida de confiança representa a probabilidade condicional da regra, sendo obtida através da divisão entre o número de casos onde A e B aparecem juntos pelo número de casos onde só aparece o B, ou seja, é o número de registos em que será possível prever B conhecendo A (Santos & Ramos, 2009).

Como exemplo, é possível determinar quais os produtos que são comprados em conjunto numa única transação e com que frequência. As organizações do setor comercial recorrem muitas vezes a Regras de Associação para reformular a posição dos produtos nas prateleiras ou em catálogos (Vercellis, 2009; Bing & Ye-bai, 2009).

O algoritmo mais utilizado é o *Apriori*, pois é o mais eficiente para extrair regras num conjunto de dados. Primeiramente, o algoritmo corre todos os dados e gera um conjunto frequente de uma forma sistemática, posteriormente extrai as regras mais fortes após receber todos os conjuntos de dados mais frequentes (Vercellis, 2009).

Avaliação

Na fase de avaliação os modelos construídos e os passos executados na sua construção são avaliados cuidadosamente, de forma a assegurar que o modelo permite atingir os objetivos do negócio considerados e, verificar também, se existe alguma variante do negócio que não foi suficientemente considerada (Chapman, et al., 2000).

Segundo Santos e Ramos (2009), o melhor modelo de previsão será aquele que apresenta a precisão mais elevada. Contrariamente, o melhor modelo de decisão é aquele que permite adquirir um conhecimento mais aprofundado e alargado dos

dados analisados. Assim, são necessárias técnicas de amostragem que permitam confrontar a precisão dos modelos concebidos, de forma a validar a sua utilidade e viabilidade.

No final desta fase decide-se o que fazer com os resultados do *Data Mining*.

Implementação

O modelo criado não significa o final do projeto. Na verdade, o objetivo do modelo é aumentar o conhecimento relativo aos dados. Porém é necessário organizar o conhecimento extraído e apresentá-lo ao utilizador para que este o possa utilizar. Dependendo dos requisitos, a fase de implementação pode ser tão simples como gerar um relatório ou pode ser tão complexa como implementar todo o processo de *Data Mining* (Chapman, et al., 2000).

2.2.4 Aplicações de *Front-End*

A prática de *Business Intelligence* passou por diversas fases evolutivas ao longo dos tempos, surgindo e extinguindo-se diversos termos relacionados com a sua prática como forma de melhorá-la. Como requisito, um sistema de *Business Intelligence* deve oferecer interfaces que facilitem ao gestor a interação e o entendimento dos dados, a fim, por exemplo, de fornecer informação adequada para a tomada de decisão (Han & Kamber, 2011), permitindo a sua manipulação, monitorização e compreensão. Uma das questões mais importantes e principal determinante do seu sucesso, a partir da perspetiva do utilizador final, é a interface "humano-computador" (Hevner, 2007 citado por (Al-Debei, 2011)). Fornecer aplicações que permitam uma interface *user-friendly* com capacidade de criação de relatórios e análises (Al-Debei, 2011) e que agreguem num único painel (acessível e de forma imediata) a informação considerada relevante para suportar o processo de tomada de decisão estratégica torna-se num fator crítico de sucesso e uma vantagem no desenvolvimento de um sistema deste tipo. É portanto indispensável disponibilizar a informação em tempo quase real (*real time*).

As tecnologias supramencionadas (subsecção 2.2.1, 2.2.2 e 2.2.3) são utilizadas para trabalhar e fornecer os dados com intuito de proporcionar aos gestores das organizações o acesso à informação. Por sua vez, as aplicações de *front-end* permitem o acesso aos dados de uma forma simples, sem grande complexidade e mais atrativa. Nestas, encontram-se portais corporativos para pesquisa, aplicações de gestão de desempenho que permitem monitorizar os KPI do negócio utilizando painéis visuais, ferramentas que permitam consultas *ad-hoc* e a análise/geração de modelos de *Data Mining* (Chaudhuri, Dayal, & Narasayya, 2011). A apresentação da informação

nestas aplicações ocorre de diversas formas como *dashboards*, tabelas, gráficos e ferramentas interativas de análise multidimensional. A Figura 2.18 ilustra um exemplo de um *dashboard* para análise de vendas, sobre diferentes perspetivas (e.g., tempo, produtos, países, etc.) e diferentes tipos de análises. Na qual é possível observar, por exemplo, o resultado das vendas trimestrais por ano representado num gráfico de barras, as vendas por categorias de produtos representadas num gráfico circular e análise das margem das vendas por produto representadas numa tabela.

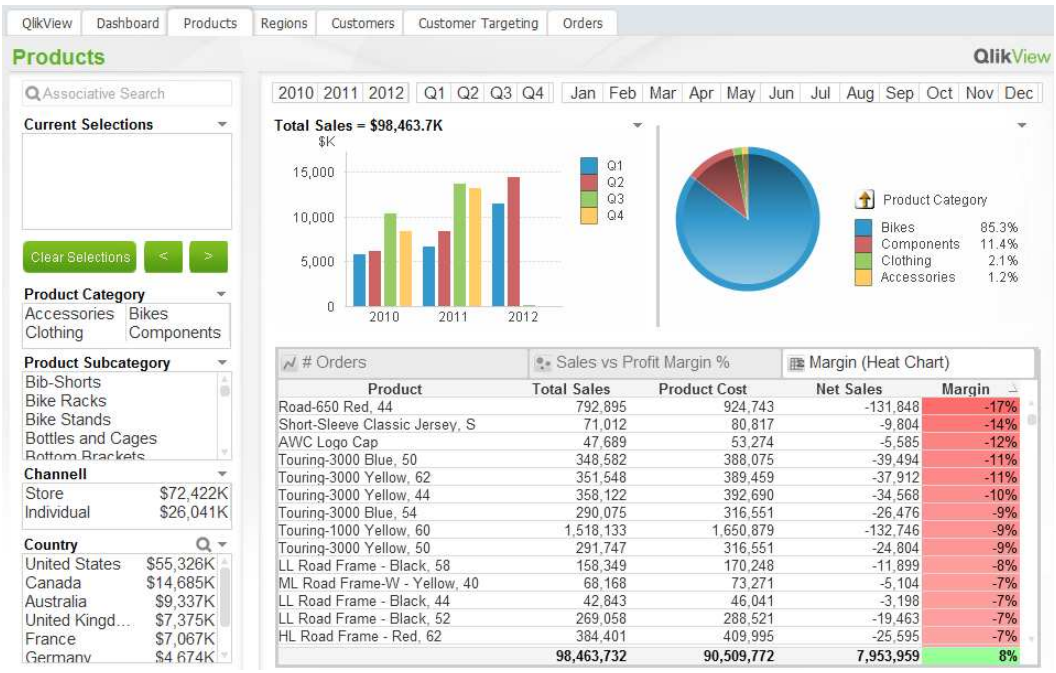


Figura 2.18: Exemplo de um *dashboard* para gestão de vendas e análise de clientes (Adaptada de (QlikView, 2012))

Os sistemas de *Business Intelligence* utilizam os KPI para avaliar o estado atual do negócio (Osadník & Landryová, 2011) e para prescrever o curso da ação face aos objetivos estratégicos das organizações. A informação que os KPI fornecem é essencial para o processo de tomada de decisão nas organizações.

Por fim, é útil referir que as aplicações de *front-end* constituem uma pequena parte, apenas 15%, de um sistema de *Business Intelligence*, enquanto 85% do trabalho efetuado está relacionado com a recuperação e transformação dos dados e o seu respetivo processamento (Laursen & Thorlund, 2010).

2.3 Aplicações de Sistemas de *Business Intelligence*

As secções anteriores proporcionaram uma introdução e revisão aos principais conceitos e tecnologias em que o *Business Intelligence* se encontra alicerçado, contribuindo desta forma para uma maior tomada de consciência, sobre o mesmo, por parte das organizações que ponderem a sua adoção. Contudo, considerou-se igualmente importante referir, ainda que de forma resumida, as possíveis aplicações destes sistemas nas organizações.

Os sistemas de *Business Intelligence* apoiam a gestão das organizações através da disponibilização de informação útil que pode ser utilizada ao longo dos três níveis da pirâmide organizacional: estratégica, tática e operacional. Ao nível da gestão estratégica, os sistemas de *Business Intelligence* tornam possível a definição de metas e objetivos estratégicos, bem como o seu respetivo acompanhamento (Olszak & Ziemba, 2007) permitindo a realização de diferentes relatórios (e.g., rentabilidade das vendas). A este nível os sistemas de *Business Intelligence* proporcionam informação sobre os diversos KPI, que permitem verificar se tais metas e objetivos estão a ser atingidos. A um nível tático possibilitam otimizar ações futuras e modificar aspetos organizacionais, financeiros ou tecnológicos do desempenho da organização (e.g., simular situações de compra e venda) a fim de ajudar a alcançar os seus objetivos estratégicos de uma forma mais eficaz. Por último, a um nível operacional, os sistemas de *Business Intelligence* são utilizados para executar análises *ad-hoc* e responder a questões relacionadas com operações das atividades da organização (e.g., clientes, vendas, fornecedores) (Olszak & Ziemba, 2007).

Segundo Chaudhuri e Narasayya (2011) será difícil encontrar uma organização bem-sucedida que não tenha aproveitado os sistemas de *Business Intelligence* para o seu negócio. Como prova, até à data, tem-se assistido a uma crescente procura de sistemas de *Business Intelligence*, e estes são normalmente aplicados em diversos sectores como os transportes, os serviços bancários, a saúde, a indústria de distribuição e produção, o retalho, as telecomunicações, e outros (Chee, et al., 2009; Olszak & Ziemba, 2006). Contudo, as necessidades das organizações variam conforme a natureza do seu negócio, e assim as necessidades dos sistemas de *Business Intelligence* variam conforme os requisitos das mesmas. Como exemplo, as necessidades de um sistema de *Business Intelligence* para uma instituição bancária são significativamente diferentes das de uma organização que atua no sector da distribuição ou da saúde.

Os efeitos no negócio mais significativos são obtidos ao utilizar as análises ofereci-

das pelos sistemas de *Business Intelligence*, como a segmentação e perfil de clientes, análise das vendas, análise da rentabilidade, análise de passivos, monitorização de custos, segmentação de mercado, análise de campanhas e de publicidade, apoio a relatórios financeiros, otimizações de logística, análises do tempo de sobrevivência, análise da lealdade dos clientes, avaliação de risco e deteção de fraudes, previsões estratégicas de desenvolvimento de processos de negócio, entre outras. (Olszak & Ziemba, 2006).

Chee et al., (2009) mostram como os sistemas de *Business Intelligence* podem ser úteis em vários setores como: produção, retalho e distribuição. Nestes sectores os sistemas de *Business Intelligence* permitem acompanhar a gestão de stocks (e.g., criando alertas para notificar índices de *stock* baixos), permitem também analisar as vendas, as compras, a distribuição e logística usando informação para agendar e consolidar as cargas de entrada e saída, permitem prever a procura por determinados serviços e/ou produtos e monitorizar a fidelidade dos clientes.

Em 2009, Zhao e Huang evidenciam como os sistemas de *Business Intelligence* podem ajudar os gestores a tomar decisões mais rápidas e precisas, através do controlo do processo de logística e da redução dos respetivos custos, através dos quais é possível analisar as vendas, tendências de mercado, comportamentos dos clientes e suas preferências, gestão do stock, inventário, produção e transporte.

No seu trabalho, Fang e Junwu (2011), implementaram um sistema de *Business Intelligence* numa organização de retalho, sobre os sistemas de CRM e ERP que esta utiliza. O principal objetivo passa por disponibilizar um sistema de apoio à gestão no processo de tomada de decisão, através de sistemas OLAP e da utilização de técnicas de *Data Mining* sobre os dados. A base de dados operacional deu origem a um *Data Warehouse* em esquema Estrela, recolhendo os dados relativos a compras, vendas, armazéns e liquidação. O sistema passou a ser utilizado pela gestão para analisar os clientes, os produtos e canais de comercialização. A utilização do sistema OLAP permitiu analisar, sobre diversas perspetivas, o lucro das vendas, relacionando as características dos diferentes tipos de clientes e dos produtos que este adquire. Sobre este projeto ainda foram utilizados algoritmos de *Data Mining* para analisar as tendências de consumo utilizando técnicas de previsão e correlação. A implementação deste sistema permitiu à organização analisar os dados de forma mais clara, promovendo desta forma a eficiência operacional, aumentando o seu lucro e aproveitando todas as oportunidades de negócio (Fang & Junwu, 2011).

Capítulo 3

Enquadramento Tecnológico

Neste capítulo é feito o enquadramento tecnológico da dissertação. Evidencia-se, de forma resumida, algumas ferramentas utilizadas no âmbito do *Business Intelligence*. Seguidamente é dado a conhecer um estudo comparativo de ferramentas de *Business Intelligence* que foram tidas em conta para o desenvolvimento do projeto de dissertação. Por último, e como o estudo incide na implementação de um sistema de *Business Intelligence*, faz sentido destacar as tecnologias utilizadas para a sua concretização.

3.1 Ferramentas de *Business Intelligence*

Depois de discutir os conceitos associados aos sistemas de *Business Intelligence*, importa agora referir as ferramentas que suportam a sua criação e gestão.

No mercado atual encontram-se uma grande diversidade de ferramentas no âmbito de *Business Intelligence* (comerciais ou não), cada uma com as suas vantagens e desvantagens em relação às demais, mas com um objetivo em comum: agrupar grandes quantidades de dados, organizando-os de uma forma unificada e clara, transformando-os em informação útil, com qualidade, para a tomada de decisão.

No que diz respeito a ferramentas de *Business Intelligence open source* verifica-se que há um número considerável de ferramentas como: *SpagoBI*[1], *OpenI*[2], *Pentaho*[3], *JasperSoft*[4] e *Palo*[5]. Todas estas ferramentas têm a capacidade de dar suporte à tomada de decisão, permitindo a estruturação e o melhoramento dos processos de negócio das organizações (Teresa & Bernardino, 2011), oferecendo um vasto con-

[1] *Http://www.spagoworld.org.*
[2] *Http://www.openi.org.*
[3] *Http://www.pentaho.com.*
[4] *Http://www.jaspersoft.com/pt.*
[5] *Http://www.palo.net.*

junto de funcionalidades (e.g., relatórios, *dashboards*, gráficos, sistemas OLAP, KPI e *Data Mining*). Por outro lado, num mercado mais competitivo, várias organizações disponibilizam ferramentas de *Business Intelligence* comerciais, como é o caso do *Oracle Business Intelligence*[6] da *Oracle*, o *Cognos*[7] da IBM, o *QlikView Business Discovery*[8] da *QlikView*, o *Business Objects Business Intelligence*[9] da SAP, SAS *Business Intelligence*[10] da SAS ou *Business Intelligence Development Studio*[11] da *Microsoft*. Para além das ferramentas referidas, que aparentemente são as mais conhecidas, muitas outras existem.

No âmbito desta análise foram identificadas diversas ferramentas de *Business Intelligence*, sendo exploradas e testadas as seguintes três: o *Pentaho* (versão *Community Edition* (CE) e versão *Enterprise Edition* (EE)) e o *QlikView*.

Partindo do principio que o desenvolvimento do sistema de *Business Intelligence* (sistema de *Data Warehousing,* sistema OLAP*, Data Mining e* aplicações de *front-end*) seria concretizado com tecnologia *Microsoft* (secção 3.2), realizaram-se estes testes de forma a encontrar alternativas viáveis e ao mesmo tempo económicas àquela que seria utilizada para o acesso e visualização da informação.

As ferramentas identificadas foram instaladas e configuradas. Posteriormente, recorrendo às tecnologias que estas possuem, foram implementadas diversas análises, através do acesso aos dados (processo de vendas) que se encontravam no *Data Warehouse*. A Figura 3.1 mostra um exemplo de um *dashboard*, implementado no *Pentaho* EE, para análise das vendas da organização. A criação do *dahsboard* implicou a construção de um cubo OLAP para o processo de vendas.

A Tabela 3.1 mostra uma análise comparativa das diversas funcionalidades que as ferramentas apresentam (consideradas no âmbito da análise). Como objetivo da análise incidiu na componente de visualização dos dados, apenas foram testadas as funcionalidades consideradas para o efeito. Ambas permitem o acesso aos dados do *Data Warehouse* concebido em SQL *Server,* e por consequente a criação de relatórios, gráficos, tabelas, cubos OLAP, consultas *ad-hoc*, *dashboards* e KPI. Relativamente à integração com o sistema OLAP implementado em tecnologia *Microsoft*, apenas o *QlikWiew* o permite fazer. Contudo, os resultados mais satisfatórios foram obtidos com a ferramenta *Pentaho EE,* seguido do *QlikWiew* e por fim o *Pentaho CE*. As primeiras duas (versões comerciais) de momento foram colocadas de lado, pelos custos a si associados. O *Pentaho CE* (versão *open-source*) foi considerada

[6] *Http://www.oracle.com/technetwork/middleware/bi-enterprise-edition/overview/index.html.*
[7] *Http://www-01.ibm.com/software/analytics/cognos.*
[8] *Http://www.qlikview.com.*
[9] *Http://www.sap.com/solutions/sapbusinessobjects/index.epx.*
[10] *Http://www.sas.com/technologies/bi.*
[11] *Http://www.microsoft.com/sqlserver/en/us/solutions-technologies/business-intelligence.aspx.*

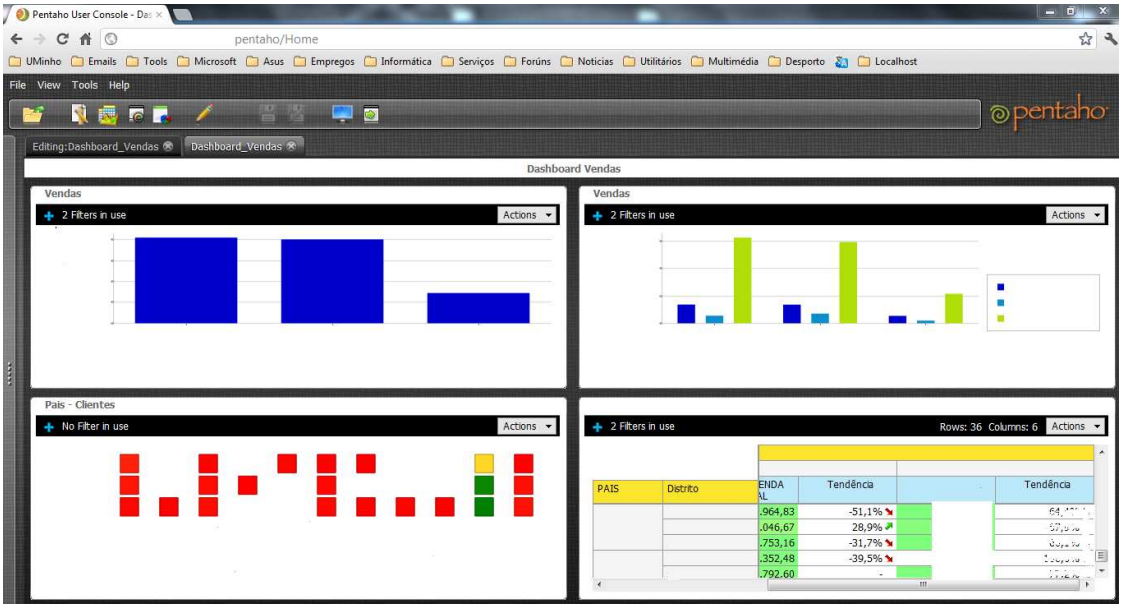


Figura 3.1: Exemplo de um *dashboard* implementado no *Pentaho* EE

e mantida no servidor de *Business Intelligence*, apesar de apresentar como limitações "o não acesso" aos cubos OLAP implementados e de requerer uma maior curva de aprendizagem para a sua utilização. No futuro esta pode vir a ser oficialmente utilizada pela organização.

Tabela 3.1: Análise comparativa das ferramentas de *Business Intelligence* exploradas

Funcionalidades Analisadas	**Ferramentas de *Business Intelligence***		
	Qlikview	***Pentaho EE***	***Pentaho CE***
Relatórios	Sim	Sim	Sim
Gráficos	Sim	Sim	Sim
Tabelas	Sim	Sim	Sim
OLAP	Sim	Sim	Sim
Consultas *ad-hoc*	Sim	Sim	Sim
KPI	Sim	Sim	Sim
Dashboards	Sim	Sim	Sim
Integração com Tecnologia *Microsoft*			
Data Warehouse	Sim	Sim	Sim
OLAP	Sim	Não	Não

De seguida descreve-se, de forma resumida, como concretizar um sistema de *Business Intelligence* com tecnologia *Microsoft*. Esta descrição é aqui realizada em virtude de ser esta a tecnologia disponível (pela organização) e utilizada no âmbito deste projeto de dissertação.

3.2 Sistema de *Business Intelligence* com Tecnologia *Microsoft*

A *Microsoft* disponibiliza uma solução de *Business Intelligence*[12], baseada numa plataforma de dados, que permite integrar e armazenar dados, efetuar análises e gerar relatórios. Trata-se de uma plataforma bastante eficiente e ao mesmo tempo intuitiva.

O SGBD *Microsoft* SQL *Server*, em conjunto com o *Microsoft Office* e o *Microsoft SharePoint* formam uma plataforma que permite armazenar e unificar os dados de uma organização, criar e gerir soluções de *Business Intelligenc*e e ainda aumentar a capacidade destas soluções de forma a dar resposta aos seus utilizadores.

A arquitetura, apresentada na Figura 3.2, baseia-se numa infraestrutura tecnológica de apoio ao *Business Intelligence*, com tecnologia *Microsoft*. Esta arquitetura

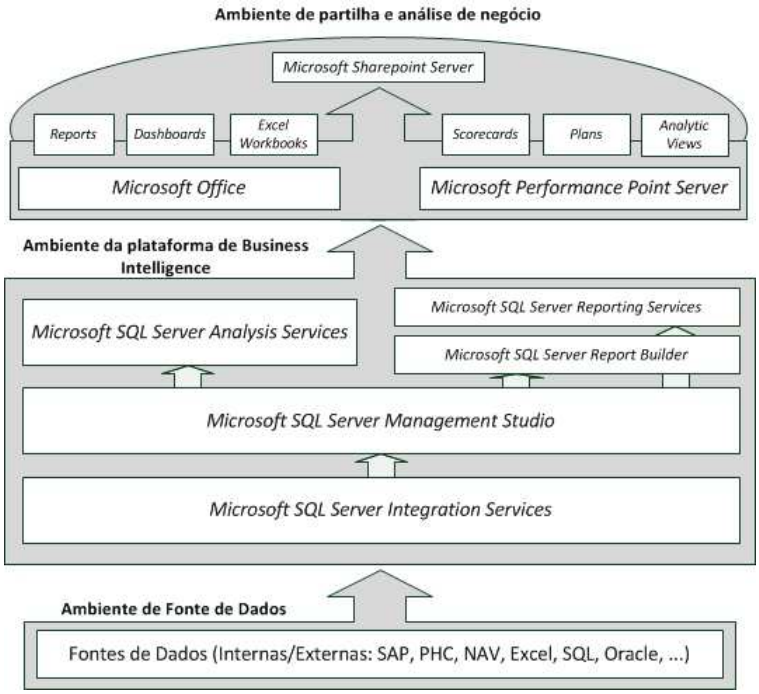


Figura 3.2: Arquitetura de *Business Intelligence* com tecnologia *Microsoft*

pode ser vista em três camadas: o ambiente de fontes de dados, o ambiente da plataforma de *Business Intelligence* e o ambiente de partilha e análise de negócio.

O ambiente de fontes de dados refere-se ao armazenamento e organização dos dados, uma vez que estes podem ter origem em vários sistemas. Podem variar de folhas de *Excel* a complexos sistemas OLTP.

[12]Na versão do *Microsoft* SQL *Server 2008* encontra-se integrada no componente *Business Intelligence Development Studio.* Na versão *do Microsoft* SQL *Server* 2012 encontra-se integrada no componente SQL *Server Data Tools.*

O ambiente da plataforma de *Business Intelligence*, neste caso *Microsoft* SQL *Server*, como o nome indica, é um servidor que disponibiliza a informação centralizada por vários pontos de uma organização. Este oferece um vasto conjunto de tecnologias que permitem criar uma plataforma completa de *Business Intelligence*: (i) as de análise e manipulação e (ii) as de publicação que utilizam a camada imediatamente a seguir para disponibilizar a informação.

O ambiente de partilha e análise de negócio é composto pela integração possível entre o *Microsoft Office* e *Microsoft Office PerformancePoin*t com o *Microsoft* SQL *Server*. Esta camada disponibiliza um vasto conjunto de ferramentas para gerir o desempenho de uma organização. Posteriormente, é possível a partilha da informação em vários formatos com recurso ao *Microsoft SharePoint Server* por toda a organização.

As tecnologias direcionadas para formar uma base de um sistema de *Business Intelligence* (Fouché & Langit, 2011) utilizadas neste projeto de dissertação são:

- *Microsoft Management Studio*;
- *Microsoft SQL Server Data Tools:*
  - ▷ *Microsoft Integration Services*;
  - ▷ *Microsoft Analysis Services*;
  - ▷ *Microsoft Reporting Services*;
- *Microsoft Report Builder*;

O *Microsoft Management Studio* constitui um ambiente integrado que permite o acesso, a configuração, a gestão e a administração dos vários componentes do *Microsoft* SQL *Server*. Possibilita o armazenamento de grandes quantidades de dados de diversas fontes. Neste componente procede-se à criação dos *Data Marts* e/ou *Data Warehouse* organizacional.

O *Microsoft Integration Services* é normalmente utilizado para a realização do processo de ETL nos dados, desde as suas origens até ao respetivo *Data Mart* ou *Data Warehouse* (Fouché & Langit, 2011). Este permite a sincronização entre o repositório de dados destino com as fontes de dados operacionais, sejam elas internas ou externas à organização, bem como a extração, limpeza e transformação dos dados, independentemente da fonte onde estes se encontram, possibilitando a unificação dos mesmos num único repositório de dados.

O *Microsoft Analysis Services* é o componente analítico que, através de vários mecanismos, permite analisar e avaliar o negócio sobre diversas perspetivas. Esta

oferece armazenamento multidimensional (MOLAP, ROLAP ou HOLAP) dos dados utilizados em cubos, assim como o seu processamento e gestão (Fouché & Langit, 2011). Segundo Fouché e Langit (2011) este componente pode ou não ser executado no mesmo servidor físico, do *Microsoft* SQL *Server*. O *Microsoft Analysis Services* permite também criar os KPI e modelos de *Data Mining* para explorar os dados de forma a encontrar padrões e realizar previsões. Possibilita a utilização de nove algoritmos de *Data Mining* (Larson, 2009), entre os quais as Árvores de Decisão, as Regras de Associação e o *Clustering*.

O *Microsoft Reporting Services* permite ao gestor criar, publicar, visualizar e distribuir relatórios detalhados sobre o negócio da organização, num ambiente *web*, através da seleção dos projetos criados no *Microsoft Analysis Services*. Por sua vez, o *Microsoft Report Builder* possibilita a criação e publicação de relatórios.

O *Microsoft* SQL Server possibilita ainda sua integração com outras ferramentas *Microsoft* como é o caso do *Microsoft Office* (*Excel*, *Visio* e *Word*) (Larson, 2009). Esta integração possibilita uma melhor e mais fácil interação entre os gestores e o sistema de *Business Intelligence*, permitindo assim a criação de interfaces gráficas para as análises OLAP, de *Data Mining* e publicação de relatórios, garantindo ações mais eficazes para a tomada de decisão.

O *Microsoft Office Excel* possibilita, de forma integrada com o *Microsoft Analysis Services* e o *Microsoft Reporting Services*, a pesquisa de informação que se encontra armazenada e, possibilita também, traduzir essa informação em KPI.

Por último, o *Microsoft SharePoint Server* surge como plataforma de colaboração e publicação, ou seja, como *front-end* para os gestores, uma vez que, de forma integrada com as anteriores, pode ser utilizada como uma plataforma de partilha de conteúdos e instrumentos de *Business Intelligence* por toda a organização (e.g., relatórios, KPI e visões analíticas sobre os dados). É importante referir que *Microsoft SharePoint Server* não foi integrado no desenvolvimento deste projeto.

Capítulo 4

Sistema de *Business Intelligence* no Suporte à Gestão Estratégica

Este capítulo incide essencialmente sobre dois tópicos que estão relacionados com o trabalho realizado. Primeiramente é efetuada uma apresentação da organização, do modelo de controlo de gestão, dos requisitos e suas necessidades de informação para a implementação de um sistema de *Business Intelligence*. Em seguida, os dados operacionais são caracterizados e apresenta-se a arquitetura do sistema de *Business Intelligence* a implementar.

4.1 Caso de Estudo

4.1.1 Descrição da Organização

A organização atua na área da Engenharia Eletrónica Industrial e apresenta como ramo de atividade o comércio por grosso de equipamentos eletrónicos.

Inicialmente, e de forma a posicionar-se estrategicamente no mercado, a organização foi pensada e assente na construção e automação industrial, mas paralelamente foi consolidando a importação e distribuição de produtos de segurança como *Closed Circuit Tele Vision* (CCTV), sistemas de alarme, sistemas de deteção de incêndio, controlos de acessos e várias áreas complementares. Porém, com o passar dos anos, a organização foi introduzindo novas áreas de negócio, acompanhando as tendências do mercado, comercializando linhas completas de CCTV, intrusão, incêndio, comunicações, automação, domótica, *digital signage* e controlo de acessos. Hoje, já com forte presença no mercado, a organização disponibiliza soluções de Engenharia de hardware e software, bem como tecnologia, consultadoria e produtos para sistemas de segurança.

Contudo, pela dimensão que atingiu, os gestores desta organização têm vindo, cada vez mais, a sentir dificuldades em tomar decisões estratégicas, baseadas na informação das atividades diárias. A utilização de simples folhas de cálculo, com informação pouco detalhada, torna este processo moroso e incompleto. E embora seja possível extrair informação de gestão diretamente a partir das fontes de dados operacionais, surgem inconvenientes e limitações de desempenho, tempo/recursos e até mesmo da instabilidade dos dados.

De forma a ultrapassar estas dificuldades, a organização procura uma solução que lhe permita alcançar melhor qualidade na tomada de decisão. A Obter informação que apoie a tomada de decisão a um nível estratégico, com implicações a um nível tático e operacional, utilizando dados históricos e atuais, e visualizando-os através de relatórios, gráficos e KPI com recurso a ferramentas de processamento analítico de dados e *Data Mining*, tornou-se um requisito fundamental na concretização da estratégia da organização.

4.1.2 Estratégia de Negócio

A estratégia da organização é orientada para a persecução e concretização dos objetivos, assentando em quatro níveis essenciais, o organizacional, o de negócio, o da tomada de decisão e o de competências. Ao nível do negócio, a estratégia da organização passa pela diversificação através da diferenciação do produto. Paralelamente, a concretização de parcerias estratégicas, quer a nível nacional ou internacional, garantem a presença dos seus produtos a uma escala mundial. A realização de ações de formação e especialização aos seus clientes, a participação em feiras tecnológicas nacionais e internacionais, os serviços *web* orientados para o cliente, assim como um serviço complementar de assistência técnica, mostram ser linhas de ação para a concretização da estratégia da organização. Ao nível do processo de tomada de decisão, a organização favorece a identificação de novas oportunidades de negócio, dando destaque à expansão de novos produtos e mercados.

No geral a estratégia da organização encontra-se focada num crescimento sustentado e na internacionalização, sendo esta concretizada ao nível do negócio através da diversificação e diferenciação orientada para o mercado. Com isto, a organização pretende desenvolver e expandir o seu negócio, através da oferta de novas soluções dentro das suas linhas de produtos, aumentando desta forma a sua quota de mercado.

4.1.3 Modelo de Controlo de Gestão Estratégica

A organização necessita de suporte ao nível da tomada de decisão, através da concretização de um sistema de *Business Intelligence*, que possibilite o acesso a informação sistematizada de compras e vendas dos seus produtos, bem como informação sobre a sua movimentação em stock (saídas e entradas) em termos de custos e quantidades, e informação relacionada com as contas correntes dos seus clientes (saldos, saldo médios, etc.). Paralelamente a organização pretende obter informação que associe a assistência técnica prestada pelos seus colaboradores aos clientes, das chamadas telefónicas efetuadas e recebidas (*help desk*), dos RMA (Retorno de Material Avariado) e de uma série de KPI contabilísticos e financeiros (enquadrados nos segmentos de análise a apresentar).

Deste modo fez-se, junto dos responsáveis da organização, um levantamento das necessidades de informação (KPI e análises mais predominantes) para suportar a tomada de decisão. O resultado das análises e KPI é apresentado Tabela 4.1.

Tabela 4.1: Principais análises e KPI

Análises e KPI	
Evolução das vendas no tempo	Crescimento das vendas
Crescimento absoluto das vendas	Vendas por períodos de tempo
Rentabilidade por área de negócio	Rentabilidade por mercado
Lucro bruto	Margem de lucro bruto
Evolução das vendas por mercado	Evolução da faturação por distrito
Variação das vendas por mercado	Variação das vendas por distrito
Variação das vendas por área de negócio	Evolução das vendas por área de negócio
Base de crescimento de clientes	Clientes ativos vs inativos
Peso top 10 cliente na faturação	Top produtos e clientes
Peso da área de negócio na faturação	Evolução da exportação
Vendas por vendedor	Top clientes por vendedor

A partir das necessidades identificadas e tomando como referência um diagnóstico realizado por parte da organização, o sistema a implementar deverá estar preparado para suportar a tomada de decisão, apurando custos e resultados nos seguintes segmentos de análise (Figura 4.1).

Os resultados por segmentos são organizados numa lógica de valor e não meramente com base em valores contabilísticos. Tendo por base este principio os indicadores devem assentar no custeio direto e na Margem de Contribuição Residual[1] (MCR) (em detrimento do resultado contabilístico). Quer isto dizer que em cada

[1]Internacionalmente assume a designação de *Economic Value Added* (EVA).

Figura 4.1: Segmentos de análise

um dos segmentos (e.g., clientes, produtos) são apenas considerados os custos diretamente gerados por esses objetos (e.g., custos dos produtos, serviços, transportes). Procura-se, deste modo, minimizar a repartição arbitrária de custos e a distorção da informação financeira.

Para alguns dos segmentos de análise (Mercado, Responsabilidade e Negócio) adota-se a MCR, indicador que, para além de considerar os proveitos e custos associados a um dado segmento, incorpora o custo de financiamento do investimento a si associado. Por exemplo, o resultado de um cliente contempla os custos e proveitos de exploração por ele gerados, mas também o custo de financiamento do crédito que lhe é concedido, num dado período. A MCR traduz o resultado do segmento, depois de remunerado o investimento.

Forma de apuramento da MCR

1. Proveitos (e.g., vendas, prestação de serviços)
2. Custos Diretos (e.g., custo do produto, transporte, comissões, custos com pessoal)
3. Margem de Contribuição (1-2)
4. Custo de financiamento do investimento (ou seja, dos ativos económicos)
5. MCR (3-4)

O custo de financiamento do investimento é apurado pela multiplicação de uma taxa de custo de capital pelo valor dos ativos económicos associados ao segmento em causa (Taxa de Custo de Capital * valor do Ativo Económico). A Taxa de Custo de Capital traduz o custo médio de financiamento da organização, que considera não apenas o custo do financiamento externo (e.g., banca), mas também o custo de oportunidade associado ao investimento dos acionistas.

O Ativo Económico contempla ativos e passivos de exploração do segmento de análise, pelo que serão normalmente do tipo:

- Imobilizado;
- Necessidades em Fundo de Maneio (crédito a clientes/fornecedores, existências em *stock* e outros ativos/passivos associados à exploração).

Nas subsecções que se seguem apresenta-se individualmente cada um dos segmentos de análise, descrevendo as necessidades informacionais inerentes a cada um deles e os principais KPI de análise.

4.1.3.1 Segmento de Negócio

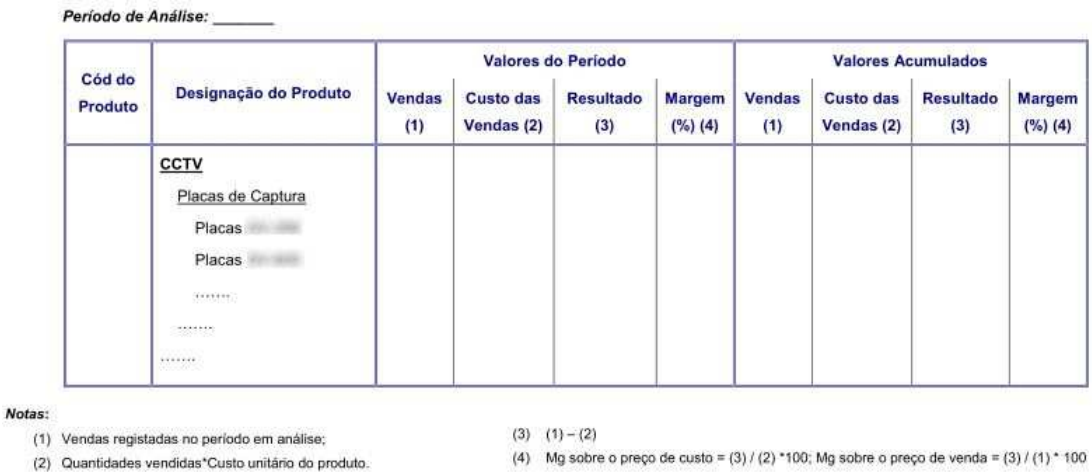
Com o segmento do Negócio pretende-se conhecer os resultados, margens e rentabilidade por área de negócio, famílias de produtos e/ou produtos ao longo do tempo. Para tal, deverão ser consideradas as vendas e os custos dos produtos vendidos. Procura-se responder a uma série de questões como: qual a rentabilidade do produto A?, Qual o produto que mais contribuiu para os resultados deste mês?, Para que clientes foram efetuadas estas vendas?, etc.

Deste modo, o sistema deverá disponibilizar ao gestor informação sobre resultados e rentabilidade por produto ou famílias (se possível, cruzando com o cliente). A Tabela 4.2 apresenta, de forma resumida, os principais KPI a considerar para o segmento de Negócio. Na Figura 4.2 evidência, como exemplo, a configuração de

Tabela 4.2: Principais KPI para o segmento de Negócio

Principais KPI do segmento do Negócio
Vendas líquidas no período de análise
Custo do produto no período de análise
Margem de Contribuição Bruta (MCB)
Custo com as áreas de negócio
Margem de Contribuição (MC)
Stock médio no período de análise
Custo de financiamento do ativo económico
Rentabilidade
MCR
Taxa de MCR

um relatório relativo às margens por produto, famílias ou área de negócio.



Período de Análise: _____

Cód do Produto	Designação do Produto	Valores do Período				Valores Acumulados			
		Vendas (1)	Custo das Vendas (2)	Resultado (3)	Margem (%) (4)	Vendas (1)	Custo das Vendas (2)	Resultado (3)	Margem (%) (4)
	CCTV								
	Placas de Captura								
	Placas								
	Placas								
	.....								
	.....								
	.....								

Notas:

(1) Vendas registadas no período em análise;

(2) Quantidades vendidas*Custo unitário do produto.

(3) (1) – (2)

(4) Mg sobre o preço de custo = (3) / (2) *100; Mg sobre o preço de venda = (3) / (1) * 100

Figura 4.2: Exemplo da estrutura de um relatório para análise de margens por área de negócio

4.1.3.2 Segmento do Mercado

Com o segmento do mercado pretende-se conhecer o resultados, margens e rentabilidades por cliente ou famílias de clientes. Para estas análises deverão ser selecionados os clientes para os quais se considera relevante conhecer, em separado, a sua contribuição e rentabilidade para a organização. Os clientes analisados, de forma isolada, serão certamente clientes recorrentes ou habituais e responsáveis por um volume de faturação significativo. Todos os restantes clientes poderão ser agrupados em famílias (e.g., localização, top 10).

A Tabela 4.3 apresenta, de forma resumida, os principais KPI a considerar para o segmento do Mercado. Neste sentido, o sistema deverá disponibilizar informação

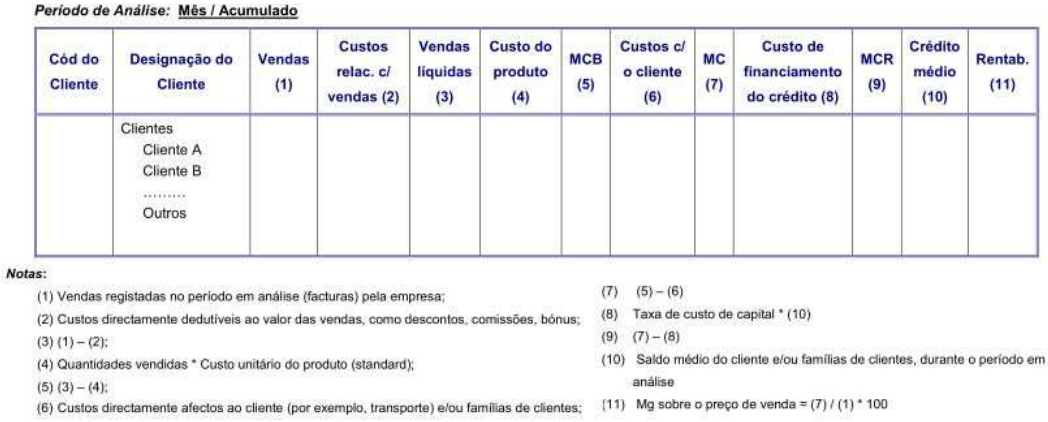
Tabela 4.3: Principais KPI para o segmento de Mercado

Principais KPI do segmento do Mercado
Vendas líquidas no período de análise
Custo do produto no período de análise
Margem de Contribuição Bruta (MCB)
Custo com os clientes
Margem de Contribuição (MC)
Crédito médio concebido no período de análise
Custo de financiamento do crédito
Rentabilidade
MCR
Taxa de MCR

da rentabilidade consolidada com base nas vendas efetuadas ao cliente, os custos

a si associados e a identificação do crédito concedido aos clientes e seu impacto financeiro.

A Figura 4.3 mostra, como exemplo, a configuração de um relatório referente aos resultados por cliente e/ou famílias de clientes.



Período de Análise: Mês / Acumulado

Cód do Cliente	Designação do Cliente	Vendas (1)	Custos relac. c/ vendas (2)	Vendas liquidas (3)	Custo do produto (4)	MCB (5)	Custos c/ o cliente (6)	MC (7)	Custo de financiamento do crédito (8)	MCR (9)	Crédito médio (10)	Rentab. (11)
	Clientes Cliente A Cliente B ..... Outros											

Notas:

(1) Vendas registadas no período em análise (facturas) pela empresa;
(2) Custos directamente dedutíveis ao valor das vendas, como descontos, comissões, bónus;
(3) (1) – (2);
(4) Quantidades vendidas * Custo unitário do produto (standard);
(5) (3) – (4);
(6) Custos directamente afectos ao cliente (por exemplo, transporte) e/ou famílias de clientes;
(7) (5) – (6)
(8) Taxa de custo de capital * (10)
(9) (7) – (8)
(10) Saldo médio do cliente e/ou famílias de clientes, durante o período em análise
(11) Mg sobre o preço de venda = (7) / (1) * 100

Figura 4.3: Exemplo da estrutura de um relatório para análise dos resultados por clientes

4.1.3.3 Segmento da Responsabilidade

O segmento da responsabilidade deverá apurar os resultados por vendedor.

Um pouco à imagem do que acontece com os anteriores segmentos de análise, é de grande utilidade a identificação de todos os custos por vendedor, das vendas e do crédito concedido aos seus clientes. A Tabela 4.4 apresenta, de forma resumida, os principais KPI a considerar para o segmento da Responsabilidade. A Figura 4.4

Tabela 4.4: Principais KPI para o segmento da Responsabilidade

Principais KPI do segmento do Mercado
Vendas líquidas no período de análise
Custo do produto no período de análise
Margem de Contribuição Bruta (MCB)
Custos comerciais com o vendedor
Margem de Contribuição (MC)
Crédito médio concebido ao cliente no período de análise
Custo de financiamento do Ativo Económico
Imobilizado líquido médio
Rentabilidade
MCR
Taxa de MCR

mostra, como exemplo, a estrutura de um relatório relativo à análise dos resultados por vendedor.

Período de Análise: **Mês / Acumulado**

Cód do Vendedor	Nome do Vendedor	Vendas (1)	Custos relac. c/ vendas (2)	Vendas liquidas (3)	Custo do produto (4)	MCB (5)	Custos Comerciais (6)	MC (7)	Custo de financiamento do crédito (8)	MCR (9)

Notas:

(1) Vendas registadas no período em análise (facturas) pelo vendedor;
(2) Custos directamente dedutíveis ao valor das vendas, como descontos, comissões, bónus;
(3) (1) – (2);
(4) Quantidades vendidas * Custo unitário do produto (standard);
(5) (3) – (4);
(6) Custos directamente afectos ao vendedor (remunerações do vendedor, alimentação, transporte e outros custos específicos);
(7) (5) – (6)
(8) Taxa de custo de capital * (7)
(9) (7) – (8)

Figura 4.4: Exemplo da estrutura de um relatório para análise dos resultados por vendedor

4.1.3.4 Segmento dos Processos

O segmento dos processos visa conhecer os custos das atividades desenvolvidas pela organização. A identificação dos custos por atividade, nomeadamente das atividades de suporte/assistência técnica (e.g., *help desk*, atendimento ao cliente) é uma análise fundamental. Deste modo, o sistema deve permitir identificar quais os clientes que "gastam" mais recursos (ao nível de suporte técnico) e que apresentam custos mais elevados nas diversas atividades da organização (e.g., técnicas, comerciais), bem como relacionar as chamadas telefónicas (tempo e quantidade) por colaboradores e entidades externas à organização, ao longo do tempo.

O modelo apresentado é a base para a conceção e implementação do sistema de *Business Intelligence*, uma vez que incide sobre diversos processos de negócio da organização. No entanto, é importante referir que muitos outros KPI são considerados. Estes dependem das dimensões que posteriormente vão ser apresentadas. O acompanhamento da evolução das vendas, o crescimento absoluto, taxas de crescimento, variações da faturação, a variação média, taxas de variação, a evolução do mercado nacional e internacional, a rentabilidade por localização geográfica, o peso que determinada área de negócio ou produto têm nas vendas, os clientes que mais contribuem para o sucesso da organização, os que menos interessam, entre outros, são fatores de análise predominantes para o processo de tomada de decisão da organização.

Na próxima secção apresentam-se as características dos Sistemas de Informação e das fontes de dados da organização.

4.2 Caracterização e Exploração dos Dados

O desenvolvimento deste projeto iniciou-se com uma abordagem *top-down* (referida na secção 2.2.1.2) depois de reunidos os requisitos necessários e as necessidades informacionais para a sua concretização.

A organização utiliza o sistema ERP PHC *Advanced* como **Sistemas de Informação** para integrar todos os dados e processos de negócio. É expectável que a informação registada num sistema deste tipo seja avultada, uma vez que gere toda a informação, desde encomendas, registo de vendas e compras, RMA, CRM, recursos humanos, movimentos de produtos e *stocks*, e outros.

A estrutura de dados do ERP encontra-se desenvolvida em SQL *Server*, o que facilita o desenvolvimento do processo de integração de dados, evitando assim constrangimentos tecnológicos. Contudo, a sua complexa arquitetura (composta por cerca de 6 (seis) mil tabelas e 10 (dez) *gigabytes* de dados) exigiu um maior esforço de análise. Visto tratar-se de uma base de dados complexa a sua análise poder-se-á tornar inexequível se não forem tomadas medidas para controlar essa complexidade. Como tal, foram implementadas *views*[2], das tabelas consideradas prioritárias, de forma a facilitar o acesso aos dados necessários ao sistema de *Business Intelligence.* A Figura 4.5 representa o modelo Entidade-Relacionamento das *views.*

Porém, com a necessidade de analisar as chamadas telefónicas registadas na organização, optou-se por integrar a respetiva base de dados (uma única tabela (Figura 4.6)) no modelo anterior (BI_Chamadas). Esta tabela regista a informação de todas as chamadas telefónicas (internas e externas), nomeadamente: o nome do colaborador (nome), a extensão do telefone do colaborador (extensão), o tipo de chamada (*way*), o número do telefone que contactou com o colaborador (*clid*), a data/hora de inicio (*start*) e de fim (*finish*), e a duração da chamada (tempo).

Como se pode verificar, na Figura 4.5, o modelo criado é composto por 30 (trinta) tabelas, entre as quais se destacam: a BI_VendasLIN (linhas dos documentos de faturação - produtos, áreas de negócio, armazém, quantidades, preço de venda, preço de custo, etc.), a BI_VendasCAB (cabeçalhos dos documentos de faturação - tipo documento, clientes, serviços, etc.), a BI_ComprasCAB (cabeçalhos dos documentos de compra), a BI_ComprasLIN (linhas dos documentos de compra - produtos, áreas

[2]Uma *view é* considerada uma vista/tabela virtual. Possibilita uma forma alternativa de visualizar as tabelas que compõe uma base de dados.

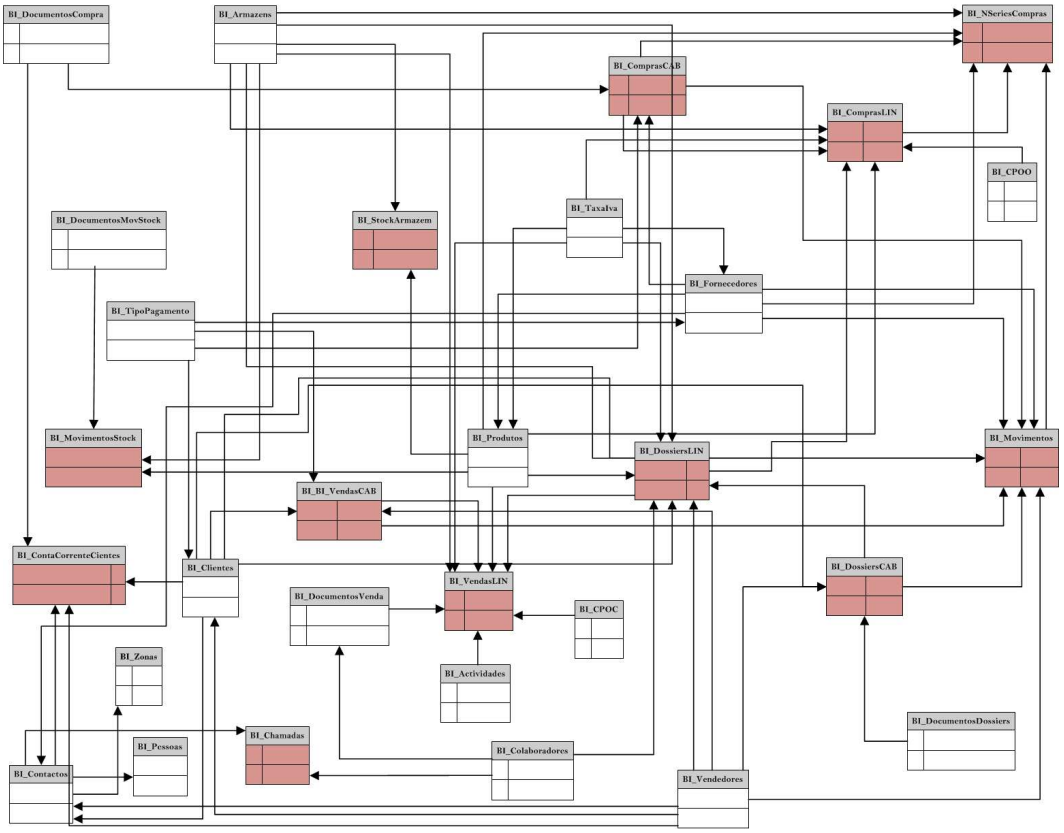


Figura 4.5: Modelo de dados Entidade-Relacionamento das *views*

de negócio, armazém, quantidades, preço de compra, etc.), a BI_DossiersCAB (cabeçalhos dos documentos internos da organização (e.g., RMA, encomendas, etc.)), a BI_DossiersLIN (linhas dos documentos internos), a BI_MovimentosStock (entradas e saídas em armazém - quantidade, custo ponderado, etc.), a BI_StockArmazem (stock dos produtos em armazém); a BI_Movimentos (movimentos de todos o número de série dos produtos); a BI_NSeriesCompras (números de serie dos produtos comprados); a BI_ContaCorrenteClientes (movimentos da conta corrente dos clientes - crédito, débito, saldo, etc.) e a BI_Chamadas (chamadas internas e externas à organização). A Figura 4.7 mostra, como exemplo, parte da estrutura e registos da tabela BI_VendasLIN.

As restantes tabelas guardam informação sobre as diversas componentes do negócio da organização. A BI_Produtos armazena a informação dos produtos e das respetivas famílias; A BI_Contactos, a BI_Clientes e a BI_Fornecedores armazenam informação dos contactos, dos clientes e dos fornecedores da organização respetivamente; A BI_CPOC e a BI_CPOO armazenam a informação relacionada com as contas do plano oficial contabilístico/sistema de normalização contabilística associadas às vendas e compras respetivamente; A tabela BI_Pessoas armazena

clid_uid	ext_uid	extensao	nome	way	clid	periods	start	finish	tempo
1346316309.110	1346316310.111	650	[illegible]	int	607	3	2012-08-30 09:45:17.000	2012-08-30 09:46:17.000	60
1346318010.229	1346318025.241	650	[illegible]	ext-in	[illegible]	1	2012-08-30 10:13:46.000	2012-08-30 10:21:36.000	470
1346322156.425	1346322228.436	650	[illegible]	int	650	7	2012-08-30 11:24:08.000	2012-08-30 11:24:30.000	22
1346322228.441	1346322228.442	650	[illegible]	int	91605	1	2012-08-30 11:24:08.000	2012-08-30 11:24:08.000	0
1346322603.463	1346322603.464	650	[illegible]	int	607	0	2012-08-30 11:30:11.000	2012-08-30 11:30:11.000	0
1346333143.1060	1346333149.1072	650	[illegible]	ext-in	[illegible]	1	2012-08-30 14:25:49.000	2012-08-30 14:31:16.000	327
1336576842.2351	1336576842.2350	651	[illegible]	ext-out	[illegible]	1	2012-05-09 16:20:52.000	2012-05-09 16:24:05.000	193
1336579792.2793	1336579792.2792	651	[illegible]	int	92614	1	2012-05-09 17:10:07.000	2012-05-09 17:13:11.000	184
1336581492.3097	1336581492.3094	651	[illegible]	ext-in	[illegible]	9	2012-05-09 17:38:34.000	2012-05-09 17:40:47.000	133
1336584178.3371	1336584178.3370	651	[illegible]	int	[illegible]	1	2012-05-09 18:23:01.000	2012-05-09 18:24:21.000	80

Figura 4.6: Exemplo da tabela BI_Chamadas

	fistamp	ftstamp	fno	ref	codproduto	familia	nmdoc	snc	ndoc	servicos	codareaneg	actividade	qtt	desconto	esltt	eecoval
349	[illegible]71466353,9700000-2	[illegible]466353,798255596	11037135	[illegible]	01030402000072	0[illegible]	Factura	1	1	0	01	ESTRUTURA	3.000	0.00	229.050000	0.250000
350	[illegible]71469115,642000004	[illegible]469115,548673817	11037141	[illegible]	01030402000072	0[illegible]	Factura	1	1	0			5.000	10.00	343.700000	0.250000
351	[illegible]41664576,2500000-1	[illegible]664576,140412350	12042069	[illegible]	01030402000072	0[illegible]	Factura	1	1	0	01	ESTRUTURA	1.000	0.00	76.350000	0.250000
352	[illegible]41366591,9320000-3	[illegible]366591,526081704	12042043	[illegible]	01030402000072	0[illegible]	Factura	1	1	0	01	ESTRUTURA	4.000	0.00	305.400000	0.250000
353	[illegible]10464920,5780000-2	[illegible]464920,468175013	12040273	[illegible]	01030402000072	0[illegible]	Factura	1	1	0	01	ESTRUTURA	2.000	0.00	152.700000	0.250000
354	[illegible]10466899,499000-11	[illegible]466695,390013846	12005078	[illegible]	01030402000072	0[illegible]	Guia de Remessa	1	2	0			1.000	10.00	68.740000	0.250000
355	[illegible]92056150,1850000-1	[illegible]056150,154716616	10030655	[illegible]	01030402000072	0[illegible]	Factura	0	1	0			1.000	0.00	77.250000	0.250000
356	[illegible]00736038,2800000-2	[illegible]736038,124124269	10030995	[illegible]	01030402000072	0[illegible]	Factura	0	1	0			4.000	10.00	278.200000	0.250000
357	[illegible]00766661,6240000-7	[illegible]766661,265009552	10031022	[illegible]	01030402000072	0[illegible]	Factura	0	1	0			2.000	0.00	154.500000	0.250000
358	[illegible]02765779,6750000-3	[illegible]2765779,472062921	10031468	[illegible]	01030402000072	0[illegible]	Factura	0	1	0			2.000	0.00	154.500000	0.250000

Figura 4.7: Exemplo da tabela BI_VendasLIN

a informação sobre os funcionários do clientes e fornecedores; A tabela BI_Zonas armazena informação relativa às localizações geográficas (concelho e distrito) das diversas entidades; As tabelas BI_DocumentosVenda, BI_DocumentosCompra e BI_DocumentosDossiers armazenam informação dos diversos documentos de venda, compra e dossieres internos da organização, respetivamente; A tabela BI_Armazem armazena a informação do vários armazéns; A BI_DocumentosMovStock armazena informação sobre os documentos que originam um movimento em stock; As tabelas BI_Colaboradores e BI_Vendedores armazenam informação dos colaboradores (funcionários internos) e dos vendedores (comerciais) da organização respetivamente; A BI_TipoPagamento armazena a informação dos possíveis tipos pagamento à organização ou clientes; A BI_TaxaIVA armazena a informação relativa às taxas de IVA inerentes ao negócio.

As *vie*ws tornam-se, neste projeto, a única fonte de dados e representam, no caso concreto da organização, um subconjunto de dados referentes ao negócio existente e que residem na base de dados do sistema OLTP. Estas visam caracterizar o negócio em termos de processos de compra e venda de produtos e/ou serviços para com as entidades externas à organização (clientes e/ou fornecedores), dos movimentos em stock dos produtos, da conta corrente dos clientes, das chamadas telefónicas e dos RMA, sendo estes os vetores de análise que serão considerados neste projeto de dissertação. No entanto, é importante referir que, o modelo apresentado pode ”crescer” à medida que novos processos de negócio vão sendo considerados e incluídos (e.g., encomendas, empréstimos, conta corrente de fornecedores, CRM e recursos humanos).

Após uma análise individual realizada a cada uma das tabelas, de forma a compreender os dados e a detetar anomalias/erros, chegou-se à conclusão de que os mesmos apresentam diversas anomalias como: diferentes representações de valor, dados em falta, erros ortográficos e existência de valores errados. Alguns exemplos destas análises encontram-se disponíveis em anexo (Anexo A - Especificação dos atributos da tabela produtos, Anexo B - Caracterização dos erros na tabela produtos). Estas anomalias encontram-se, em grande percentagem, nos dados cuja data é anterior a Março de 2009. Isto deve-se ao facto da organização ter alterado o seu sistema ERP (integração de três bases de dados operacionais).

As políticas de tratamento e correção dos dados foram definidas com a ajuda da organização. Assim junto dos responsáveis achou-se conveniente trabalhar com dados posteriores a 01 de Janeiro de 2010 e estabeleceram-se duas politicas de correção e tratamento: (1) para erros nos dados com possibilidade de correção no sistema ERP e (2) para os restantes que não possam ser corregidos no sistema. Para os primeiros fez-se o levantamento das anomalias/erros de forma a que estes possam ser corregidos no sistema ERP. Os segundos são tratados no processo de ETL (secção 5.2.2).

Nesta fase inicial concluiu-se que más decisões ou uma má interpretação dos dados poderão comprometer o sucesso de todo o projeto. Como tal, deve-se ter em atenção eventuais erros/problemas de interpretação, com o objetivo de evitar constrangimentos futuros e o redesenho da solução.

De momento, e como referido anteriormente, as *views* implementadas são a única fonte de dados. Contudo, em trabalho futuro, outras fontes complementares poderão ser acrescentadas ao modelo tendo em vista o crescimento do *Data Warehouse*, como por exemplo: os sistemas de picagem de presenças dos colaboradores, os sistemas de processamento de *tickets* de suporte técnico com o cliente e/ou colaborador, sistema de formação e repositório de informação (*moodle*), bem como uma outra qualquer área da organização.

4.3 Arquitetura do Sistema de *Business Intelligence*

Nesta secção é apresentada a arquitetura do sistema, assim como as tecnologias utilizadas para a sua implementação e exploração.

Como descrito anteriormente, a base tecnológica da organização assenta essencialmente em tecnologias *Microsoft.* O sistema ERP encontra-se alojado num servidor que possui o sistema operacional *Windows Server* 2003 em conjunto com SQL *Server*

2005. Deste modo, sugere-se um outro servidor para suportar o sistema de *Business Intelligence*. A Figura 4.8 ilustra a decomposição física dos respetivos servidores (servidor de dados e o servidor de *Business Intelligence*).

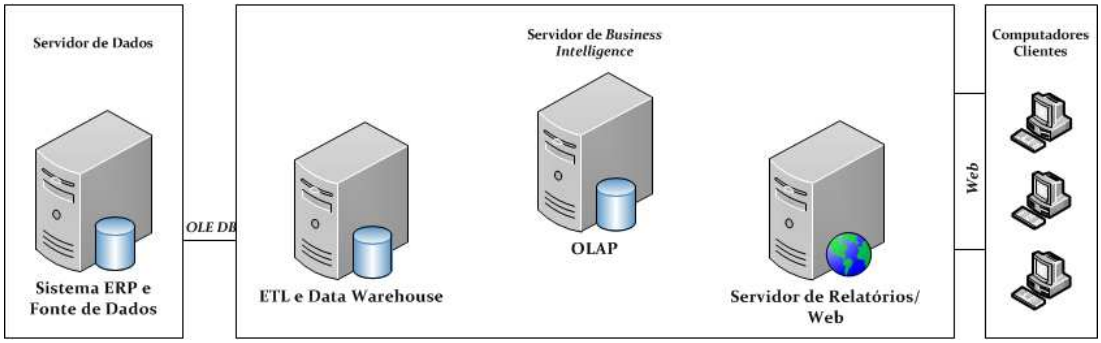


Figura 4.8: Arquitetura física do sistema

Já a Figura 4.9 mostra a arquitetura (tecnologias e componentes) do sistema de *Business Intelligence* a implementar. Esta arquitetura baseia-se numa arquitetura genérica apresentada na secção 2.2.

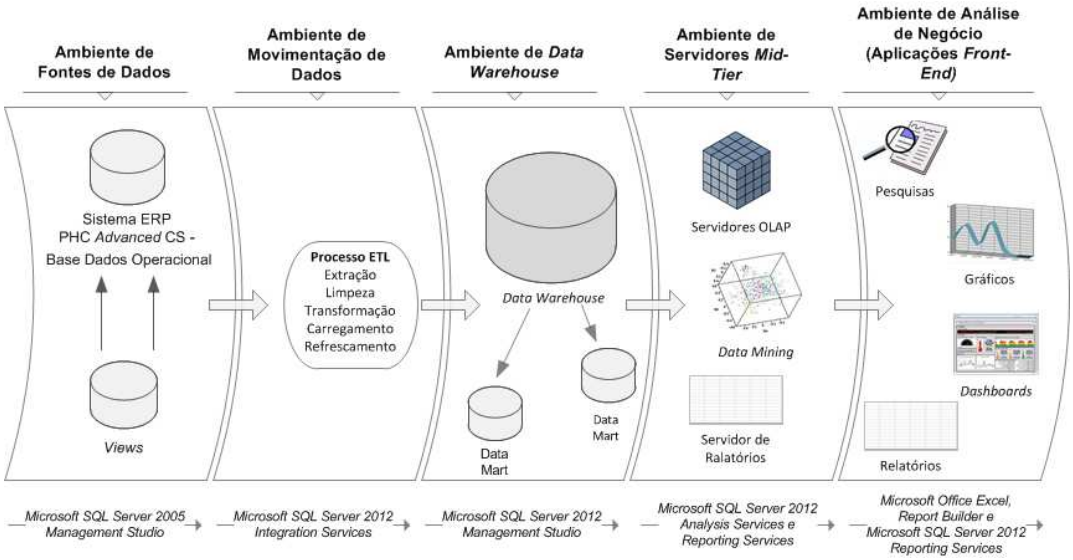


Figura 4.9: Arquitetura do sistema implementado

Atendendo ao que foi referido anteriormente e como é possível verificar na Figura 4.9, a base de dados da organização é originária do seu sistema de ERP e consequentemente foram criadas as *views* para facilitar o acesso aos dados. Seguindo esta lógica, é a partir destas que o processo de ETL será executado, o que permitirá a extração, a transformação e o carregamento dos dados para o respetivo *Data Warehouse*. O *Data Warehouse* serve de suporte ao sistema OLAP, que possibilita a análise dos dados sobre as diversas perspetivas consideradas aquando a modelação

multidimensional dos dados, e às técnicas de *Data Mining* que serão aplicadas para análise dos dados. As aplicações de *front-end* vão permitir uma análise mais intuitiva, mais acessível e fácil da informação, possibilitando a análise e criação de relatórios, bem como o acompanhamento dos KPI considerados.

No geral, esta plataforma deve conceder, de uma forma rápida, automatizada e flexível, toda a informação de gestão necessária, permitindo ao gestor o acompanhamento do negócio sobre diversas perspetivas e diferentes níveis de detalhe.

Capítulo 5

Implementação do Sistema de *Business Intelligence*

Neste capítulo é apresentado o sistema de *Business Intelligence* concebido e implementado na organização. Inicia-se o capítulo com a descrição da conceção e implementação do *Data Warehouse* que vai servir de suporte ao processo de tomada de decisão. Segue-se a descrição individual de cada uma das tabelas de factos e das dimensões que constituem o modelo do *Data Warehouse*. Posteriormente é apresentado o processo de ETL realizado para transformar e integrar os dados no respetivo modelo. Ainda neste capítulo é descrita a forma de como foi realizada a criação e modelação dos cubos de dados que são a base para o sistema OLAP. Após as análises OLAP são apresentadas as técnicas utilizadas para o processo de *Data Mining*, nomeadamente a origem e o processo de integração dos dados, bem como os algoritmos explorados. O capítulo termina com a apresentação das aplicações de *front-end* que permitem o acesso e a exploração da informação.

5.1 Sistema de *Data Warehousing*

Depois de realizada a caracterização dos dados operacionais e das respetivas fontes de dados, a modelação do *Data Warehouse* assim como o processo de ETL são apresentados nesta secção.

5.1.1 Modelo Dados do *Data Warehouse*

A modelação multidimensional é considerada uma das etapas mais importantes e é um dos fatores críticos de sucesso num sistema de *Data Warehousing.* Neste sentido, é importante referir que a modelação do *Data Warehouse* foi definida com base nas

necessidades de informação e no processo de tomada de decisão da organização. O modelo de dados proposto assenta na construção de uma estrutura que suporte os principais indicadores de negócio, partindo dos requisitos e dos objetivos pretendidos para as análises do negócio.

A modelação multidimensional assenta num esquema em Constelação e é representado pela Figura 5.1. Neste modelo é possível verificar todas as entidades (factos

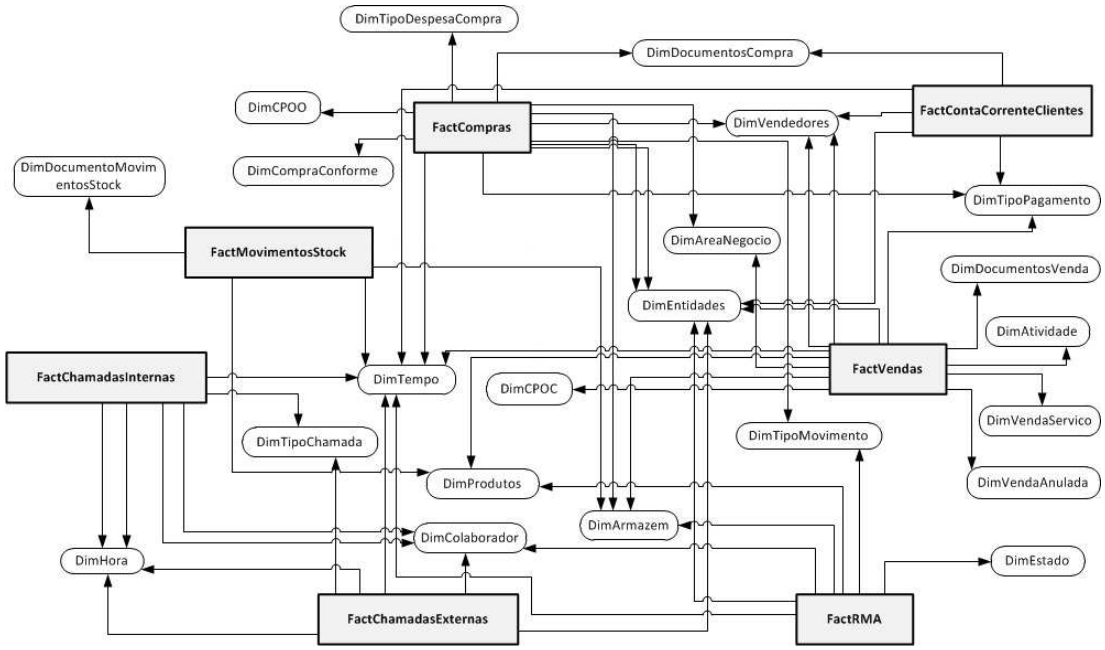


Figura 5.1: Modelo de dados do *Data Warehouse*

e dimensões) presentes no *Data Warehouse* e os relacionamentos entre elas.

Sendo o modelo de dados do *Data Warehouse* baseado num esquema em Constelação, que por definição é um conjunto de esquemas em estrela, ou seja, composto por múltiplas tabelas de factos, unidas por diversas dimensões, procura-se detalhar o modelo apresentando as estrelas que o compõe (inicialmente as tabelas de facto e posteriormente as tabelas de dimensão).

O modelo de *Data Warehouse* integra 7 (sete) tabelas de facto e 22 (vinte e duas) tabelas de dimensão.

As tabelas de factos são representadas por: **FactVendas**, **FactCompras**, **FactContaCorrenteClientes**, **FactMovimentosStock**, **FactRMA**, **FactChamadasInternas** e **FactChamadasExternas**.

A tabela de factos **FactVendas** permite o armazenamento da informação sobre as vendas efetuadas pela organização. Esta tabela está relacionada com diversas dimensões (DimEntidades, DimProdutos, DimVendedores, DimTempo, DimTipoPagamento, DimAreaNegocio, DimAmazem, DimVendaServico, DimVendaAnulada,

DimCPOC, DimAtividade e DimDocumentosVenda), o que permite analisar as vendas sobre diversas perspetivas (Figura 5.2).

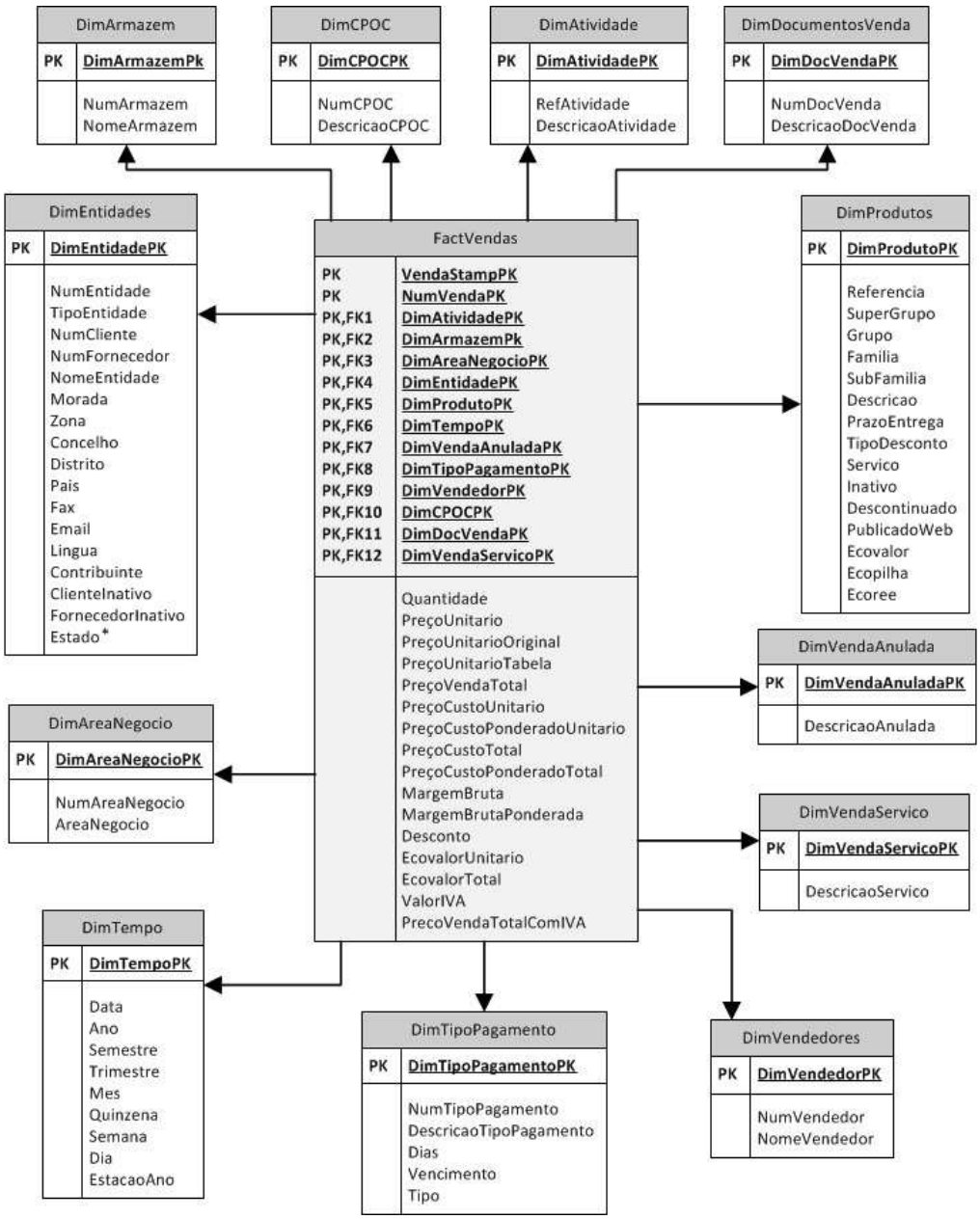


Figura 5.2: Modelo de dados em Estrela da FactVendas

A Tabela 5.1 detalha os diversos atributos que integram a FactVendas. Destacam-

Tabela 5.1: Tabela de factos FactVendas do *Data Warehouse*

FactVendas				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
VendaStampPK	Char(25)	PK	-	Identificador único de uma linha
NumVendaPK	numeric(10,0)	PK	-	Identificador de uma transação
DimEntidadePK	int	PK e FK	-	ID da DimEntidade
DimTempoPK	int	PK e FK	-	ID da DimTempo
DimDocVendaPK	int	PK e FK	-	ID da DimDocumentosVenda
DimVendedorPK	int	PK e FK	-	ID da DimVendedores
DimVendaAnuladaPK	int	PK e FK	-	ID da DimVendaAnulada
DimVendaServicoPK	int	PK e FK	-	ID da DimVendaServico
DimTipoPagamentoPK	int	PK e FK	-	ID da DimTipoPagamento
DimArmazemPK	int	PK e FK	-	ID da DimArmazem
DimProdutoPK	int	PK e FK	-	ID da DimProdutos
DimCPOCPK	int	PK e FK	-	ID da DIM CPOC
DimAtividadePK	int	PK e FK	-	ID da DimCPOC
DimAreaNegocioPK	int	PK e FK	-	ID da DimAreaNegocio
Quantidade	int	N/A	-	Quantidade vendida
PrecoUnitario	numeric(19,2)	N/A	-	Preço unitário do produto
PrecoUnitarioOriginal	numeric(19,2)	N/A	-	Preço original do produto
PreçoUnitarioTabela	numeric(19,2)	N/A	-	Preço em tabela do produto
PrecoVendaTotal	numeric(19,2)	N/A	-	Preco de venda
PrecoCustoUnitario	numeric(19,2)	N/A	-	Preço de custo unitário do produto
PreçoCustoTotal	numeric(19,2)	N/A	-	Preco de custo
PrecoCustoPonderadoUnitario	numeric(19,2)	N/A	-	Preço de custo unitário com serviços
PrecoCustoPonderadoTotal	numeric(19,2)	N/A	-	Preço de custo total com serviços
MargemBruta	numeric(19,2)	N/A	-	PrecoVendaTotal - PrecoCustoTotal
MargemBrutaPonderada	numeric(19,2)	N/A	-	PreçoVendaTotal - PrecoCustoPonderadoTotal
Desconto	numeric(19,2)	N/A	-	Desconto da venda
EcovalorUnitario	numeric(19,2)	N/A	-	Ecovalor unitário do produto
EcovalorTotal	numeric(19,2)	N/A	-	Ecovalor total do produto
ValorIVA	numeric(19,2)	N/A	-	Valor de IVA (Taxa * PreçoVendaTotal)
PreçoVendaTotalComIVA	numeric(19,2)	N/A	-	Preço de venda com IVA (PreçoVendaTotal+ValorIVA)

se as chaves estrangeiras (FK) que ligam as diversas dimensões, bem como as duas chaves primárias (PK) (VendaStampPK e NumVendaPK) que permitem identificar uma venda em particular. Para além das chaves mencionadas, a tabela de factos contém diversos factos que permitem avaliar a performance das vendas. É importante referir que na concretização do sistema OLAP várias medidas podem ser calculadas tendo por base os factos apresentados (e.g., Margem de Lucro Bruto).

A FactVendas inclui todos os registos das vendas, independente da sua origem e

do documento associado (e.g., faturas, notas de crédito, notas de débito, vendas a dinheiro, guias de remessa, fatura pro-forma, etc.). Essa distinção é realizada ao nível das diversas dimensões (e.g., DimDocumentosVenda, DimCPOC, etc.). Contudo, na implementação do sistema OLAP opta-se por criar uma vista da FactVendas incluindo apenas os documentos de faturação.

A tabela de factos **FactCompras** permite o armazenamento de toda a informação associada às compras (de produtos e serviços) efetuadas pela organização. Esta tabela relaciona-se com diversas dimensões (DimEntidades, DimProdutos, DimTempo, DimAreaNegocio, DimTipoDespesaCompra, DimArmazem, DimCPOO, DimDocumentosCompra, DimVendedor, DimTipoMovimentos e DimCompraConforme), o que permite analisar as compras sobre diversas perspetivas (Figura 5.3).

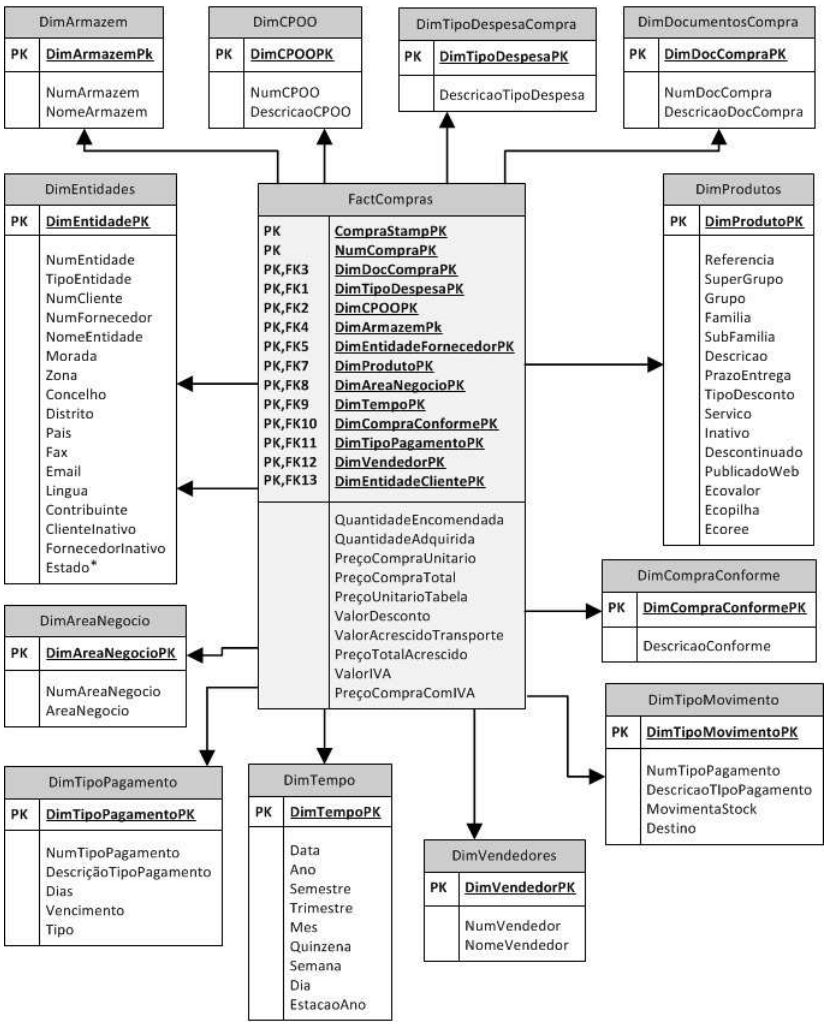


Figura 5.3: Modelo de dados em Estrela da FactCompras

A Tabela 5.2 detalha os atributos que integram a FactCompras.

Tabela 5.2: Tabela de factos FactCompras do *Data Warehouse*

FactCompras				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
CompraStampPK	Char(25)	PK	-	Identificador único de uma linha
NumCompraPK	numeric(10,0)	PK	-	Identificador de uma transação
DimEntidadeFornecedorPK	int	PK e FK	-	ID da DimEntidades
DimTempoPK	int	PK e FK	-	ID da DimTempo
DimDocCompraPK	int	PK e FK	-	ID da DimDocCompra
DimVendedorPK	int	PK e FK	-	ID da DDimVendedor
DimEntidadeClientePK	int	PK e FK	-	ID da DimEntidade
DimTipoMovimentoPK	int	PK e FK	-	ID da DimTipoMovimento
DimTipoPagamentoPK	int	PK e FK	-	ID da DimTipoPagamento
DimArmazemPK	int	PK e FK	-	ID da DimArmazem
DimProdutoPK	int	PK e FK	-	ID da DimProdutos
DimCPOOPK	int	PK e FK	-	ID da DimCPOO
DimTipoDespesaPK	int	PK e FK	-	ID da DimTipoDespesaCompra
DimAreaNegocioPK	int	PK e FK	-	ID da DimAreaNegocio
QuantidadeEncomendada	int	N/A	-	Quantidade encomendada
QuantidadeAdquirida	int	N/A	-	Quantidade comprada
PrecoCompraUnitario	numeric(19,2)	N/A	-	Preço unitário
PrecoCompra	numeric(19,2)	N/A	-	Preço de compra
PrecoUnitarioTabela	numeric(19,2)	N/A	-	Preço em tabela do produto
ValorDesconto	numeric(19,2)	N/A	-	Desconto da compra
ValorAcrescidoTransporte	numeric(19,2)	N/A	-	Custo dos serviços/transporte
PreçoTotalAcrescido	numeric(19,2)	N/A	-	Preço total com serviços/transporte
ValorIVA	numeric(19,2)	N/A	-	Valor de IVA (Taxa de IVA*PrecoCompra)
PrecoCompraTotalComIVA	numeric(19,2)	N/A	-	Preço de compra com IVA (PreçoCompra+ValorIVA)

A FactCompras contém todos o registos das compras (gastos) desde a aquisição de equipamento a fornecedores às despesas internas (e.g., luz e água). Destacam-se as chaves estrangeiras (FK) que ligam as diversas dimensões, bem como as duas chaves primárias (PK) (CompraStampPK e NumCompraPK) que permitem identificar uma compra em particular.

A tabela de factos **FactContaCorrenteClientes** armazena a informação relacionada com a conta corrente dos clientes. A par das anteriores, esta tabela de factos também se encontra relacionada com diversas tabelas de dimensão (DimEntidades, DimTempo, DimTipoPagamento, DimVendedores, DimDocumentosCompra e DimTempo), o que permite a análise da conta corrente sobre diversas perspetivas (Figura 5.4).

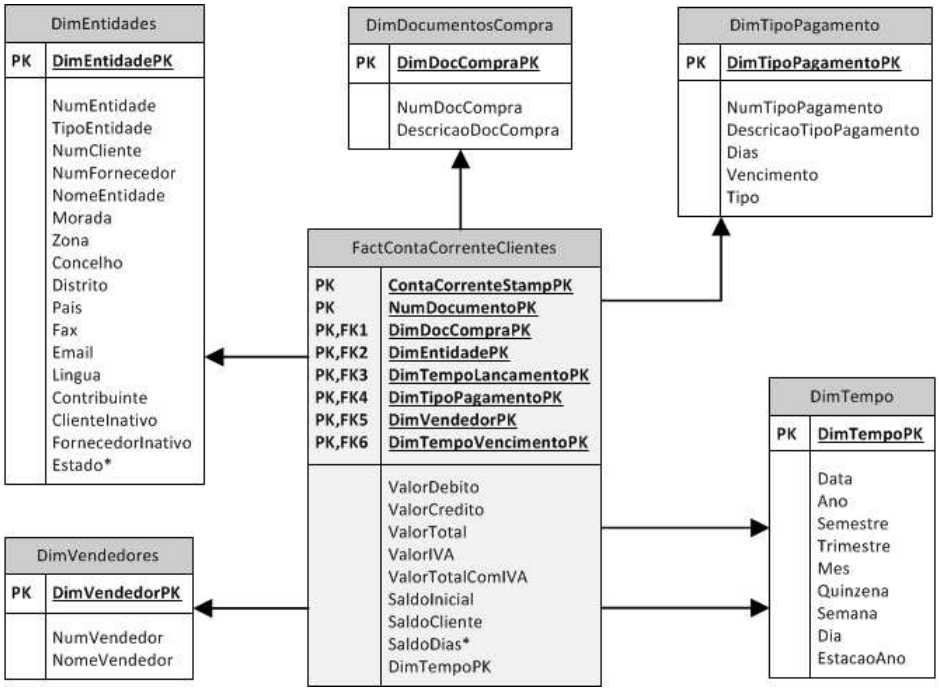


Figura 5.4: Modelo de dados em Estrela da FactContaCorrenteClientes

O principal KPI de análise está relacionado com o saldo do cliente, em particular o saldo médio do cliente. O que vai permitir, posteriormente, estudar a rentabilidade e o crédito médio concedido ao cliente, num determinado período de tempo, assim como o custo de financiamento do ativo económico.

A Tabela 5.3 detalha os diversos atributos que integram FactContaCorrenteClientes.

Tabela 5.3: Tabela de factos FactContaCorrenteClientes do *Data Warehouse*

FactContaCorrenteClientes				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
ContaCorrenteStampPK	Char(25)	PK	-	Identificador único de uma linha
NumDocumentoPK	numeric(10,0)	PK	-	Identificador de um documento
DimEntidadePK	int	PK e FK	-	ID da DimEntidade
DimTempoLancamentoPK	int	PK e FK	-	ID da DimTempo
DimTempoVencimentoPK	int	PK e FK	-	ID da DimTempo
DimDocCompraPK	int	PK e FK	-	ID da DimDocumentosCompra
DimVendedorPK	int	PK e FK	-	ID da DimVendedor
DimTipoPagamentoPK	int	PK e FK	-	ID da DimTipoPagamento
ValorDebito	numeric(19,2)	N/A	-	Valor do débito do cliente
ValorCredito	numeric(19,2)	N/A	-	Valor do crédito do cliente
ValorTotal	numeric(19,2)	N/A	-	Diferença entre o Crédito e o Débito
ValorIVA	numeric(19,2)	N/A	-	Valor de IVA a pagar
ValorTotalComIVA	numeric(19,2)	N/A	-	ValorTotal+ValorIVA
SaldoInicial	numeric(19,2)	N/A	-	Saldo inicial do cliente
SaldoCliente	numeric(19,2)	N/A	-	Saldo do cliente
SaldoDias*	numeric(19,2)	N/A	-	SaldoCliente * NºDias

O atributo que está assinalado com * (SaldoDias) é considerado apenas para efeitos de cálculo futuro na construção do sistema OLAP e dos KPI relacionados com a conta corrente dos clientes.

A tabela de factos **FactMovimentosStock** armazena a informação relativa aos movimentos de stock, nomeadamente das entradas e saídas (em *stock*) dos produtos. Esta tabela encontra-se relacionada com diversas dimensões (DimCPOC, DimDocumentosMovimentosStock, DimArmazem, DimProdutos e DimTempo), o que permite analisar os movimentos sobre diversas perspetivas (Figura 5.5).

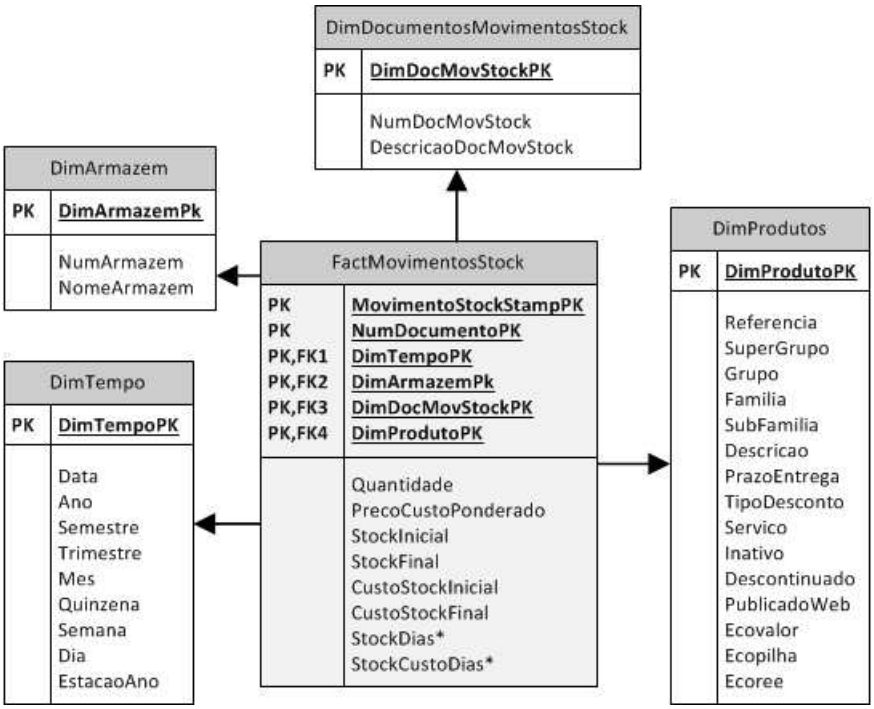


Figura 5.5: Modelo de dados em Estrela da FactMovimentosStock

Um dos principais KPI reside no apuramento do custo do *stock* médio de uma determinada área de negócio (produtos), bem como o *stock* médio a si associado, num determinado período de tempo.

A Tabela 5.4 detalha os diversos atributos que integram a FactMovimentosStock.

Tabela 5.4: Tabela de factos FactMovimentosStock do *Data Warehouse*

FactMovimentosStock				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
MovimentoStockStampPK	Char(25)	PK	-	Identificador único de uma linha
NumDocumentoPK	numerica(10,0)	PK	-	Identificador de um documento
DimTempoPK	int	PK e FK	-	ID da DimTempo
DimDocMovStockPK	int	PK e FK	-	ID da DimDocumentosMovimentosStock
DimArmazemPK	int	PK e FK	-	ID da DimArmazem
DimProdutoPK	int	PK e FK	-	ID da DimProduto
Quantidade	int	N/A	-	Quantidade movimentada
PrecoCustoPonderado	numeric(19,2)	N/A	-	Custo do produto com serviços
StockInicial	numeric(19,2)	N/A	-	Stock inicial do produto
StockFinal	numeric(19,2)	N/A	-	Stock final depois do movimento
CustoStockInicial	numeric(19,2)	N/A	-	Custo do stock inicial do produto
CustoStockFinal	numeric(19,2)	N/A	-	Custo do stock final depois do movimento
StockDias*	numeric(19,2)	N/A	-	StockFinal * NºDias
StockCustoDias*	numeric(19,2)	N/A	-	StockCustoFinal * NºDias

Os atributos que se encontram assinalados com * (StockCustoDias e StockDias) são considerados apenas para efeitos de cálculo futuro aquando da criação do sistema OLAP e dos KPI relacionados com os movimentos em *stock.*

As tabelas de factos supra-mencionadas (**FactVendas**, **FactCompras**, **FactContaCorrenteClientes**, **FactMovimentosStock**) foram consideradas como prioritárias para a concretização dos objetivos inerentes às análises de negócio, bem como para suportar os principais KPI identificados. Contudo, modelaram-se mais três tabelas de factos (**FactRMA**, **FactChamadasInternas** e **FactChamadasExternas**).

A tabela de factos **FactRMA** armazena toda a informação dos RMA da organização. O retorno de material avariado acontece quando um cliente devolve um determinado produto. Com isto é possível identificar quais os produtos que dão mais problemas, quais os clientes que efetuam mais RMA e os custos a si associados. Esta tabela encontra-se relacionada com diversas dimensões, o que permite analisar os RMA sobre diversas perspetivas (DimEntidades, DimProdutos, DimTempo, DimColaborador, DimEstado, DimArmazem, DimTipoMovimento) (Figura 5.6).

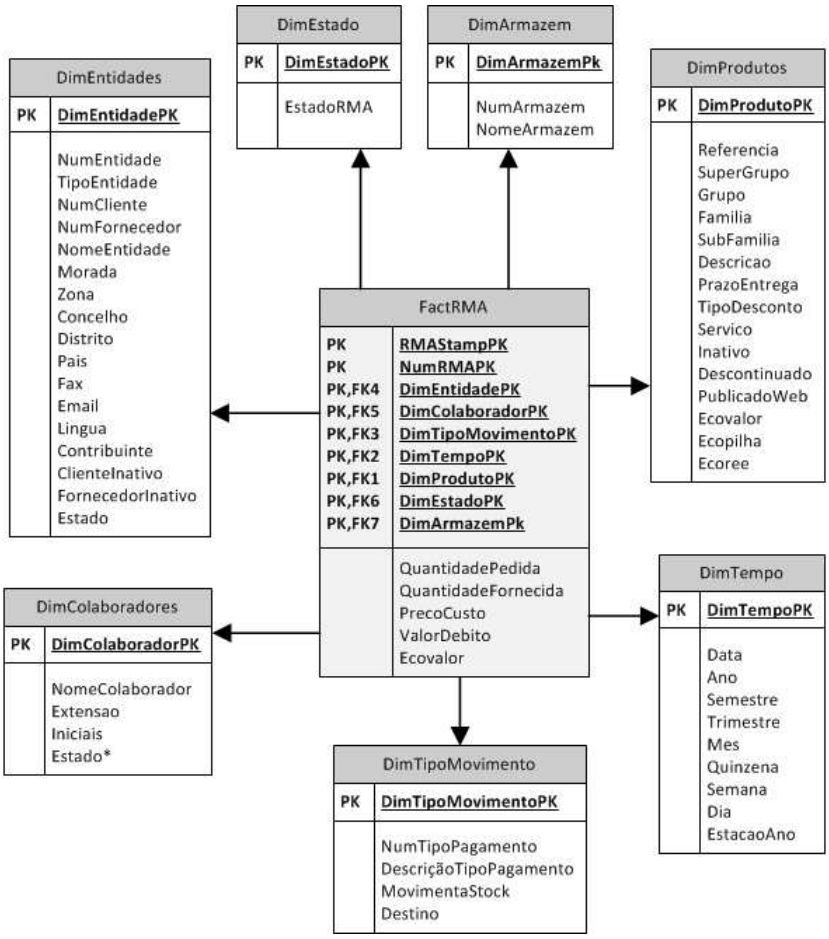


Figura 5.6: Modelo de dados em Estrela da FactRMA

A Tabela 5.5 detalha os diversos atributos que integram a FactRMA.

Tabela 5.5: Tabela de factos FactRMA do *Data Warehouse*

FactRMA				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
RMAStampPK	Char(25)	PK	-	Identificador único de uma linha
NumRMAPK	varchar(20)	PK	-	Identificador de um documento de RMA
DimTempoPK	int	PK e FK	-	ID da DimTempo
DimTipoMovimentoPK	int	PK e FK	-	ID da DimTipoMovimento
DimArmazemPK	int	PK e FK	-	ID da DimArmazem
DimEstadoPK	int	PK e FK	-	ID da DimEstado
DimEntidadePK	int	PK e FK	-	ID da DimEntidades
DimProdutoPK	int	PK e FK	-	ID da DimProdutos
DimColaboradorPK	int	PK e FK	-	ID da DimColaboradores
QuantidadePedida	int	N/A	-	Quantidade de RMA
QuantidadeFornecida	int	N/A	-	Quantidade de produtos recebidos
PrecoCusto	numeric(19,2)	N/A	-	Custo do RMA
ValorDebito	numeric(19,2)	N/A	-	Valor de débito do cliente
Ecovalor	numeric(19,2)	N/A	-	Ecovalor do produto

A tabela de factos **FactChamadasInternas** armazena toda a informação relativa às chamadas telefónicas (tempos e quantidades) que são efetuadas internamente entre os colaboradores da organização, enquanto que a tabela de factos **FactChamadasExternas** regista as chamadas (tempos e quantidades) estabelecidas com o exterior entre os colaboradores da organização e as entidades externas (clientes, fornecedor ou contactos). Estas tabelas encontram-se associadas a diversas dimensões, o que permite analisar as chamadas sobre várias perspetivas (Figura 5.7).

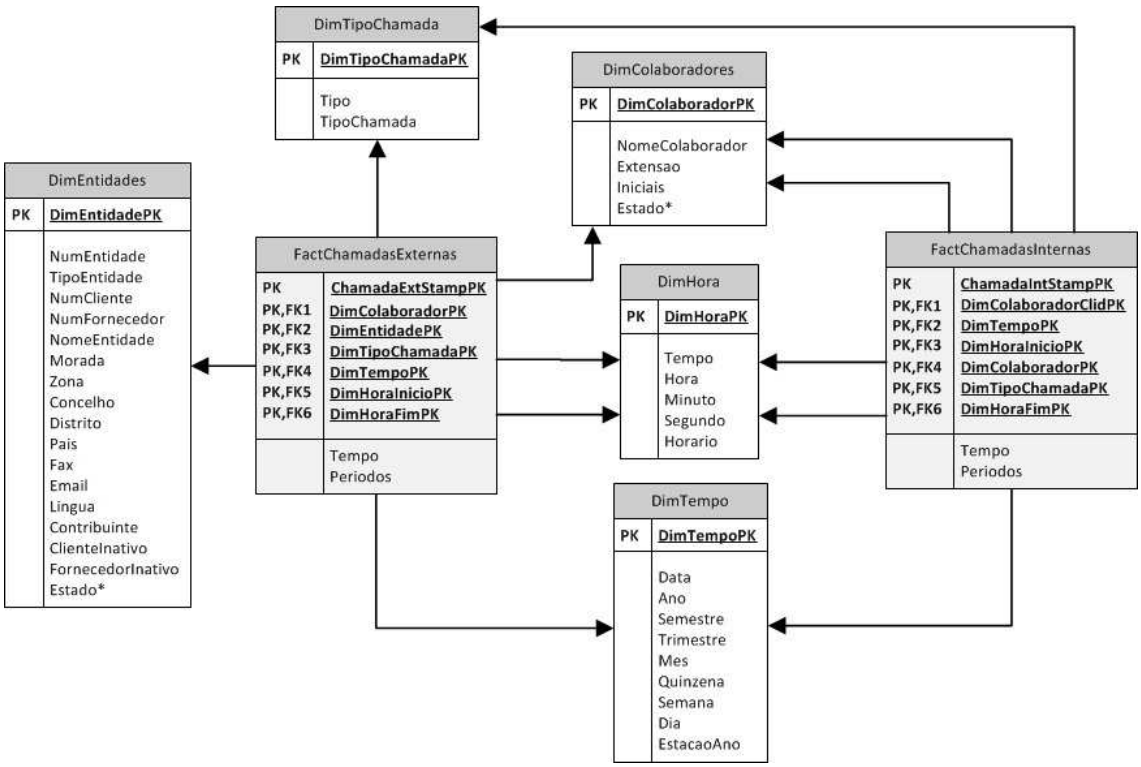


Figura 5.7: Modelo de dados em Constelação das Chamadas

A construção deste modelo baseou-se nas necessidades informacionais para as chamadas telefónicas registadas. Compreender e identificar quais os colaboradores mais ativos, analisar quais as entidades que normalmente estão em contacto com a organização, e acima de tudo, a possibilidade de cruzar a informação das chamadas com KPI financeiros. O possível cruzamento desta informação (e.g, chamadas efetuadas por clientes com informação das vendas num determinado período de tempo) obtendo indicadores que caracterizem os clientes, é um fator de análise importante para a organização. É então possível responder a questões como: "Quais os clientes que recebem mais assistência e quais destes os mais rentáveis?".

O sistema OLTP que faz o controlo das chamadas telefónicas na organização têm vindo, ao longo do tempo, a sofrer algumas alterações. No entanto este sistema carece ainda de informação que permita caracterizar os custos, as chamadas não

atendidas, entre outros indicadores.

As Tabelas 5.6 e 5.7 detalham os diversos atributos que integram as FactChamadasInternas e a FactChamadasExternas respetivamente.

Tabela 5.6: Tabela de factos FactChamadasInternas do *Data Warehouse*

FactChamadasInternas				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
ChamadaIntStampPK	Char(25)	PK	-	Identificador único de uma linha
DimColaboradorClidPK	int	PK e FK	-	ID da DimColaboradores
DimTempoPK	int	PK e FK	-	ID da DimTempo
DimHoraFimPK	int	PK e FK	-	ID da DimHora
DimHoraInicioPK	int	PK e FK	-	ID da DimHora
DimTipoChamadaPK	int	PK e FK	-	ID da DimTipoChamada
DimColaboradorPK	int	PK e FK	-	ID da DimColaboradores
Tempo	int	N/A	-	Duração da chamada em segundos
Periodos	int	N/A	-	Periodos de tempo uma chamada

Tabela 5.7: Tabela de factos FactChamadasExternas do *Data Warehouse*

FactChamadasExternas				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
ChamadaExtStampPK	Char(25)	PK	-	Identificador único de uma linha
DimTempoPK	int	PK e FK	-	ID da DimTempo
DimHoraFimPK	int	PK e FK	-	ID da DimHora
DimHoraInicioPK	int	PK e FK	-	ID da DimHora
DimTipoChamadaPK	int	PK e FK	-	DimTipoChamada
DimColaboradorPK	int	PK e FK	-	ID da DimColaborador
DimEntidadePK	int	PK e FK	-	ID da DimEntidades
Tempo	int	N/A	-	Duração da chamada em segundos
Periodos	int	N/A	-	Periodos de tempo uma chamada

No geral, as tabelas de factos apresentadas são constituídas por métricas mensuráveis (factos) e por chaves (FK) para cada uma das dimensões com as quais se encontram relacionadas. No entanto, as tabelas de factos incluem, para além das chaves da dimensão, uma chave primária (PK) que identifica unicamente um registo, garantindo desta forma a integridade da entidade, e uma chave degenerada (PK) que identifica o número do documento (e.g., fatura, guia de remessa). A decisão de utilizar a chave degenerada é baseada no desejo de fornecer uma referência direta sistema OLTP.

O modelo do *Data Warehouse*, para além das 7 (sete) tabelas de facto apresentadas, é composto por 22 (vinte e duas) tabelas de dimensão, representadas por:

- DimAreaNegocio - tabela que guarda informação sobre as diversas áreas de negócio da organização (Tabela 5.8);

Tabela 5.8: Tabela de dimensão DimAreaNegocio do *Data Warehouse*

DimAreaNegocio				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimAreaNegocioPK	int	PK	-	Chave substituta
NumAreaNegocio	varchar(20)	-	1	Número da área de negócio
AreaNegocio	varchar(81)	-	1	Descrição da área de negócio

- DimArmazem - tabela que guarda a informação sobre os armazéns (nome e número de armazém) onde os produtos transacionados dão entrada ou saída (Tabela 5.9);

Tabela 5.9: Tabela de dimensão DimArmazem do *Data Warehouse*

DimArmazem				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimAmazemPK	int	PK	-	Chave substituta
NumAmazem	numeric(5,0)	-	1	Chave original
NomeArmazem	varchar(81)	-	1	Descrição do armazém

- DimCompraConforme - esta tabela guarda a informação relativa à conformidade das compras do produtos, ou seja, se o produto adquirido está em conformidade para com os requisitos da organização (Tabela 5.10);

Tabela 5.10: Tabela de dimensão DimCompraConforme do *Data Warehouse*

DimCompraConforme				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimCompraConformePK	int	PK	-	Chave substituta
DescricaoConforme	varchar(3)	-	-	Descrição da compra conforme

- DimAtividade - esta tabela armazena informação sobre a descrição das atividades ao qual uma venda se encontra associada (e.g., comercial, estrutura) (Tabela 5.11);

Tabela 5.11: Tabela de dimensão DimAtividade do *Data Warehouse*

DimAtividade				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimAtividadePK	int	PK	-	Chave substituta
RefAtividade	varchar(20)	-	1	Chave original
DescricaoAtividade	varchar(70)	-	1	Descrição da atividade

- DimColaboradores - esta tabela guarda a informação sobre os colaboradores da organização, nomeadamente o seu nome, iniciais e extensão telefónica (Tabela 5.12).

Tabela 5.12: Tabela de dimensão DimColaboradores do *Data Warehouse*

DimColaboradores				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimColaboradorPK	int	PK	-	Chave substituta
NomeColaborador	varchar(30)	-	1	Chave original
Extensao	varchar(25)	-	2	Extensão interna do telefone
Iniciais	varchar(3)	-	1	Iniciais do nome
Estado*	varchar(25)	-	2	Descrição do estado do colaborador

O SCD do atributo "Extensão" é do tipo 2, já o atributo "Estado" permite fazer o controlo das atualizações. Este reflete o estado do colaborador (*current* (currente) ou *expired* (expirado)) caso ocorram atualizações no atributo "Extensão".

- DimCompraConforme - esta tabela guarda a informação relativa à conformidade das compras dos produtos (Tabela 5.13);

Tabela 5.13: Tabela de dimensão DimCompraConforme do *Data Warehouse*

DimCompraConforme				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimCompraConformePK	int	PK	-	Chave substituta
DescricaoConforme	varchar(3)	-	-	Descrição da compra conforme

- DimCPOC - esta tabela armazena informação sobre o plano oficial de contabilidade relativo às vendas (e.g., venda de mercadorias) (Tabela 5.14);

Tabela 5.14: Tabela de dimensão DimCPOC do *Data Warehouse*

Tabela de dimensão DimCPOC				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimCPOCPK	int	PK	-	Chave substituta
NumCPOC	numeric(6,0)	-	1	Chave original
DescricaoCPOC	varchar(50)	-	1	Designação da CPOC

- DimCPOO - esta tabela, assim como a anterior, armazena informação sobre o plano oficial de contabilidade relativo às compras (e.g., despesas de transporte) (Tabela 5.15);

Tabela 5.15: Tabela de dimensão DimCPOO do *Data Warehouse*

DimCPOO				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimCPOOPK	int	PK	-	Chave substituta
NumCPOO	numeric(6,0)	-	1	Chave original
DescricaoCPOO	varchar(50)	-	1	Designação da CPOO

- DimDocumentosVenda - esta tabela guarda a informação sobre os documentos de venda (e.g., fatura, nota de crédito) (Tabela 5.16);

Tabela 5.16: Tabela de dimensão DimDocumentosVenda do *Data Warehouse*

DimDocumentosVenda				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimDocVendaPK	int	PK	-	Chave substituta
NumDocVenda	numeric(3,0)	-	1	Chave original
DescricaoDocVenda	varchar(20)	-	1	Designação do documento de venda

- DimDocumentosCompra - esta tabela guarda a informação sobre os documentos de compra (e.g., guia de entrada) (Tabela 5.17);

Tabela 5.17: Tabela de dimensão DimDocumentosCompra do *Data Warehouse*

DimDocumentosCompra				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimDocCompraPK	int	PK	-	Chave substituta
NumDocCompra	numeric(3,0)	-	1	Chave original
DescricaoDocCompra	varchar(20)	-	1	Designação do documento de compra

- DimDocumentosMovimentosStock - esta tabela guarda informação sobre os documentos relacionados com os movimentos de produtos (Tabela 5.18);

Tabela 5.18: Tabela de dimensão DimDocumentosMovimentosStock do *Data Warehouse*

DimDocumentosMovimentosStock				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimDocMovStockPK	int	PK	-	Chave substituta
NumDocMovStock	numeric(5,0)	-	1	Chave original
DescriçãoDocMovStock	varchar(20)	-	1	Designação do movimento de stock

- DimEstado - esta tabela guarda informação relativa ao estado em que se encontra o RMA (e.g., entregue ao fornecedor, devolvido ao cliente) (Tabela 5.19);

Tabela 5.19: Tabela de dimensão DimEstado do *Data Warehouse*

DimEstado				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimEstadoPK	int	PK	-	Chave substituta
EstadoRMA	varchar(40)	-	1	Descrição do estado do RMA

- DimTipoChamada - esta tabela guarda a informação sobre o tipo de chamada (e.g., interna, externa, efetuada, recebida, etc.) (Tabela 5.20);

Tabela 5.20: Tabela de dimensão DimTipoChamada do *Data Warehouse*

DimTipoChamada				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimTipoChamadaPK	int	PK	-	Chave substituta
Tipo	varchar(25)	-	1	Chave original
TipoChamada	varchar(25)	-	1	Descrição do tipo de chamada

- DimEntidades - esta tabela guarda informação sobre os clientes, fornecedores e contactos da organização (Tabela 5.21). Estes são diferenciados através do atributo "Tipo de Entidade". Optou-se por integrar nesta a informação geográfica da entidade (e.g., morada, zona, concelho, distrito e país). O SCD do atributo "Morada", "Zona", "ClienteIantivo" e "FornecedorInativo" são do tipo 2, quer isto dizer que caso ocorra uma atualização nos registos destes atributos o seu histórico é guardado, sendo que o "Estado" do registo atualizado

passa a ser "*Expired*" e é inserido um novo registo na DimEntidades atualizado e onde o "Estado" passa a ser "*Current",*

Tabela 5.21: Tabela de dimensão DimEntidades do *Data Warehouse*

DimEntidades				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimEntidadePK	int	PK	-	Chave substituta
NumEntidade	numeric(10,0)	-	1	Chave original da entidade
NumCliente	numeric(10,0)	-	1	Número de cliente
NumFornecedor	numeric(10,0)	-	1	Número de fornecedor
TipoEntidade	varchar(25)	-	1	Descrição do tipo de entidade
NomeEntidade	varchar(55)	-	1	Nome da entidade
Morada	varchar(55)	-	2	Morada da entidade
Zona	varchar(30)	-	2	Descrição da zona geográfica
Concelho	varchar(30)	-	1	Concelho da entidade
Distrito	varchar(30)	-	1	Distrito da entidade
Pais	varchar(20)	-	1	Distrito da entidade
Lingua	char(20)	-	1	Designação do idioma da entidade
Fax	varchar(60)	-	1	Número do Fax da entidade
Email	varchar(60)	-	1	Email da entidade
Contribuinte	varchar(20)	-	1	Número de contribuinte da entidade
ClienteInativo	varchar(10)	-	2	Descreve se entidade cliente está ativo ou não
FornecedorInativo	varchar(10)	-	2	Descreve se entidade fornecedor está ativo ou não
Estado*	varchar(25)	-	2	Descrição do estado da entidade

- DimHora - esta tabela armazena informação relativa a tempos em horas, minutos e segundos das chamadas telefónicas (Tabela 5.22);

Tabela 5.22: Tabela de dimensão DimHora do *Data Warehouse*

DimHora				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimHoraPK	int	PK	-	Chave substituta
Tempo	time(7)	-	-	Descrição do tempo(HH:mm:ss)
Hora	tinyint	-	-	Descrição da hora(HH)
Minuto	tinyint	-	-	Descrição do minuto(mm)
Segundo	tinyint	-	-	Descrição do segundo(ss)
Horário	varchar(10)	-	-	Descrição da parte do dia

A DimHora contém todas as horas, minutos e segundos possíveis. A informação da dimensão é gerada a partir de um *script* (implementado em SQL). O *script* encontra-se disponível em anexo (Anexo C - Criação da tabela DimHora).

- DimProdutos - esta tabela armazena a informação dos produtos e as respetivas áreas de negócio aos quais de encontram associados (e.g., super grupo, grupo, família, sub-família, nome, referência, etc.) (Tabela 5.23);

Tabela 5.23: Tabela de dimensão DimProdutos do *Data Warehouse*

DimProdutos				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimProdutoPK	int	PK	-	Chave substituta
Referencia	varchar(18)	-	1	Referência do produto
SuperGrupo	varchar(81)	-	1	Descrição do super grupo do produto
Grupo	varchar(81)	-	1	Descrição do grupo do produto
Familia	varchar(81)	-	1	Descrição da familia do produto
SubFamilia	varchar(81)	-	1	Descrição da subfamilia do produto
Designacao	varchar(60)	-	1	Designação/nome do produto
PraozEntrega	numeric(3,0)	-	1	Prazo de entrega para o produto
TipoDesconto	varchar(20)	-	1	Tipo de desconto do produto
Servico	bit	-	1	Se o produto é um serviço
Inativo	bit	-	1	Se o produto está inativo
Descontinuado	bit	-	1	Se o produto está descontinuado
PublicadoWeb	bit	-	1	Se o produto está publicado em serviços web
Ecovalor	bit	-	1	Se o produto tem ecovalor
Ecopilha	bit	-	1	Se o produto tem ecopilha
Ecorre	bit	-	1	Se o produto tem ecorre

- DimTipoMovimento - esta tabela armazena a informação dos tipo de movimentos dos RMA e das compras (e.g., encomenda a fornecedores) (Tabela 5.24);

Tabela 5.24: Tabela de dimensão DimTipoMovimento do *Data Warehouse*

DimTipoMovimento				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimTipoMovimentoPK	int	PK	-	Chave substituta
NumTipoMovimento	numeric(3,0)	-	1	Chave original
DescriçãoTipoMovimento	varchar(24)	-	1	Descrição do tipo de movimento
MovimentaStock	bit	-	1	Se movimenta stock
Destino	varchar(2)	-	1	Descrição do destino do movimento

- DimVendaAnulada - esta tabela representa uma *flag* ("Sim" ou "Não") e que permite identificar se uma venda registada acabou por ser anulada (Tabela 5.25);

Tabela 5.25: Tabela de dimensão DimVendaAnulada do *Data Warehouse*

DimVendaAnulada				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimVendaAnuladaPK	int	PK	-	Chave substituta
DescricaoAnulada	varchar(3)	-	1	Se a venda foi anulada

- DimTipoDespesaCompra - esta tabela armazena informação sobre o tipo de despesa a que uma compra está relacionada (Tabela 5.26);

Tabela 5.26: Tabela de dimensão DimTipoDespesaCompra do *Data Warehouse*

DimTipoDespesaCompra				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimTipoDespesaPK	int	PK	-	Chave substituta
DescriçãoTipoDespesa	varchar(25)	-	1	Descrição da despesa

- DimTipoPagamento - esta tabela guarda a informação sobre os tipos de pagamentos (e.g., descrição, dias, vencimento, etc.). Um tipo de pagamento está associado a um determinado cliente (Tabela 5.27);

Tabela 5.27: Tabela de dimensão DimTipoPagamento do *Data Warehouse*

DimTipoPagamento				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimTipoPagamentoPK	int	PK	-	Chave substituta
NumTipoPagamento	char(25)	-	1	Chave original
DescriçãoTipoPagamento	varchar(55)	-	1	Descrição do tipo de pagamento
Dias	numeric(3,0)	-	1	Número de dias de pagamento
Vencimento	numeric(1,0)	-	1	Tipo de vencimento
Tipo	numeric(1,0)	-	1	Tipo de pagamento

- DimVendaServiço - esta tabela armazena informação sobre as vendas, nomeadamente se esta foi ou não uma prestação de serviço (Tabela 5.28);

Tabela 5.28: Tabela de dimensão DimVendaServiço do *Data Warehouse*

DimVendedor				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimVendaServicoPK	int	PK	-	Chave substituta
DescricaoServico	varchar(3)	-	-	Se a venda foi um serviço

- DimTempo - esta tabela guarda informação sobre o tempo (e.g., ano, semestre, trimestre, mês, dia, etc.) em que ocorre determinado facto (Tabela 5.29);

Tabela 5.29: Tabela de dimensão DimTempo do *Data Warehouse*

DimTempo				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimTempoPK	int	PK	-	Chave substituta
Data	date	-	-	Desginação da data (yyyy.mm.dd)
Ano	int	-	-	Descrição do ano (yyyy)
Semestre	varchar(50)	-	-	Descrição do semestre
Trimestre	varchar(50)	-	-	Descrição do trimestre
Mes	varchar(50)	-	-	Descrição do mês
Quinzena	varchar(50)	-	-	Descrição da quinzena do mês
Semana	int	-	-	Descrição da semana do ano
Dia	int	-	-	Descrição do dia do mês
EstacaoAno	varchar(10)	-	-	Descrição da estacao do ano

A DimTempo contém todas as datas possíveis para um calendário. A informação do tempo é gerada inicialmente a partir de um *script* (Anexo D - Criação da tabela DimTempo), encontrando-se a atual tabela preparada até à data atual. Para posteriores carregamentos, o *script* foi incluído no processo de ETL, de forma a incrementar diariamente um registo com a data atual.

- DimVendedor - esta tabela armazena informação sobre os vendedores da organização (e.g., nome e número) (Tabela 5.30);

Tabela 5.30: Tabela de dimensão DimVendedor do *Data Warehouse*

DimVendedor				
Atributo	**Tipo**	**Chave**	**SCD**	**Descrição**
DimVendedorPK	int	PK	-	Chave substituta
NumVendedor	numeric(4,0)	-	1	Chave original
NomeVendedor	varchar(3)	-	1	Descrição do vendedor

É importante referir que por questões de desempenho não são utilizadas chaves compostas, optando-se por utilizar uma chave substituta (*Surrogate Key*) definida como chave primária, para cada uma das dimensões, baseada num número inteiro que é atribuído sequencialmente (0, 1, 2, 3, ...) à medida que novos registos vão sendo inseridos. As chaves substitutas são utilizadas para relacionar as tabelas de factos com as tabelas de dimensão. Desta forma o *Data Warehouse* não fica vulnerável a problemas de duplicação de chaves durante o seu refrescamento (Rainardi, 2008).

Contudo, optou-se por guardar a chave original de algumas dimensões, de forma relacionar as mesmas com as tabelas de factos aquando o processo de integração dos dados. Porém este atributo não fica visível ao utilizador no sistema OLAP.

O modelo de *Data Warehouse* foi modelado a pensar na sua evolução futura, à medida que novas tabelas de factos e dimensões vão sendo adicionadas.

5.1.2 Processo de ETL

Após a caracterização e exploração dos dados operacionais e da implementação do modelo de *Data Warehouse,* anteriormente apresentado, procedeu-se à concretização do processo de ETL. Esta fase tem como principal objetivo extrair os dados dos sistemas OLTP para o ambiente do *Data Warehouse.* A ferramenta utilizada para concretizar o processo de ETL foi o SQL *Server Integration Services*, que permitiu a construção de *packages* de fluxos de dados e tarefas de extração, transformação, integração e carregamento dos dados.

De forma a exemplificar o processo procura-se, através de alguns exemplos, mostrar como todo o processo foi construído. A Figura 5.8 ilustra a metodologia seguida na concretização do processo de ETL.

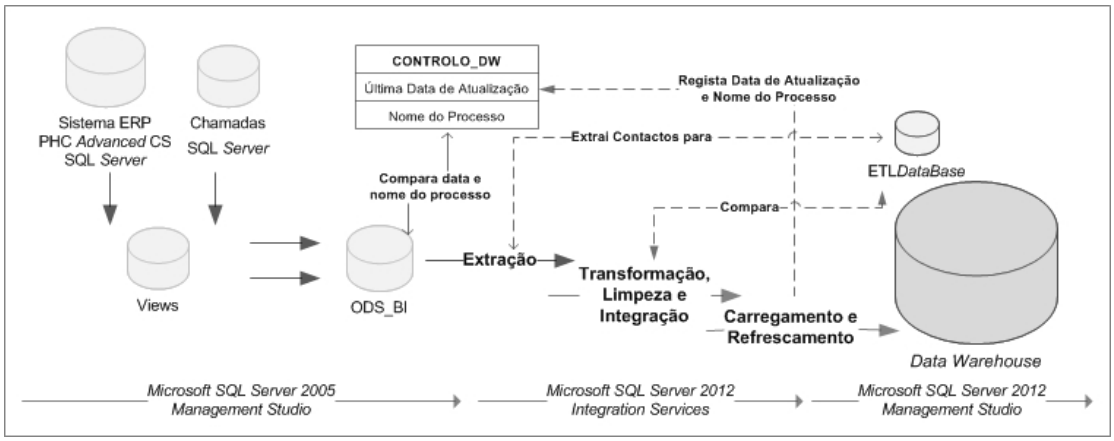


Figura 5.8: Metodologia do processo de ETL

Primeiramente é útil referir que a ”interface” entre o ambiente de fonte de dados operacionais e o ambiente de *Data Warehouse* é feito através das *views* que os sistemas OLTP disponibilizam. Contudo, optou-se por criar uma base de dados (ODS_BI) que armazena todos estes dados numa estrutura relacional e através da qual se evitam constrangimentos a nível do desempenho do sistema OLTP no momento de extrair os dados. O carregamento dos dados para a ODS_BI é efetuado diariamente através de *scripts* SQL. Neste processo inicial de extração são

simultaneamente efetuadas transformações nos dados (e.g., terminologia utilizada nas tabelas e nos atributos que as compõe).

A ETL*DataBase* surge como uma base de dados intermédia de suporte ao processo de ETL. Esta é utilizada para armazenar os contactos (telefone, telemóvel e fax) das entidades externas à organização. A tabela "CONTROLO_DW" encontra-se no modelo *Data Warehouse* e têm como principal tarefa apoiar o processo de extração e refrescamento dos dados.

No geral o processo de ETL é iniciado mediante um processo de carregamento dos dados diários na ODS_BI. Seguidamente, o processo de ETL é responsável por extrair os dados provenientes desta fonte de dados para a DSA onde estes vão ser transformados. Os dados resultantes destas transformações são carregados no *Data Warehouse*.

Depois de apresentada a metodologia utilizada para o processo de ETL, será de seguida descrito este processo com mais detalhe.

O primeiro passo foi desenvolver os processos de ETL das tabelas de dimensão e, posteriormente, os processos de ETL para as tabelas de factos. A Figura 5.9 ilustra o *package* geral do processo de ETL.

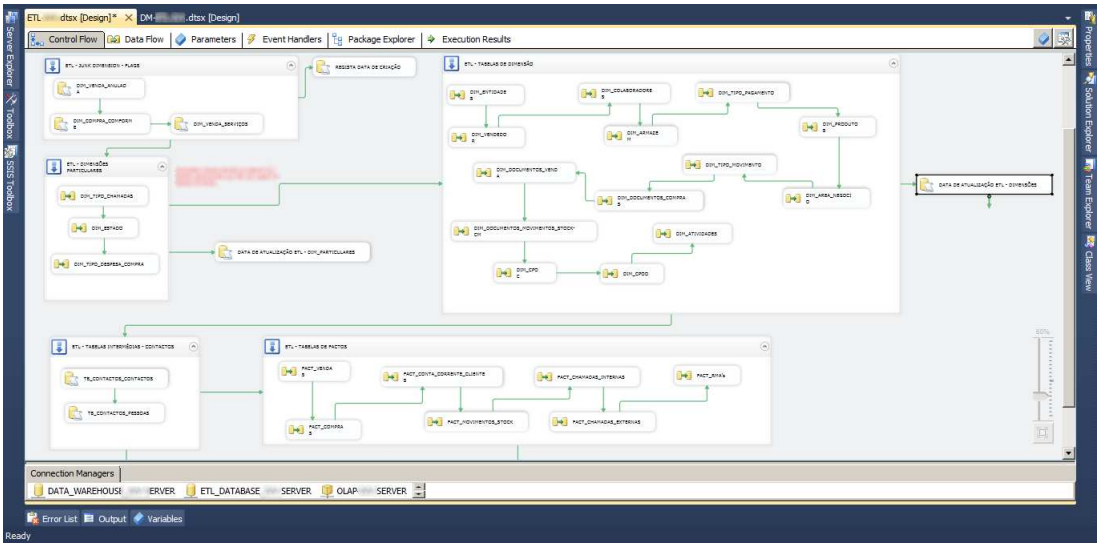


Figura 5.9: *Package* geral do processo de ETL

É importante referir que cada uma das tabelas de factos está relacionada com um conjunto de tabelas de dimensões através das chaves estrangeiras. Por este motivo, as tabelas de factos são sempre as últimas a ser carregadas e refrescadas no processo de ETL.

A extração dos dados é efetuada diariamente. Neste sentido, é necessário um mecanismo que permita apenas extrair da ODS_BI os dados que não se encontram

no *Data Warehouse*. Para que fosse efetivamente possível optou-se por criar uma tabela de controlo designada por "CONTROLO_DW". Esta tabela permite o armazenamento da última data de atualização e o nome do processo executado (e.g., "DIM_PRODUTOS"). A estrutura da tabela "CONTROLO_DW" é apresentada na Tabela 5.31.

Tabela 5.31: Tabela de controlo CONTROLO_DW

Tabela de controlo CONTROLO DW			
Atributo	**Tipo**	**Chave**	**Descrição**
ControloUpdateDWPK	int	PK	Chave original
DataAtualização	Datetime	-	Guarda a ultima data/hora de atualização
NomeProcesso	Date	-	Guarda o nome do processo

O processo de ETL (de uma tabela de dimensão ou uma tabela de factos) inicia-se com a extração dos dados. Nesta extração procede-se a uma verificação que vai permitir apenas extrair os novos registos. As tabelas da ODS_BI contêm um atributo ("dataultimamod") que permite o armazenamento da última data de atualização da tabela. Posteriormente esta data é comparada com a que se encontra registada na tabela "CONTROLO_DW" relativa ao último carregamento.

Como exemplo, a Figura 5.10 ilustra o processo (*script* SQL) de extração dos dados da tabela "BI_Produtos" (tabela que armazena a informação dos produtos da organização) para a DSA da DimProdutos. Os atributos da tabela "BI_Produtos" a

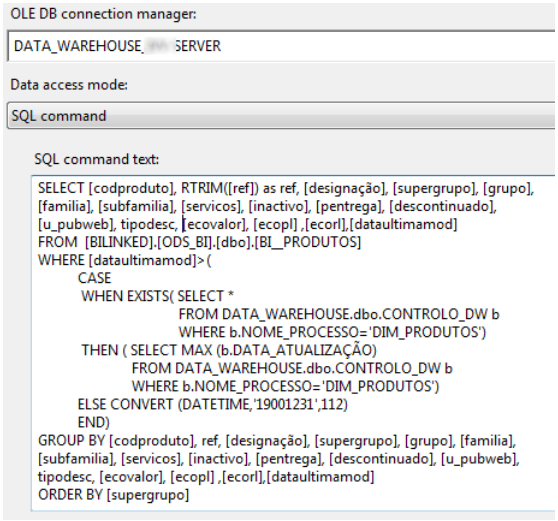


Figura 5.10: Exemplo da extração dos dados da tabela BI_Produtos

carregar na DimProdutos são selecionados, bem como o atributo "dataultimamod" utilizado apenas para o processo de extração.

Caso a data da tabela fonte seja superior à data de atualização do respetivo processo, os dados são extraídos e canalizados na DSA para percorrer todo o fluxo de ETL. Se a data da tabela fonte é inferior à data de atualização do processo, os dados não são extraídos e por consequente o processo de ETL não é executado.

Partindo do princípio que os dados são devidamente extraídos, prossegue-se com o processo de transformação e limpeza dos dados. A Figura 5.11 ilustra o fluxo de ETL completo para a DimProdutos. Através deste são efetuadas todas as transformações necessárias para garantir a qualidade e integridade dos dados (como referido na secção 4.2). Desta forma procurou-se assegurar que o processo de ETL tenha a capacidade de corrigir e ignorar dados rejeitando e/ou substituíndo valores por defeito. Porém, a correção pode não ser possível por inúmeras razões (e.g., falta de informação). No caso dos atributos *null* insere-se sempre valores equivalentes a "desconhecido". Optou-se também por ter um registo "desconhecido" nas tabelas de dimensão de forma a garantir que os dados sejam todos considerados na integração das várias tabelas (tabelas de dimensão e tabelas de factos).

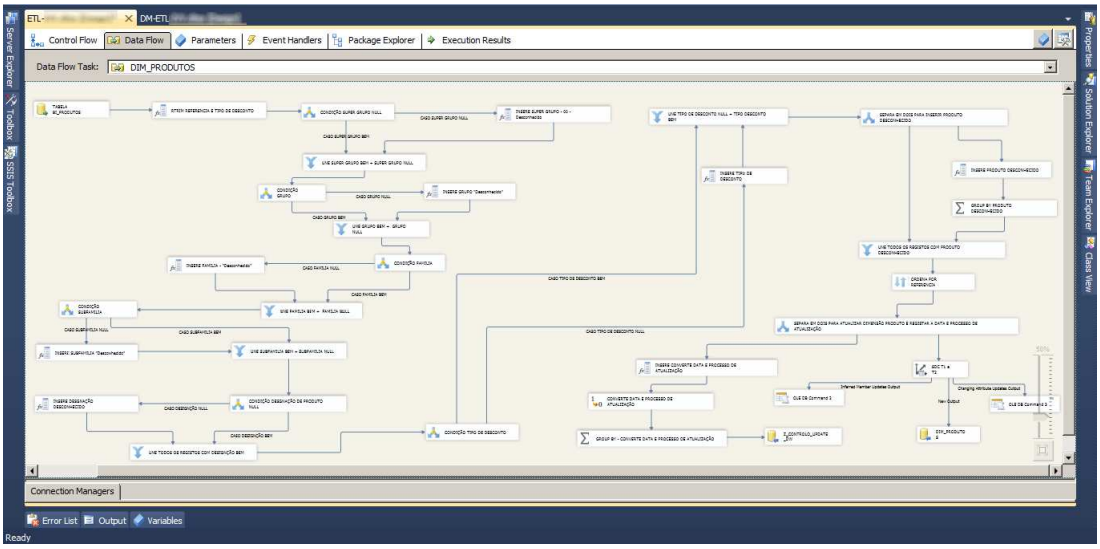


Figura 5.11: Processo ETL da DimProdutos

Na fase de carregamento dos dados é importante confirmar se o registo que se pretende colocar na tabela de dimensão já existe e, se existir, verificar os valores dos atributos que diferem do registo que se pretende inserir. Nestas situações, se o atributo se referir a um valor que pode ser alterado ao longo do tempo (SCD do tipo 1 ou 2) é atualizado, caso contrário, se este se refere a um atributo que não pode sofrer alterações na dimensão, então este é simplesmente ignorado. A Figura 5.12 ilustra como este processo é executado. O carregamento dos dados implica que seja também registado, na tabela "CONTROLO_DW", a data e hora do momento

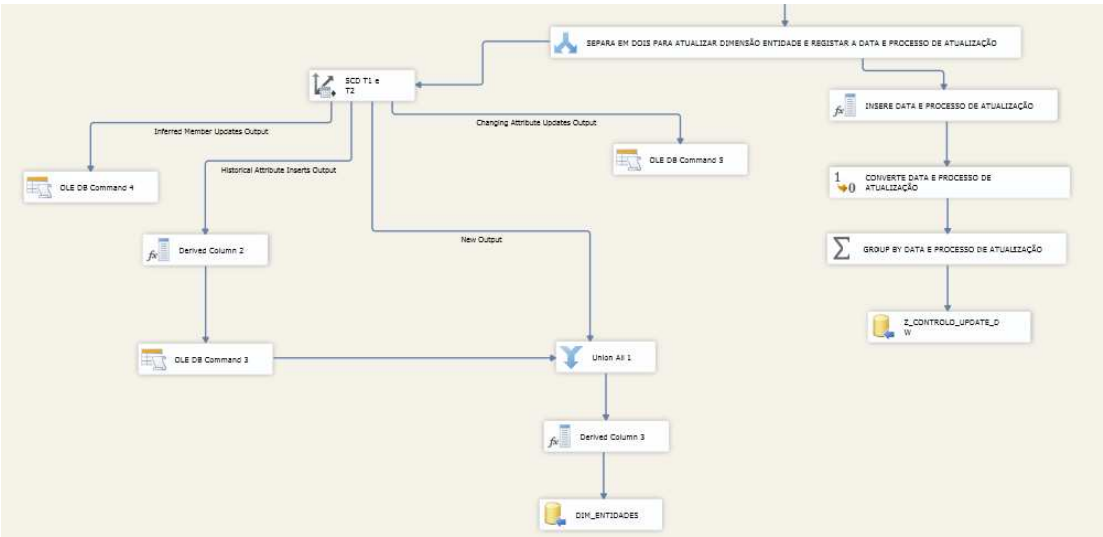


Figura 5.12: Carregamento dos dados na DimEntidades

em que ocorreu o último carregamento (Figura 5.13). Em todos os casos (tabelas

	CONTROLO_UPDATE_DW_PK	NOME_PROCESSO	DATA_ATUALIZAÇÃO
16	16	DIM_DOCUMENTOS_MOV_STOCK	2012-08-07 18:15:48.467
17	17	DIM_CPOC	2012-08-07 18:15:48.713
18	18	DIM_CPOO	2012-08-07 18:15:48.943
19	19	DIM_ATIVIDADES	2012-08-07 18:15:49.240
20	20	ETL - TABELAS DE DIMENSÃO	2012-08-07 18:15:49.310
21	21	FACT_VENDAS	2012-08-07 18:16:55.067
22	22	FACT_COMPRAS	2012-08-07 18:17:44.020
23	23	FACT_CONTA_CORRENTE_CLIENTES	2012-08-07 18:18:16.663
24	24	FACT_MOVIMENTOS_STOCK	2012-08-07 18:19:41.130
25	25	FACT_CHAMADAS_INTERNAS	2012-08-07 18:20:47.427
26	26	FACT_CHAMADAS_EXTERNAS	2012-08-07 18:21:03.260
27	27	FACT_RMA	2012-08-07 18:21:18.210

Figura 5.13: Tabela CONTROLO_DW

de dimensão e tabelas de facto), o fluxo de ETL de qualquer um dos processos, subdivide-se em dois fluxos (Figura 5.12), através do qual é possível a duplicação dos dados que vão ser carregados. No fluxo da esquerda os dados são carregados normalmente no *Data Warehouse*. No fluxo da direita, os dados são agregados num único registo, ao qual é atribuído o nome do processo (e.g., "DIM_PRODUTOS") e a data/hora de quando o processo ocorre.

Uma vez realizadas todas as transformações e encontrando-se as tabelas de dimensão devidamente carregadas/atualizadas, os dados ficam disponíveis para serem integrados e carregados nas tabelas de factos. No entanto, para as tabelas de factos, o processo de ETL é mais complexo em comparação com o das tabelas de dimensões. As tabelas de factos têm chaves estrangeiras (FK) para as várias dimensões, bem como os factos numéricos que requererem um tratamento especial (e.g., conversões de valores, cálculos de vários atributos, etc.). O processo para extrair os dados, das

tabelas que dão origem à tabela de factos, ocorre de igual modo ao das tabelas de dimensão. A Figura 5.14 mostra, como exemplo, a extração dos dados da tabela BI_VendasLIN.

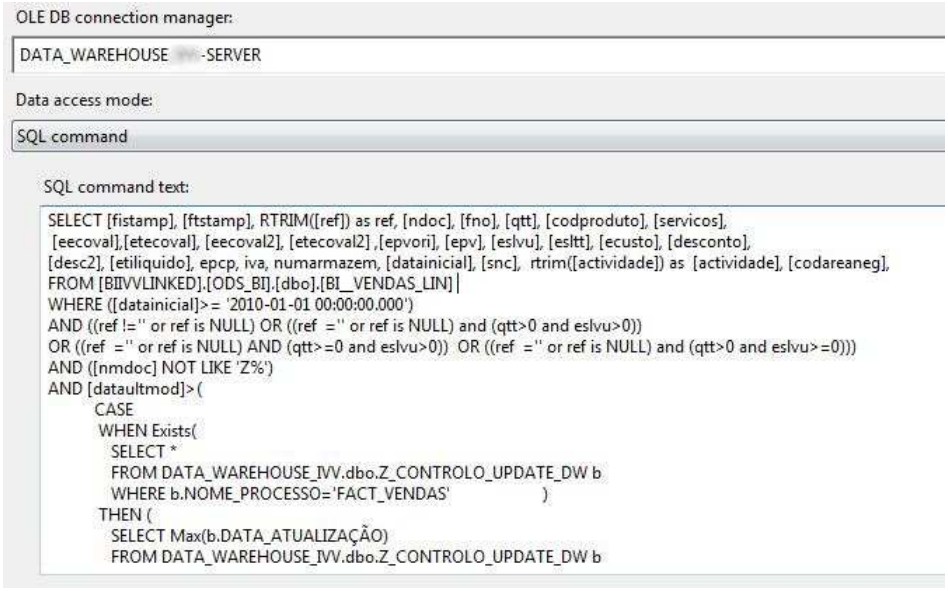


Figura 5.14: Exemplo da extração dos dados da tabela BI_VendasLIN

Como exemplo, as tabelas BI_VendasLIN e BI_VendasCAB vão dar origem à tabela de factos FactVendas. Nesta tabela existem chaves estrangeiras para as seguintes dimensões: DimEntidades, DimProdutos, DimVendedores, DimTempo, DimTipoPagamento, DimAreaNegocio, DimAmazem, DimVendaServico, DimVendaAnulada, DimCPOC, DimAtividade e DimDocumentosVenda. Portanto é necessário carregar as tabelas de factos não só com os factos, mas também com as chaves estrangeiras destas tabelas. Assim, após a extração, os dados são canalizados na DSA para percorrer todo o fluxo de transformação. Aqui os dados sofrem as transformações necessárias e são integrados. A Figura 5.15 mostra a junção, das tabelas BI_VendasCab e BI_VendasLIN.

Antes de os dados serem carregados, é necessário verificar se os atributos, que referenciam a dimensão, são válidos face às tabelas de dimensão existentes. Para tal utiliza-se a chave original que é guardada na tabela de dimensão. Na Figura 5.15 é possível observar o relacionamento da DimProdutos com a FactVendas. Para as restantes tabelas de dimensão o procedimento é o mesmo. Por outro lado, os factos são valores numéricos e obtidos maioritariamente partir do cálculo de vários atributos, para determinar por exemplo, o lucro bruto ou o valor de IVA (Figura 5.16). Tais cálculos são efetuados nos fluxos das tabelas de factos. Concluído todo processo de ETL os dados são carregados no modelo *Data Warehouse*.

Figura 5.15: Exemplo da integração da tabela FactVendas

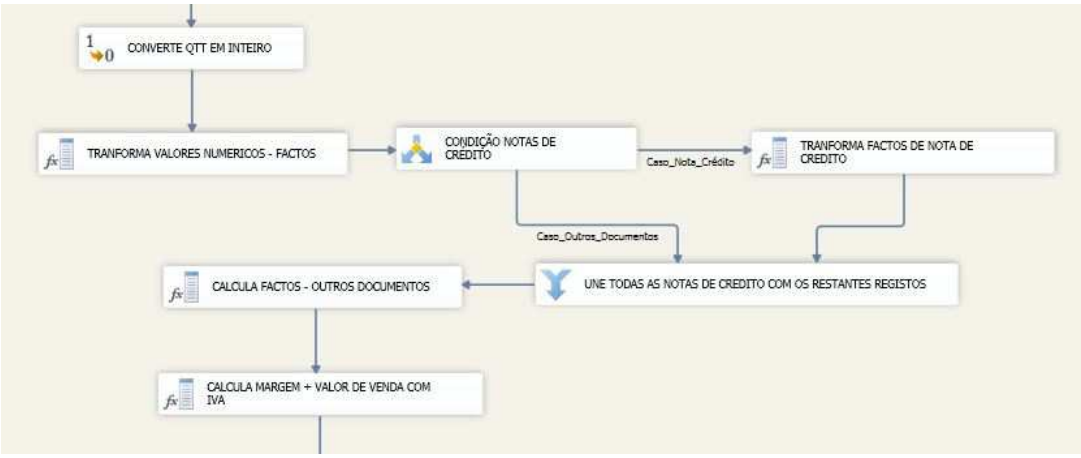


Figura 5.16: Parte do fluxo de ETL para transformação dos factos da FactVendas

Por fim, criou-se uma rotina (*Job*), no SQL *Server Agent*, responsável pela execução do processo de ETL, nomeadamente dos *packages* implementados. Esta rotina é executada com uma periodicidade diária, durante a noite, no servidor onde se encontra armazenado o *Data Warehouse*, não condicionando desta forma a performance dos sistemas. Contudo, o sistema de *Data Warehousing* encontra-se preparado para que tal periodicidade seja alterada, por exemplo, por semana, mês, ou mesmo por hora.

De um forma geral, a execução do processo de ETL implica que os dados sejam extraídos, transformados e carregados para o *Data Warehouse,* ficando depois disponíveis e acessíveis para serem trabalhados recorrendo aos sistemas OLAP e a técnicas de *Data Mining*.

5.2 Sistema de Processamento Analítico

Depois de apresentado o modelo do *Data Warehouse* e o respetivo processo de ETL, esta secção apresenta os resultados obtidos na análise dos dados utilizando os sistemas OLAP.

Após o armazenamento dos dados no *Data Warehouse,* os cubos OLAP que permitem efetuar consultas analíticas são implementados. A componente utilizada nesta etapa do projeto foi SQL *Analysis Services.*

Através da pré-definição das análises e das consultas pretendidas geraram-se vários cubos OLAP de análise, definindo-se numa primeira fase as fontes de dados dos cubos (ligação OLE DB ao *Data Warehouse*) e gerando-se as diferentes vistas dessas mesmas fontes (opção *data source view*). As particularidades do negócio da organização influencia as vistas. No caso concreto das vendas, foi necessário criar duas vistas: (1) uma com os documentos de faturação não anulados - FactVendasCF e (2) outra com todos os documentos de venda - FactVendas. As tabelas (de dimensão e de factos) do modelo físico mantêm-se inalteráveis, apenas sendo acrescentadas duas vistas diferentes das vendas.

Partindo das necessidades organizacionais verificou-se que seria necessário a construção de 8 (oito) cubos OLAP. Este processo iniciou-se com a seleção das tabelas de dimensão e de factos que vão constar nos cubos OLAP, sendo depois implementados os seguintes oito cubos OLAP:

- DM_CHAMADAS - engloba a Constelação Chamadas (FactChamadasInternas e FactChamadasExternas) o que permite analisar as chamadas sobre as diversas perspetivas que lhe estão subjacentes;

- DM_CONTA_CORRENTE - engloba a Estrela FactContaCorrenteClientes e permite a análise da conta corrente dos clientes sobre as diversas perspetivas que lhe estão subjacentes;

- DM_MOVIMENTOS_STOCK - engloba a Estrela FactMovimentosStock e permite a análise dos movimentos em stock sobre as diversas perspetivas que lhe estão subjacentes;

- DM_COMPRAS - engloba a Estrela FactCompras e permite a análise das compras sobre as diversas perspetivas que lhe estão subjacentes;

- DM_RMA - engloba a Estrela FactRMA e permite a análise dos RMA sobre as diversas perspetivas que lhe estão subjacentes;

- DM_VENDAS - engloba a Estrela FactVendas e permite a análise das vendas sobre as diversas perspetivas que lhe estão subjacentes;
- DM_VENDAS_CF - engloba a Estrela FactVendasCF e permite a análise das vendas sobre as diversas perspetivas que lhe estão subjacentes;
- DW_OLAP - engloba todas as tabelas de dimensão e tabelas de factos, o que permite a análise do negócio como um todo.

Após a definição da estrutura e construção dos cubos OLAP, os nomes dos atributos, métricas, dimensões e factos foram ajustados de forma proporcionar uma interpretação mais intuitiva, do ponto de vista dos gestores, aquando a utilização dos cubos OLAP. A Figura 5.17 ilustra o cubo DM_VENDAS_CF, criado especificamente para as análises das vendas (apenas com documentos de faturação).

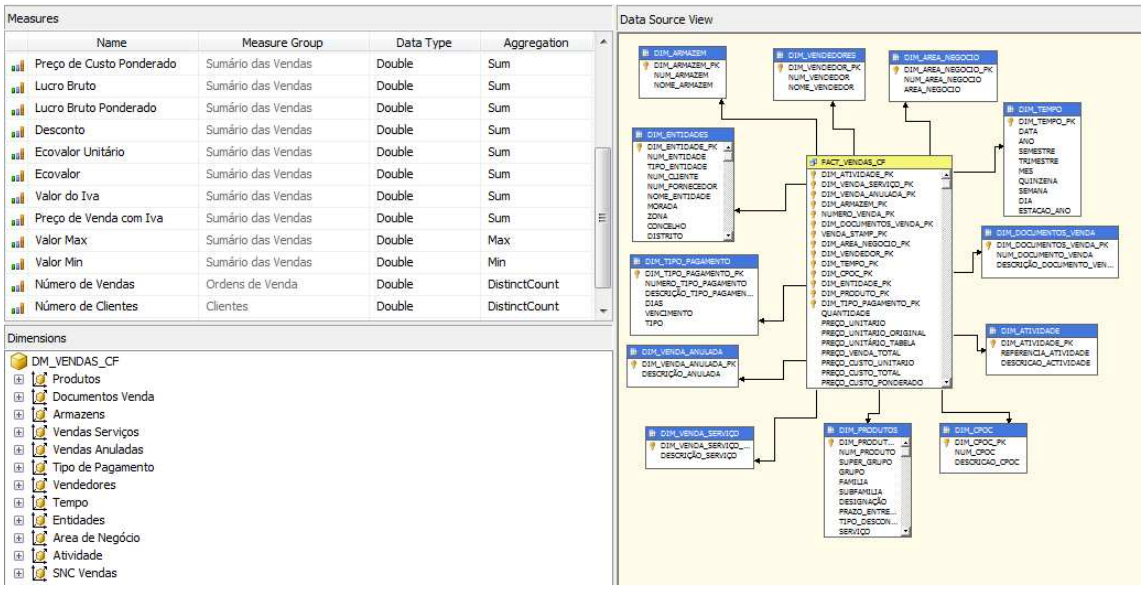


Figura 5.17: Cubo OLAP DM_VENDAS_CF

É importante referir que para cada uma das dimensões foram definidos os atributos que vão ser visualizados pelos gestores (ocultando por exemplo as chaves substitutas) e foram também criadas hierarquias que permitem o acesso à informação com maior ou menor detalhe.

As tabelas de dimensão que incluem hierarquias são:

- DimProdutos;
- DimEntidades;
- DimColaboradores;

- DimTipoChamada;
- DimTempo;
- DimHora;
- DimTipoMovimento, e;
- DimTipoPagamento.

As Figuras 5.18 e 5.19 mostram os atributos e as hierarquias definidas para a DimProdutos e DimTempo respetivamente.

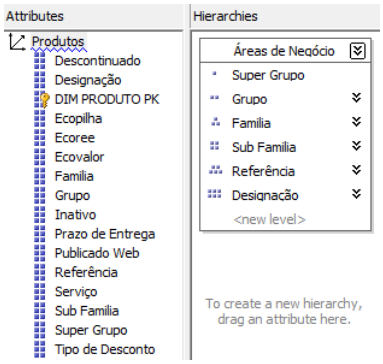


Figura 5.18: Hierarquia da tabela de dimensão DimProdutos

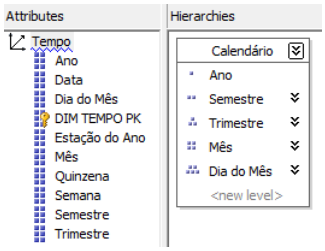


Figura 5.19: Hierarquia da tabela de dimensão DimTempo

A Figura 5.20 ilustra uma pesquisa ao longo da hierarquia (Localização Geográfica) definida para a DimEntidades.

A informação que é apresentada quando o gestor realiza consultas em níveis hierárquicos menos detalhados é assente em valores agregados. A utilização da hierarquia Localização Geográfica {Pais - Distrito - Concelho - Morada - Cliente} permite, por exemplo, analisar as vendas por País e, através da operação de *Drill-Down,* detalhar a análise a um nível inferior e saber especificamente qual o cliente.

Adicionalmente foram implementadas medidas calculadas e KPI com base nos factos (*measures*) considerados na modelação do *Data Warehouse*. Estas medidas

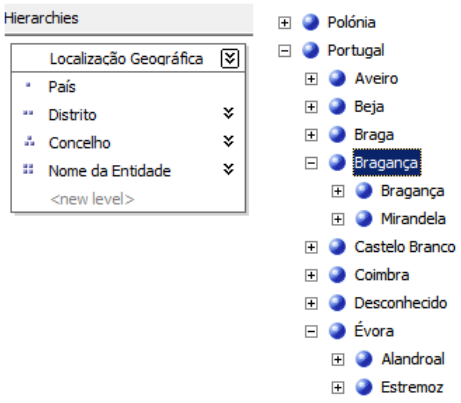


Figura 5.20: Hierarquia da tabela de dimensão DimEntidades

podem ser aditivas, semi-aditivas ou não-aditivas e representam KPI para a organização. A linguagem utilizada para a criação das medidas e dos KPI foi o MDX[1] (*MultiDimensional eXpressions*).

Na Figura 5.21 e 5.22 apresentam-se algumas das medidas calculadas e como exemplo a criação da medida "Variação da Faturação Percentagem" e "Saldo Médio". A primeira permite acompanhar a variação da faturação, em percentagem, ao longo da hierarquia de tempo (Calendário). A segunda permite o calculo do saldo médio do cliente.

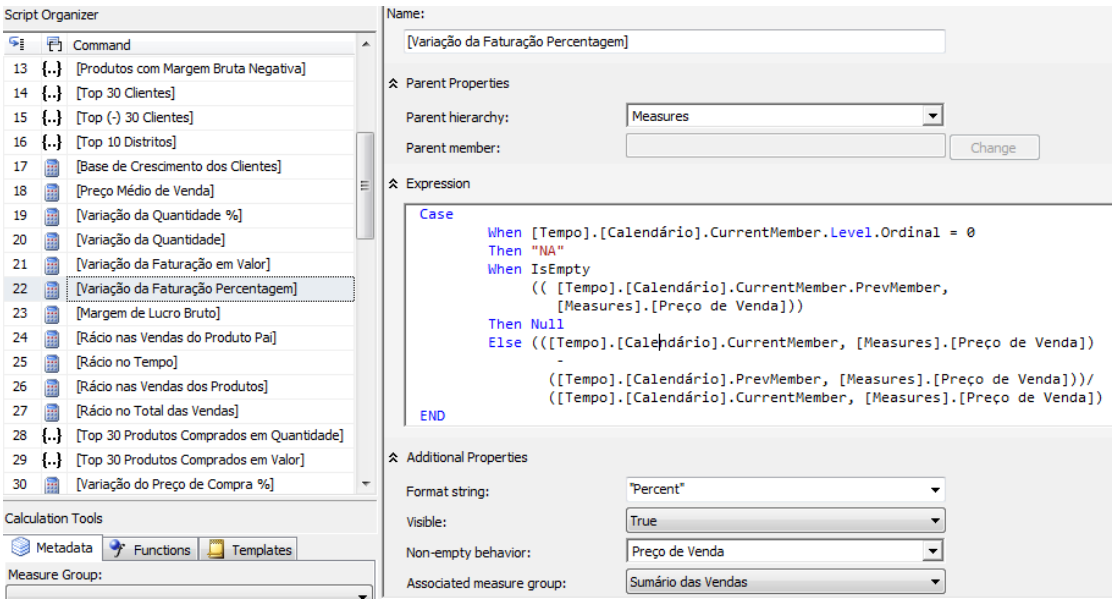


Figura 5.21: Medida calculada Variação da Faturação Percentagem

No geral foram criados os KPI considerados como prioritários para a gestão

[1]O MDX é uma extensão do SQL orientada a sistemas OLAP. Permite consultar e manipular os dados multidimensionais armazenados cubos OLAP.

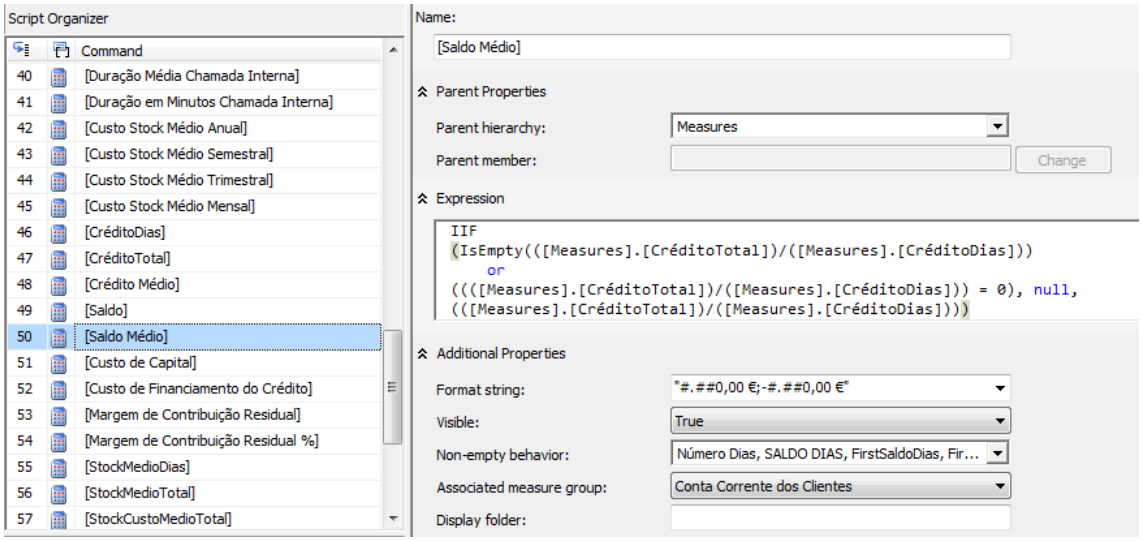


Figura 5.22: Medida calculada Saldo Médio do Cliente

estratégica da organização, entre o quais se destacam: margem de lucro por área de negócio, a rentabilidade do cliente (Figura 5.23), MCR por cliente, vendedor e área de negócio, o custo de stock médio por área de negócio, o saldo médio por cliente, o custo de financiamento do ativo económico, o custo do financiamento do crédito, a variação da faturação, peso da área de negócio nas vendas, e outros indicadores.

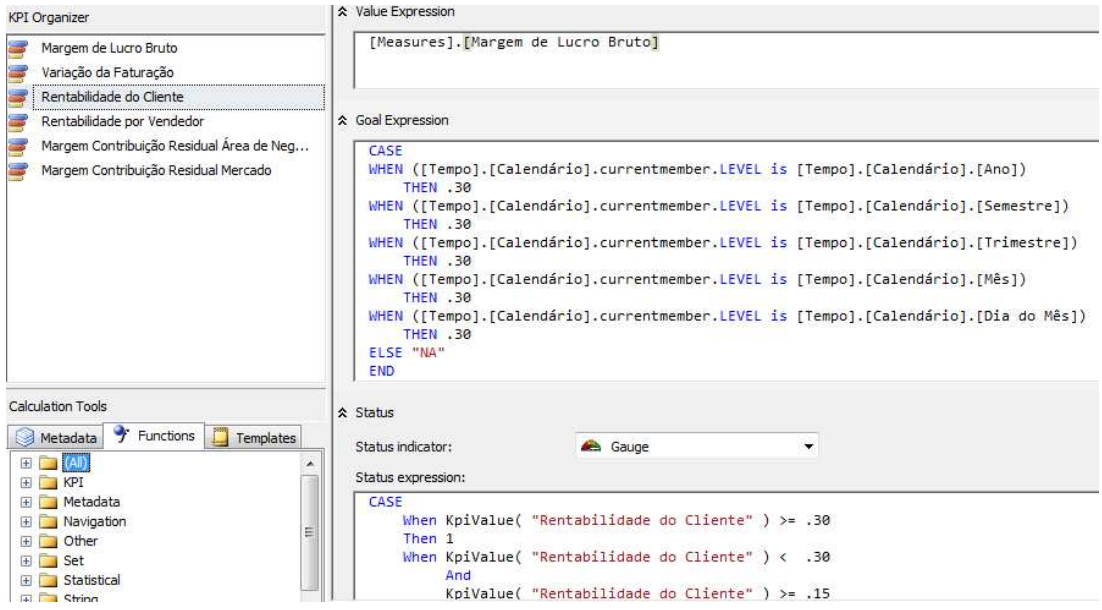


Figura 5.23: Exemplo do KPI Rentabilidade do Cliente

O sistema OLAP assenta numa arquitetura MOLAP (Han & Kamber, 2011). O seu excelente desempenho e a rápida indexação de dados pré-processados leva a crer que esta é a arquitetura mais adequada para a concretização dos objetivos do projeto.

O processamento do sistema OLAP ocorre diariamente após a conclusão do processo de ETL. Desta forma os dados encontram-se acessíveis em qualquer momento.

A implementação do sistema OLAP possibilitou a utilização de mecanismos de pesquisa e análise dos dados, através dos quais se podem criar gráficos, tabelas e novos KPI recorrendo a ferramentas de criação/exploração de relatórios como o *Microsoft Reporting Services*, do *Microsoft Report Builder* e do *Microsoft Office Excel.* Na secção 5.4 são apresentados exemplos da exploração do sistema OLAP.

5.3 *Data Mining*

As limitações dos gestores, em analisar e detetar relações, tendências e padrões em grandes quantidades de informação, leva a que estes procurem por soluções que possibilitem a exploração destes conjuntos de dados. Uma dessas soluções é o *Data Mining,* como foi apresentado na secção 2.2.3.

As técnicas de *Data Mining* a utilizar dependem dos objetivos inerente às análises a efetuar. As necessidades da organização são específicas do seu negócio e procurar/encontrar padrões/tendências nas vendas é um objetivo. Desta forma pretende-se, como objetivos para o *Data Mining,* a identificação de modelos que permitam:

1. Encontrar relações nos dados das vendas sobre os produtos vendidos, identificando quais aqueles que frequentemente são vendidos em conjunto, quer seja no global das vendas, por cliente ou mesmo por área de negócio;

2. Prever as vendas atendendo a fatores como desconto, serviços, prazo de entrega ou localização;

3. Encontrar similaridades nos dados dos clientes mais frequentes, quer seja por localização, por área de negócio ou no tempo.

A metodologia seguida, neste projeto, para o processo de *Data Mining* foi a CRISP-DM. As duas atividades iniciais "Compreensão do Negócio" e "Compreensão dos Dados" foram executas aquando da conceção e implementação do sistema de *Data Warehousing.* Continua-se aqui o processo a partir da atividade "Preparação dos Dados".

A ferramenta utilizada para identificar modelos de *Data Mining* foi o SQL *Analysis Services* que permite a preparação dos dados e a utilização de vários algoritmos.

5.3.1 Preparação dos Dados

A preparação dos dados envolve todas as atividades (seleção de tabelas, registos e atributos, assim como a transformação e limpeza dos dados) necessárias para a construção de um conjunto de dados final para ser analisado através de algoritmos de *Data Mining*.

A preparação dos dados envolve todas as atividades (seleção de tabelas, registos e atributos, assim como a transformação e limpeza dos dados) necessárias para a construção de um conjunto de dados final para ser analisado através de algoritmos de *Data Mining*.

Sendo o objetivo principal encontrar padrões e tendências nas vendas, recorre-se à tabela de factos FactVendas do *Data Warehouse* e às dimensões a si associadas. O resultado da junção das tabela de factos com as várias dimensões resulta numa tabela achatada. Esta tabela apenas contém os atributos que podem ser relevantes para o processo de descoberta de conhecimento. Neste caso em particular foram considerados apenas atributos com interesse das tabelas de dimensão (excluindo as chaves substitutas e as chaves originais) e todos os factos presente na tabela de factos. Optou-se por considerar apenas os documentos de faturação e não considerar as vendas anuladas. No entanto, os dados podem posteriormente ser trabalhados e os atributos podem, ou não, ser tidos em consideração nos modelos.

Para além dos atributos das várias dimensões e da tabela de factos, criaram-se classes de agrupamento para os factos numéricos (e.g, preço de venda, preço de custo, quantidade, margem bruta, etc.).

Foi criado um processo automatizado, com recurso ao SQL *Server Integration Services*, que permite a extração dos dados do *Data Warehouse* (com as condições supra-mencionadas), a junção das tabelas, a criação das classes e o carregamento dos mesmos na tabela de achatada.

5.3.2 Modelação e Validação

Neste trabalho estava previsto considerar as seguintes três abordagens: **Associação**, **classificação** e **segmentação**. Para o modelos de associação (objetivos 1), foi testado o algoritmo *Microsoft Association Rules*. Para os modelos de classificação (objetivos 2), estava previsto testar o algoritmo *Microsoft Decision Trees.* No que diz respeito aos modelos de segmentação (objetivo 3), estava previsto testar o algoritmo de *Microsoft Clustering.*

No entanto, face às limitações temporais, apenas foi concretizado o primeiro objetivo de *Data Mining*.

Regras de Associação

Para encontrar grupos de itens que ocorrem simultaneamente nas transações, determinando quais os produtos que fazem parte do mesmo centro de interesse recorre-se a uma abordagem de associação, utilizando a técnica *Microsoft Association Rules.*

É expectável que uma transação possa conter mais do que um produto, sendo que, aqueles produtos que são vendidos com maior frequência apresentem uma maior relação entre si. Parte-se do princípio que as vendas na organização ocorrem de uma forma linear, nas diversas áreas de negócio, isto é, as vendas de um produto A (e.g., central de alarme) implica sempre a venda do produto B (e.g., suporte da central) e do produto C (e.g., cabo de dados). Isto ocorre frequentemente, pois um produto raramente é vendido isoladamente.

Foi necessário definir uma estratégia de análise de dados, tendo a mesma sido realizada de três formas:

1. Análise por transações efetuadas, procurando identificar quais os produtos vendidos em conjunto numa mesma transação;

2. Análise da relação dos produtos vendidos por cliente ao longo do tempo, e;

3. Análise da relação dos produtos vendidos por área de negócio (por transação e/ou cliente);

O algoritmo utilizado necessita, como requisito, de uma chave única que identifica uma transação, de um atributo a prever e dos atributos de *input* que serão utilizados.

Para a estratégia de análise de dados definida, foram realizadas as seguintes operações:

(1) Para identificar quais os produtos vendidos em conjunto numa mesma transação foram inicialmente considerados todos os dados disponíveis, incluindo todas as áreas de negócio. Dos atributos que se encontram no conjunto de dados apenas se considera o número da transação (NumVendaPK) como chave única de identificação de uma transação e a referência do produto (Produto) que é utilizado como atributo de previsão e também como input do modelo (Figura 5.24).

(2) Para identificar a relação dos produtos vendidos por clientes ao longo do tempo, a chave única para identificar uma venda é o número de cliente (NumClientePK) e a referência do produto (Produto) que é utilizado como atributo de previsão e também como input do modelo (Figura 5.24).

(3) Para identificar a relação dos produtos vendidos por área de negócio, o conjunto de dados utilizado contém unicamente as transações das vendas por área de

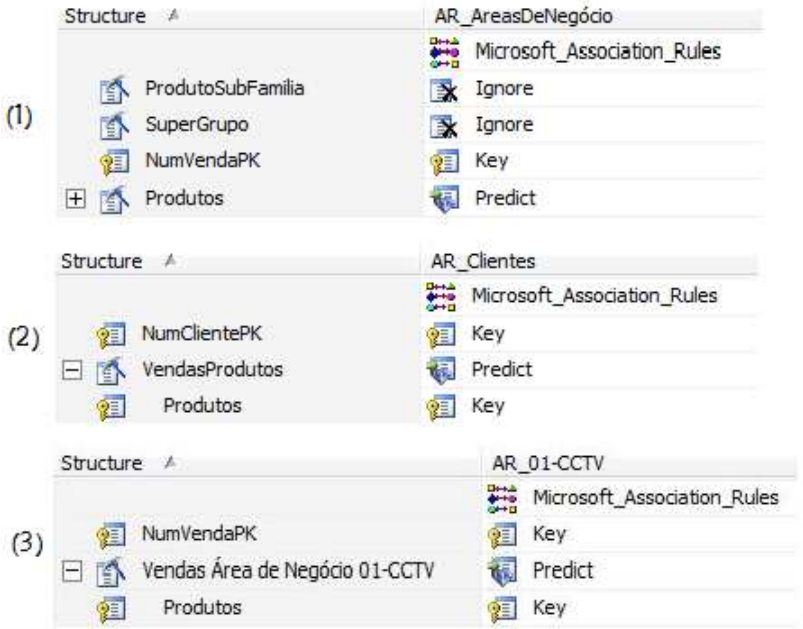


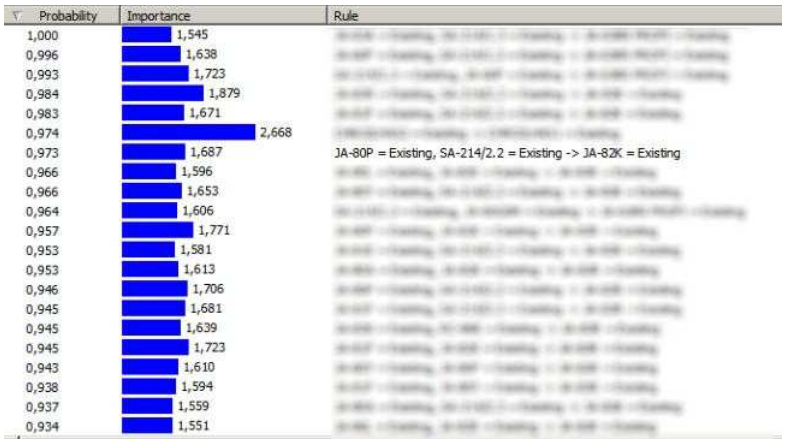
Figura 5.24: Atributos utilizados nas Regras de Associação

negócio, por exemplo, para análise da área de negócio CCTV são utilizadas somente as transações de CCTV. Dos atributos que se encontram no conjunto de dados apenas se considera o número da transação (NumVendaPK) como chave única de identificação de uma transação e a referência do produto (Produto) que é utilizado como atributo de previsão e também como input do modelo (Figura 5.24).

É de realçar que os parâmetros do algoritmo foram alterados limitando o número mínimo de suporte a ”0.01” e o mínimo de probabilidade a ”0.5”, de forma a que o modelo devolve-se o maior número de regras possíveis, mas com uma probabilidade mínima de 50%. Optou-se por considerar (após realizados alguns testes para determinar qual a divisão ideal) que 70% dos dados vão para treino e os restantes 30% dos dados para teste.

Os resultados obtidos resultam num conjunto de regras que associam os produtos vendidos em simultâneo. Contudo, os modelos geraram demasiadas regras e parte delas triviais para a organização.

Na primeira análise obtiveram-se inúmeras regras de associação que relacionam a venda dos produtos. A Figura 5.25 apresenta os resultados obtidos, ordenados por probabilidade (confiança) das regras, após o processamento do algoritmo. Os resultados obtidos permitiram identificar 163 regras com uma probabilidade compreendida entre 50% e 100%. Olhando para a regra que se encontra visível, é possível afirmar que se ocorrer uma venda do produto ”JA-80P” e do produto ”SA-214/2.2” então o produto ”JA-82K” é vendido com uma confiança de 97%.

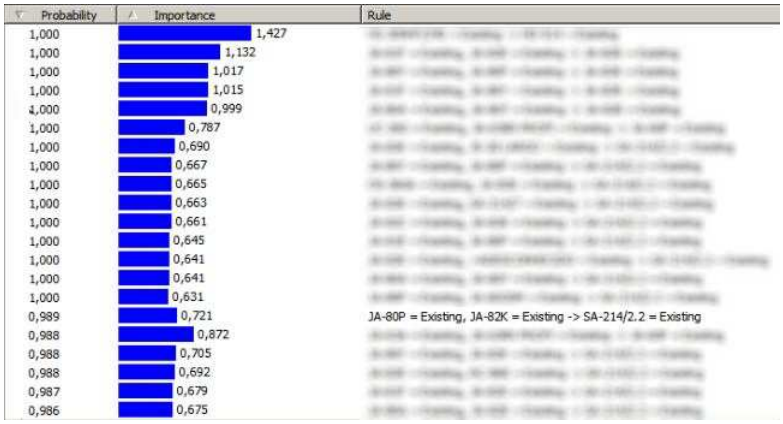


Probability	Importance	Rule
1,000	1,545	
0,996	1,638	
0,993	1,723	
0,984	1,879	
0,983	1,671	
0,974	2,668	
0,973	1,687	JA-80P = Existing, SA-214/2.2 = Existing -> JA-82K = Existing
0,966	1,596	
0,966	1,653	
0,964	1,606	
0,957	1,771	
0,953	1,581	
0,953	1,613	
0,946	1,706	
0,945	1,681	
0,945	1,639	
0,945	1,723	
0,943	1,610	
0,938	1,594	
0,937	1,559	
0,934	1,551	

Figura 5.25: Conjunto de Regras de Associação identificadas para as vendas dos produtos (1)

Relembra-se que a maioria dos resultados estão ocultos devido à confidencialidade dos dados tratados.

O resultado da segunda análise difere daquele que foi obtido após a primeira iteração. Nesta, as regras baseiam-se no contexto do cliente, ou seja, o relacionamento dos produtos comprados por clientes. A Figura 5.26 apresenta os resultados obtidos e olhando para a regra que se encontra visível, é possível afirmar que se ocorrer uma venda do produto "JA-80P" e do produto "JA-82K" então o produto "SA-214/2.2" é vendido com uma confiança de 99%.



Probability	Importance	Rule
1,000	1,427	
1,000	1,132	
1,000	1,017	
1,000	1,015	
1,000	0,999	
1,000	0,787	
1,000	0,690	
1,000	0,667	
1,000	0,665	
1,000	0,663	
1,000	0,661	
1,000	0,645	
1,000	0,641	
1,000	0,641	
1,000	0,631	
0,989	0,721	JA-80P = Existing, JA-82K = Existing -> SA-214/2.2 = Existing
0,988	0,872	
0,988	0,705	
0,988	0,692	
0,987	0,679	
0,986	0,675	

Figura 5.26: Conjunto de Regras de Associação identificadas para as vendas dos produtos (2)

Na primeira e na segunda análises, as regras evidenciaram o conhecimento do domíno existente. Duas áreas negócio predominam nas vendas e os produtos vendidos nestas áreas, são produtos já com alguma relação entre si, como foi explicado no início desta secção. Contudo, procurou-se discriminar esta relação, e como tal, foi estabelecida uma terceira análise para relacionar os produtos vendidos por área

de negócio (particionar o conjunto de dados filtrando-o por uma área de negócio especifica). Os resultados obtidos são naturalmente diferentes por áreas de negócio. A Figura 5.27 mostra os resultados obtidos para a área de negócio CCTV. Foram identificadas 6 regras com probabilidades compreendidas entre 50% e 100%.

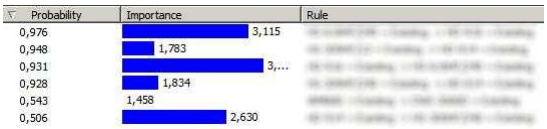


Figura 5.27: Conjunto de Regras de Associação identificadas para as vendas dos produtos (3)

Os modelos identificados visam encontrar relações de associação entre os produtos que são vendidos em conjunto. Desta forma, foram definidas as três estratégias, de análise aos dados, referidas anteriormente. Acredita-se, no entanto, que os modelos identificados possam ser refinados e como consequência melhorar os resultados obtidos.

5.3.3 Implementação

Num projeto de *Data Mining* a criação de um modelo não significa o final do projeto. Naturalmente, o conhecimento extraído deve ser organizado e apresentado para que possa ser utilizado. Assim, face aos requisitos definidos, geraram-se relatórios (secção 5.4) dos modelos identificados anteriormente (Regras de Associação). Estes relatórios ficam disponíveis e acessíveis aos gestores através da plataforma *web.*

Na próxima secção são apresentadas as aplicações de *front-end* de acesso à informação, bem como alguns exemplos de relatórios implementados.

5.4 Aplicações de *Front-End*

Um dos objetivos deste projeto de dissertação foi disponibilizar uma plataforma de *Business Intelligence* integrada e que permitisse o acesso à informação de uma forma fácil e intuitiva.

Nesta secção faz-se referência às aplicações de *front-end* que constituem a componente mais visual do sistema implementado, na qual os gestores vão poder aceder à informação sobre o negócio, identificando tendências e observando os KPI de negócio que a organização pretende monitorizar.

As aplicações permitem o acesso *ad-hoc* aos dados, fazer operações de *drill-down* e *roll-up* e, analisar e acompanhar a informação dos KPI através de um ambiente

web. Os gestores podem ainda criar os seus próprios relatórios e as suas consultas partindo de qualquer um dos cubos OLAP implementados, e publicar os mesmos no servidor de *Business Intelligence* onde ficam disponíveis para posteriores análises.

A ferramenta utilizada para a construção dos relatórios foi o SQL *Reporting Services* e como recurso adicional o SQL *Report Builder,* através da qual foram criados relatórios das análises ditas predominantes e necessárias para a gestão estratégica da organização.

Na construção dos diversos relatórios foram tidos em consideração os seguintes aspetos:

- Manter um interface simples, com recurso a gráficos e tabelas sem grande variação do tipo e cores utilizados; e
- Permitir a definição de parâmetros (filtros) nas análises a realizar.

Primeiramente foram implementados os relatórios inerentes às análises frequentemente realizadas pela organização. Posteriormente, com base nos KPI identificados, foram implementados relatórios associados a cada um dos segmentos de análise.

De seguida são apresentados diversos exemplos que visam evidenciar os diversos relatórios implementados para analisar as diversas vertentes do negócio.

A Figura 5.28 ilustra um relatório relativo à faturação anual da organização.

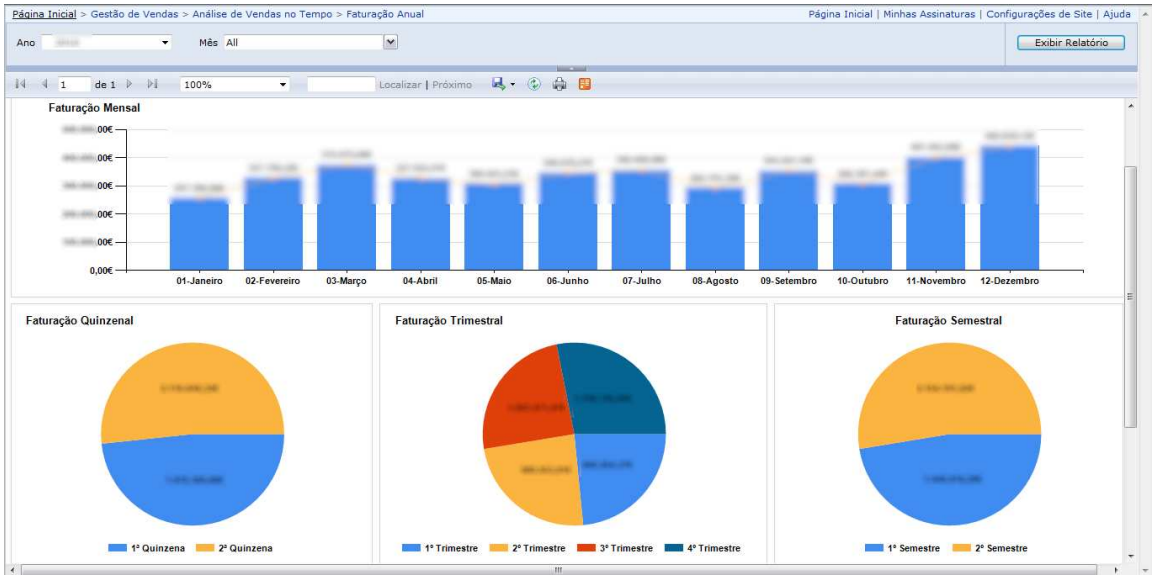


Figura 5.28: Relatório da Faturação Anual

Através deste relatório é possível realizar diversas análises que permitem acompanhar a evolução da faturação quinzenal, mensal, trimestral, semestral ou anual.

Por sua vez, a Figura 5.29 ilustra uma análise da variação mensal num período de tempo.

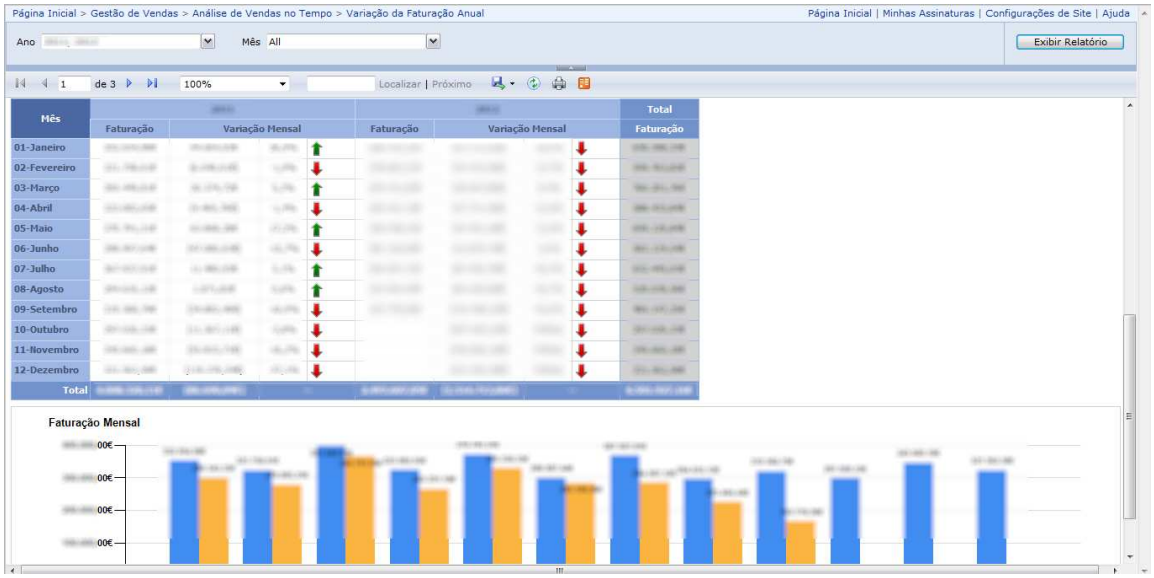


Figura 5.29: Relatório com a Variação da Faturação Mensal

As colunas da tabela representam o ano, já cada linha da tabela representa o respetivo mês. É assim possível analisar a faturação e variação mensal (em valor e percentagem) ao longo dos anos, sendo acompanhado com um indicador que permite avaliar o desempenho mensal de uma forma mais eficaz. O gráfico representa a faturação mensal dos anos em análise. O que permite acompanhar e verificar a tendência da faturação ao longo do tempo.

Para uma análise mais personalizada, como é aquela que é apresentada na Figura 5.30, referente a análise MCR do segmento de Mercado, foi necessário a utilização do cubo DW_OLAP, que engloba todas as tabelas de factos e dimensões. Como é possível observar na Figura 5.30 esta análise refere-se ao top 30 clientes do período em análise, neste caso semestral (editável nos parâmetros), através da qual se pode analisar e acompanhar diversos KPI como: as Vendas Liquidas, os Custo do Produto, MCB, Rentabilidade, Custos com o Cliente, MC, Crédito Médio concebido ao cliente, Custo Financiamento do Crédito, MCR de mercado e a Taxa de MCR.

Olhando a linha visível, pode-se verificar que o cliente, que se encontra no top 30 clientes do ano e semestre em análise, apresenta o volume de faturação na ordem dos 46.928,98€, um custo com os produtos adquiridos na ordem dos 33.261,70€, o que reflete uma MCB no valor dos 13.667,28€ e numa rentabilidade de 29,1%. A MCR do cliente é de 28,5%, abaixo dos 30%, condicionando desta forma a cor do indicador.

A Figura 5.31 apresenta uma outra análise mensal referente ao segmento de Mercado. Esta análise é continuação da anterior, mas neste caso especifico refere-se a um período de tempo menor (mês).

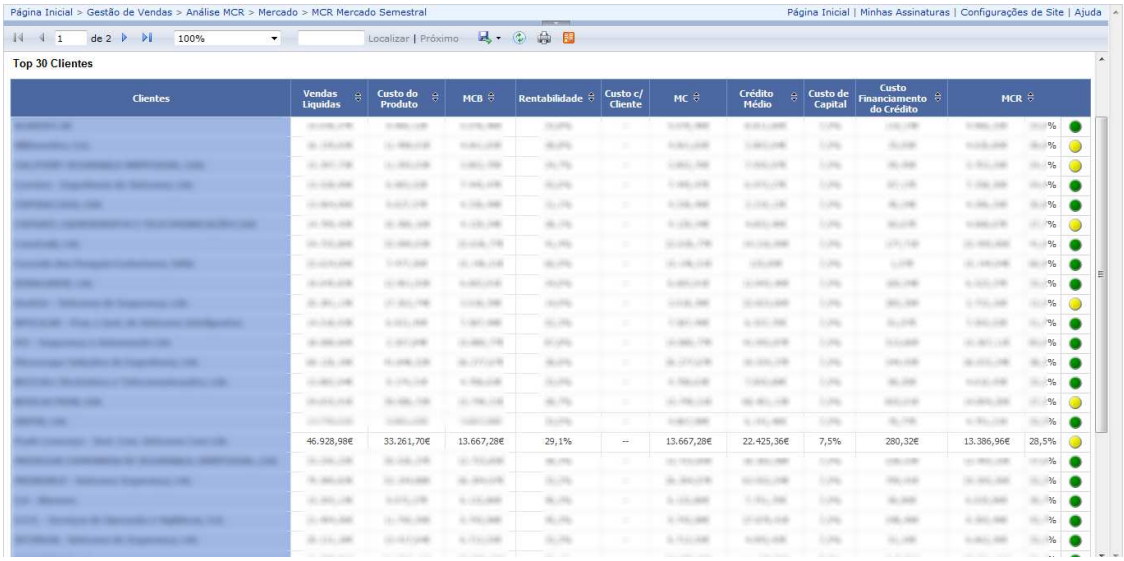


Figura 5.30: Relatório da análise MCR do Mercado

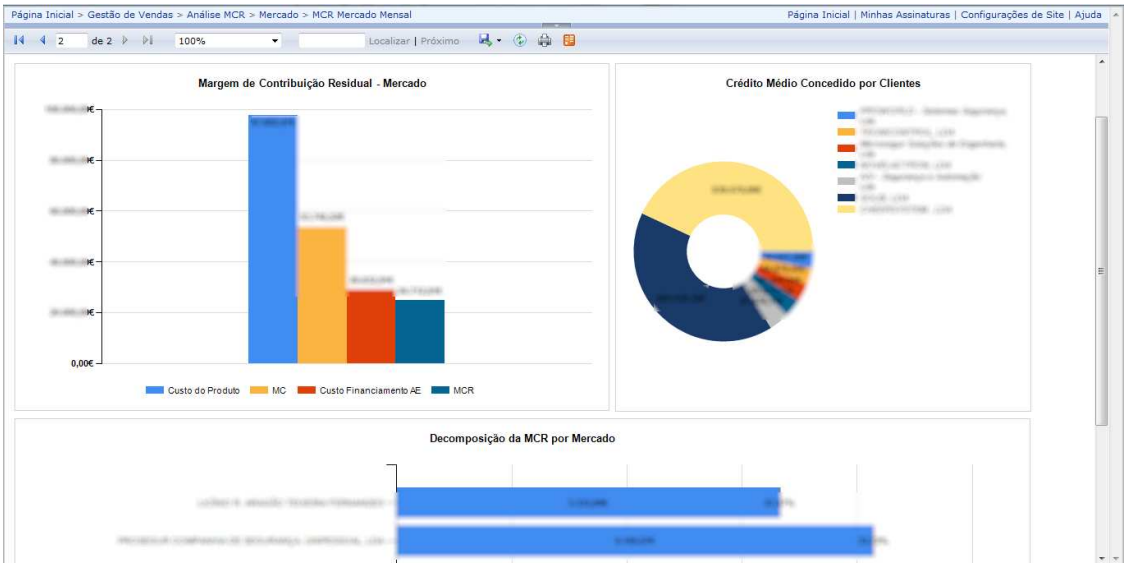


Figura 5.31: Relatório da análise MCR do Mercado Mensal

Nesta é possível analisar a MCR no período em análise, recorrendo a gráficos circulares e de barras, que permitem medir o impacto dos top 30 de clientes no global das vendas.

A Figura 5.32 apresenta um relatório da análise MCR Mensal por Responsabilidade. Nesta é possível observar que a análise efetuada difere muito pouco da anterior, mas neste caso em concreto permite analisar as vendas por cada um dos vendedores da organização (3 vendedores) quer seja no global da vendas ou no top 30 de clientes do período em análise. Os gestores têm a possibilidade de manipular estas análises (recorrendo às opções dos parâmetros) selecionando o período de

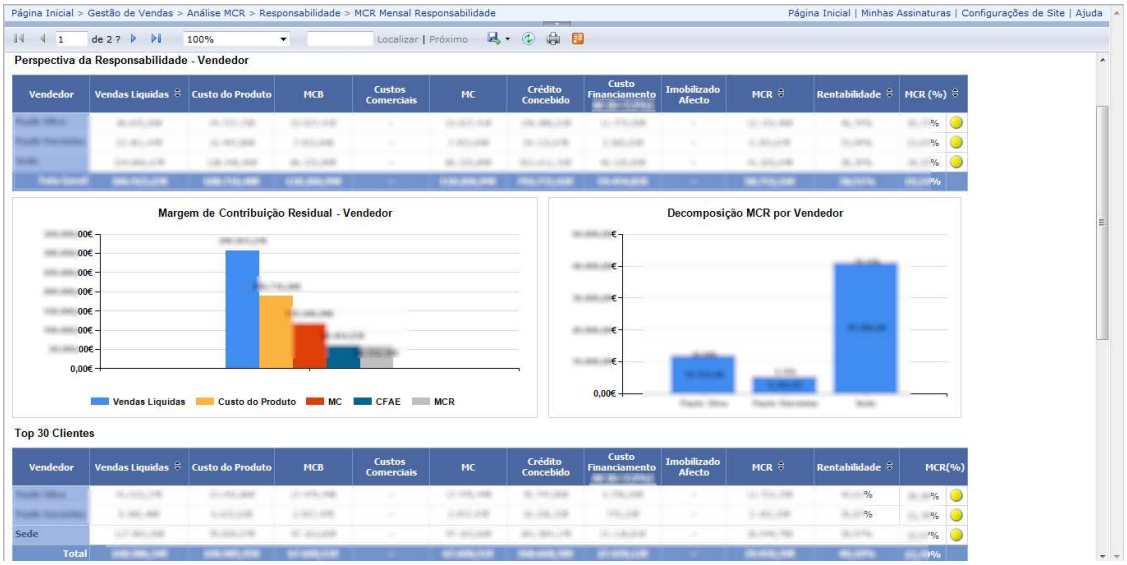


Figura 5.32: Relatório da análise MCR Mensal por Responsabilidade

tempo que se pretende analisar, assim como a possibilidade de ordenar os resultados por KPI.

A Figura 5.33 e a Figura 5.34, apresentam análises que permite acompanhar a

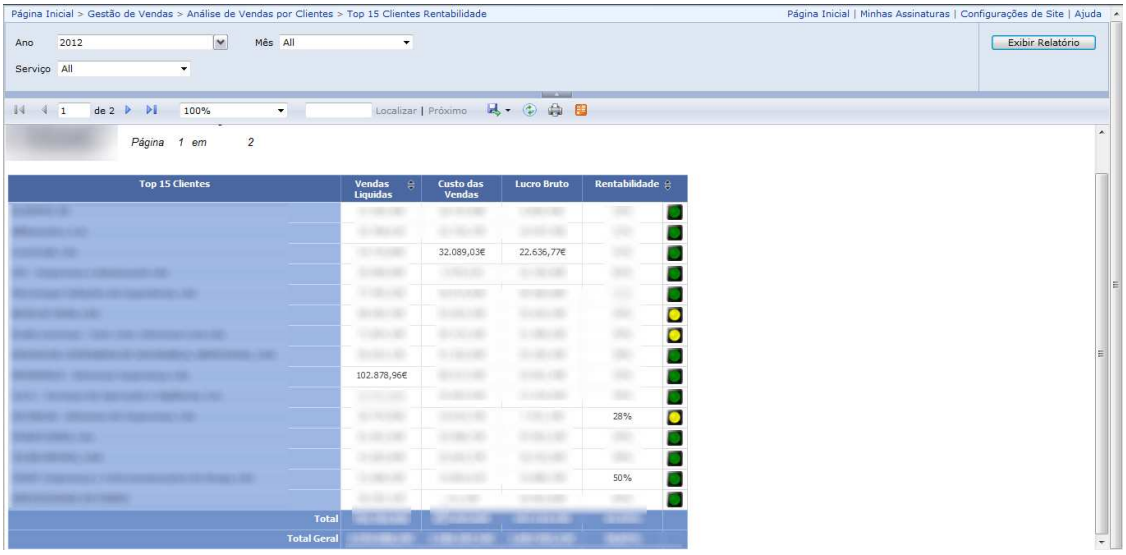


Figura 5.33: Relatório da análise Rentabilidade do Cliente

rentabilidade por cliente (top 15) no período em análise ou acompanhar a rentabilidade por localização geográfica, ao longo da hierarquia Localização Geográfica.

Na Figura 5.35 é possível observar uma análise das chamadas telefónicas de um determinado colaborador, através da qual é possível analisar a duração e número de chamadas externas e internas por período do dia ou acompanhar as mesmas

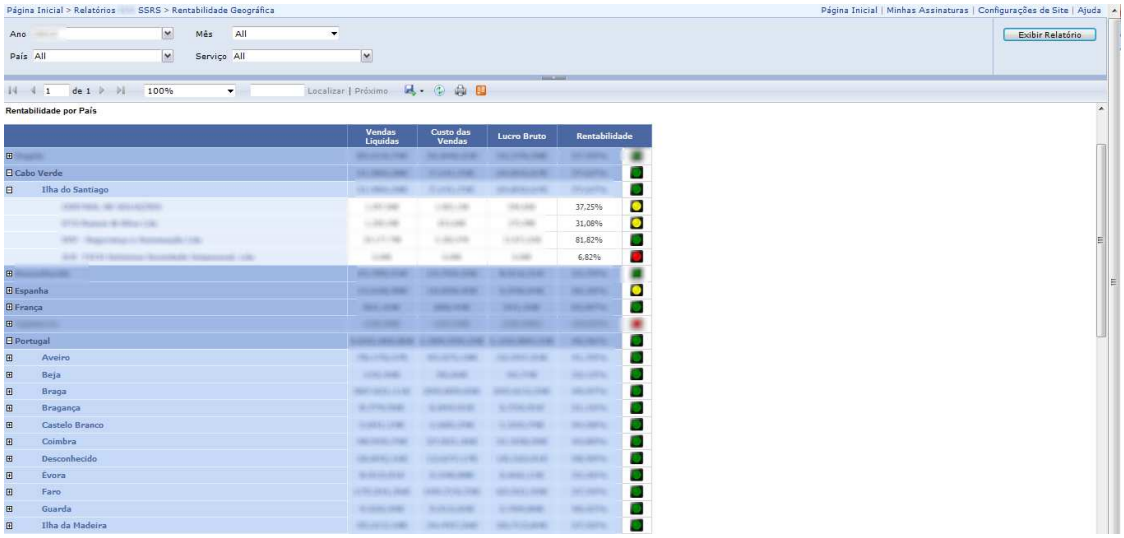


Figura 5.34: Relatório da análise Rentabilidade por Localização Geográfica

diariamente ao longo do mês, entre outras.

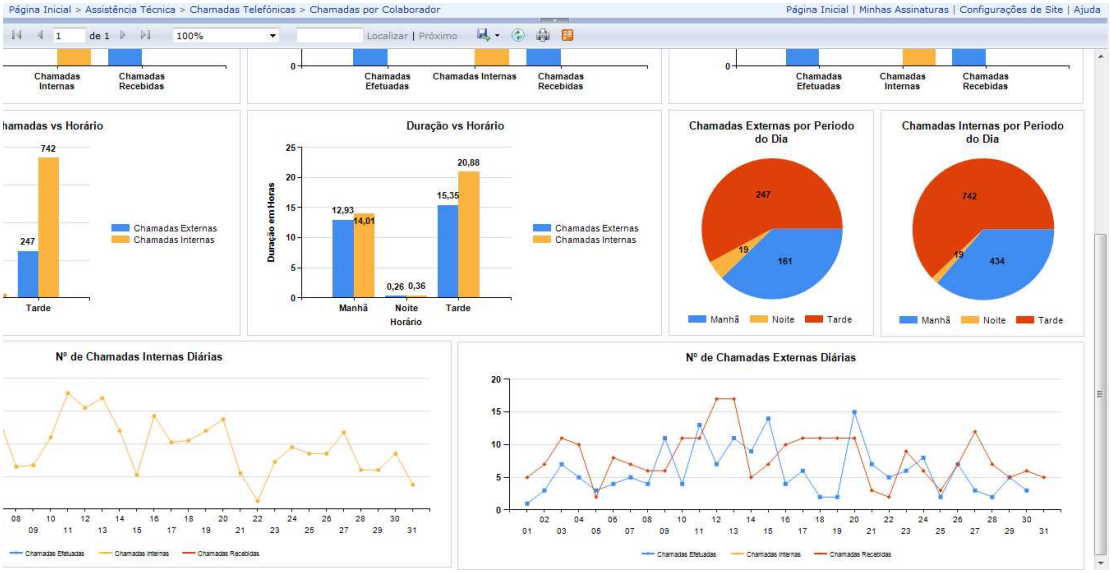


Figura 5.35: Relatório da análise das Chamadas por Colaborador

Através das aplicações de *front-end* procurou-se também apresentar os resultados das análises efetuadas recorrendo ao modelos de *Data Mining* identificados. A Figura 5.36 ilustra, como exemplo, um relatório relativo às Regras de Associação por área de negócio (Intrusão).

Apesar das análises que é possível realizar, algumas limitações são apontadas às aplicações de *front-end*. Estas apesar de permitirem o acesso aos dados e partilhar os mesmos num ambiente *web*, demonstram algumas debilidades em comparação, por exemplo, com outras plataformas de *Business Intelligence* exploradas. Os relatórios

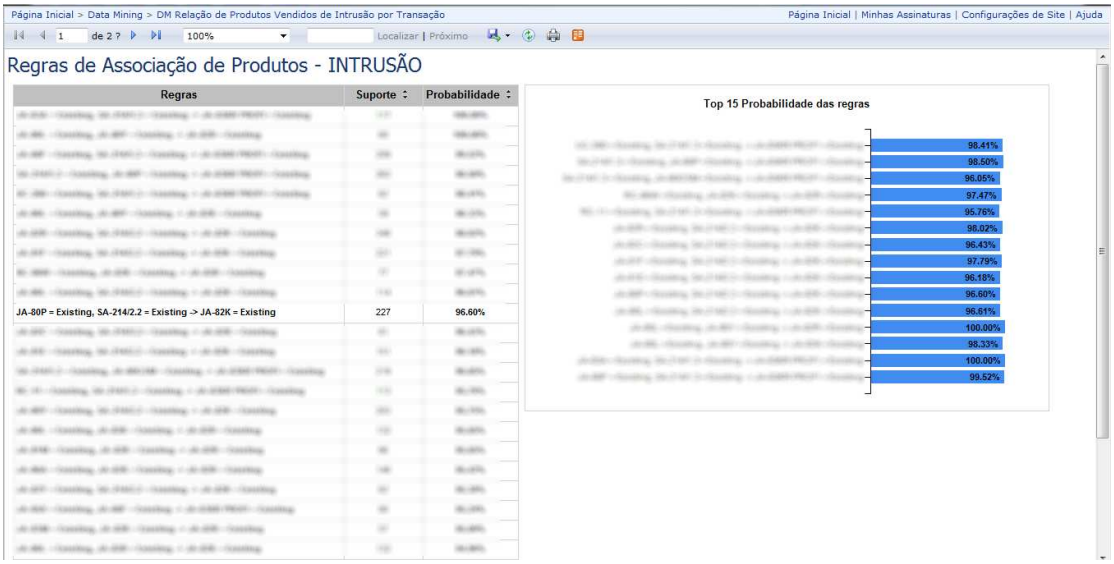


Figura 5.36: Relatório da Regras de Associação por Área de Negócio

são pré-definidos e são demasiado estáticos, limitando a manipulação da informação de forma pouco dinâmica.

Adicionalmente, como plataforma "alternativa" e de exploração da informação de uma forma *ad-hoc*, recorre-se ao *Microsoft Office Excel*, através do qual é possível explorar os cubos OLAP de uma forma mais dinâmica, possibilitando o cruzamento de dados entre as diversas dimensões disponíveis, filtrar dados a partir dessas mesmas dimensões e a monitorização dos diversos KPI. A Figura 5.37 apresenta uma análise da MCR e rentabilidade anual por cliente, efetuada através do *Microsoft Office Excel*.

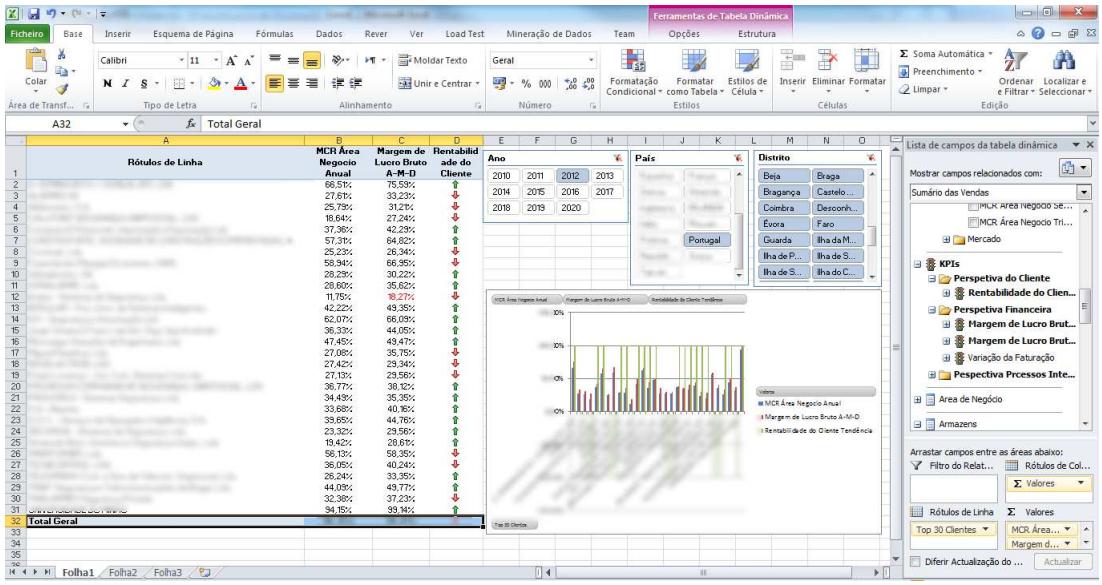


Figura 5.37: Análise MCR e Rentabilidade Anual por Cliente

Por fim é importante referir que a segurança de acesso à plataforma de *Business Intelligence* é concretizada com base no *Active Directory*[2], que foi integrado no servidor e que permite a gestão das contas de utilizador. Apenas os gestores e o administrador de sistemas têm acesso ao sistema implementado, salvaguardando e preservando desta forma a segurança da informação, crucial para a gestão estratégica da organização.

[2] *Http://www.microsoft.com/en-us/download/details.aspx?displaylang=en&id=14683.*

Capítulo 6

Conclusões

Este capítulo descreve o trabalho realizado, apresentando uma síntese dos resultados e das contribuições obtidas, atendendo à finalidade e aos objetivos que foram inicialmente propostos. Tendo como objetivo dar continuidade ao trabalho aqui iniciado são também identificados e descritos projetos de trabalho futuro.

6.1 Síntese

Na busca à resposta da questão de investigação, o objetivo principal deste projeto de dissertação consistiu na conceção e implementação de um sistema de *Business Intelligence* que suportasse a gestão estratégica de uma organização que atua na área da Engenharia Eletrónica Industrial, no ramo do comércio e distribuição de equipamentos eletrónicos por grosso. Os objetivos base para a sua concretização contribuem para a resolução de um problema real de uma organização.

Os objetivos inicialmente propostos para a concretização do projeto de dissertação assentaram no estudo da solução que melhor se enquadrava à organização, bem como a definição de uma arquitetura para o sistema de *Business Intelligence* a implementar. Foi necessário a definição e implementação de um modelo multidimensional que servisse de suporte ao armazenamento dos dados e dos principais KPI do negócio. Outro objetivo foi a definição de um processo de ETL automatizado, que permitisse a extração e transformação dos dados, das diversas fontes de dados, e os canalizasse para o *Data Warehouse* implementado. Foi definido e implementado um sistema de processamento analítico de dados, e identificados modelos de análise dos dados recorrendo a técnicas avançadas de *Data Mining*. Por último concretizou-se uma aplicação de *front-end* que permitisse o acesso, a exploração e a visualização da informação. É importante referir que não fazia parte deste trabalho a análise dos dados, na ótica do utilizador.

A revisão de literatura permitiu identificar os principais conceitos e tecnologias em que o *Business Intelligence* se encontra alicerçado, contribuindo desta forma para uma maior tomada de consciência, sobre o mesmo, por parte das organizações que ponderem a sua adoção.

O projeto desenvolvido foi realizado na organização alvo do estudo. Para facilitar o seu entendimento e auxiliar a sua condução, recorreu-se à metodologia de ciclo de vida dimensional do negócio.

Os requisitos para a implementação de um sistema de *Business Intelligence* são alicerçados pelas necessidades de informação requeridas pela organização. Estas vão de encontro a quatro segmentos de análise: Mercado, Negócio, de Responsabilidade e dos Processos. Os KPI inerentes a cada um dos segmentos foram identificados, bem como as restantes análises e KPI que frequentemente a organização realiza para suportar a tomada de decisão.

Face às necessidades especificadas efetuou-se uma análise detalhada das fontes de dados. Os dados são provenientes dos Sistemas de Informação da organização. A estrutura de dados que suporta o sistema ERP apresenta uma arquitetura complexa. Esta complexidade exigiu um maior esforço de análise e implicou que fossem tomadas medidas para controlar tal complexidade. A criação de *views* representa, no caso concreto da organização, um subconjunto de dados referentes ao negócio existente e que residem na base de dados dos sistemas OLTP.

Atendendo à identificação das necessidades e à análise dos sistemas OLTP, propôs-se, com base na revisão de literatura realizada, uma arquitetura tecnológica de suporte ao sistema de *Business Intelligence* desenvolvido.

O modelo de dados do *Data Warehouse* concebido baseia-se num esquema em constelação, constituído por sete tabelas de facto e trinta tabelas de dimensão. Tal modelo suporta os principais KPI do negócio da organização. As tabelas de facto (FactVendas, FactCompras, FactContaCorrenteClientes, FactMovimentosStock, FactRMA, FactChamadasInternas e FactChamadasExternas) encontram-se ligadas a diversas tabelas de dimensão que permitem analisar os dados sob diversas perspetivas.

Para concretizar o armazenamento dos dados no *Data Warehouse*, implementou-se um processo de ETL que, diariamente, permite extrair e transformar os novos dados do sistema OLTP e os canalize no *Data Warehouse*, de uma forma automatizada.

O sistema de processamento analítico dos dados (sistema OLAP) permitiu acesso e exploração do *Data Warehouse*. Foram implementados oito cubos OLAP para cada uma das tabelas de facto e respetivas tabelas de dimensão. Em cada um deles foram

implementados os diversos KPI que suportam a gestão estratégica da organização.

O algoritmo utilizado na análise de *Data Mining*, para identificação de modelos de associação, tem como base o objetivo de análise. Para um problema de associação, encontrar relações nos dados das vendas sobre os produtos vendidos, identificando quais aqueles que frequentemente são vendidos em conjunto, quer seja no global das vendas, por cliente ou mesmo por área de negócio foi utilizada a técnica de *Microsoft Association Rules*.

As aplicações de *front-end* concretizadas permitem o acesso à informação de uma forma rápida e intuitiva. A plataforma *web* (*Microsoft Reporting Services*) permite aos gestores da organização aceder a relatórios no âmbito das análises requeridas. Por sua vez, a necessidade de explorar a informação, de uma forma mais dinâmica e de forma *ad-hoc* requer a utilização da componente *Microsoft Office Excel*.

6.2 Contribuições

Os sistemas *Business Intelligence* orientados à tomada de decisão surgem como um processo de recolha, tratamento, análise e utilização de informação estratégica para as organizações. A área da distribuição, bem como diversas áreas (Chee, et al., 2009) onde os sistemas de *Business Intelligence* são normalmente aplicados, mostra necessidades em combinar a recolha de dados de forma rápida, através de ferramentas de análise, com o principal objetivo de disponibilizar informação precisa para apoiar a tomada de decisão.

O projeto desenvolvido reflete um estudo novo, sem antecedentes, para a organização alvo do estudo. A introdução dos conceitos de *Business Intelligence* na organização para suportar a gestão estratégica torna-se numa mais-valia para a mesma, pois visa uma plataforma de recolha, tratamento, armazenamento e processamento dados, e de exploração de informação útil para a organização. O sistema desenvolvido possibilita o acesso, em tempo útil, à informação que vai suportar a tomada de decisão na organização. Até ao momento, para analisar informação e tomar decisões, seria necessário aceder diretamente às bases de dados do sistema OLTP, agora e através do sistema desenvolvido a informação encontra-se num único ambiente e de fácil acesso.

Apenas com informação acerca do negócio, é possível tomar as melhores decisões e garantir uma melhor gestão da organização. O valor do sistema de *Business Intelligence* para o negócio é predominante, conseguindo atingir todos os níveis da organização, proporcionando um recurso capaz de suportar a tomada de decisão, sendo expresso pelo facto de disponibilizar informação que pode servir como base

para a realização de mudanças significativas.

O projeto desenvolvido mostrou que, com a implementação de sistemas de *Business Intelligence* nas organizações, torna-se mais fácil extrair informação útil através dos dados que se encontram dispersos numa organização, transformando-os e integrando-os num *Data Warehouse*.

É importante referir que este projeto foi o primeiro trabalho desenvolvido, pelo autor, na área do *Business Intelligence*. O facto de ser desenvolvido num ambiente organizacional e dada a sua abrangência, considera-se que a participação neste, apesar das dificuldades, revelou ser uma mais-valia na sua formação, enriquecendo-o ao nível pessoal e profissional. Relativamente ao trabalho futuro, referido na secção 6.3, existe a possibilidade de dar seguimento ao trabalho até aqui iniciado.

Por último importa realçar que o este trabalho deu origem a uma publicação científica em conferência de âmbito nacional: - Sérgio Costa e Maribel Y. Santos. Sistema de *Business Intelligence* no suporte à Gestão Estratégica: Caso prático no comércio de equipamentos eletrónicos. *Atas da Conferência Anual da Associação Portuguesa de Sistemas de Informação - CAPSI'2012*, Guimarães, Portugal, Setembro, 2012[1].

6.3 Trabalho Futuro

O trabalho desenvolvido, apesar de consistir num sistema funcional e que se adapta às necessidades especificadas pela organização para a introdução dos conceitos associados aos sistemas de *Business Intelligence* no suporte à gestão estratégica, tem ainda espaço para diversas melhorias, bem como para superar algumas limitações atualmente existentes. Destaca-se, desta forma, um conjunto de pontos a ter em consideração no desenvolvimento de projetos de trabalho futuro que visem dar continuidade ao trabalho aqui iniciado.

- Definição e implementação de novos KPI que suportem a gestão estratégica da organização.

- Estender o modelo do *Data Warehouse,* acrescentando novas tabelas de facto e de dimensões, de forma a abranger os diversos processos organizacionais, bem como a construção do respetivo processo de ETL e a construção dos cubos OLAP. Está previsto a inclusão da conta corrente dos fornecedores, dos empréstimos de equipamentos, das encomendas de clientes e do sistema de *tickets.*

[1] *http://www3.dsi.uminho.pt/CAPSI2012/CD/submissions/capsi2012_submission_16.pdf.*

- Integrar novas ferramentas para criar os *Dashboards*, visto que as ferramentas utilizadas apresentam algumas limitações a este nível. Uma solução pode passar pela integração do sistema implementado com a plataforma *Microsoft Sharepoint Server* para potencializar o mesmo, proporcionando novas formas de visualização e exploração dos dados, de uma forma mais dinâmica, recorrendo ao *Microsoft Performance Point* e *Power View*, o que vai permitir tirar maior partido do sistema OLAP e dos modelos de *Data Mining*.

- Exploração de novas técnicas de *Data Mining*, abrangendo todos os processos de negócio para retirar mais informação dos dados fornecidos e consequentemente melhorar a informação disponibilizada.

Referências

Al-Debei, M. M. (2011). Data Warehouse as a Backbone for Business Intelligence: Issues and Challenges. *European Journal of Economics, Finance and Administrative Sciences*, 153-166.

Ariyachandra, T., & Watson, H. J. (2006). Key Factors in Selecting a Data Warehouse Architecture. *Business Intelligence Journal*, 10(2).

Ariyachandra, T., & Watson, H. J. (2011). Which Data Warehouse Architectures Most Successful?. *Business Intelligence Journal. Vol. 11, No. 1*, 4-6.

Ariyachandraa, T., & Watson, H. (2010). Key organizational factors in data warehouse architecture selection. *Decision Support Systems 49*, 200–212.

Berry, M., & Linoff, G. (2004). *Data Mining Techniques for Marketing, Sales and Customer Relationship Management (2 ed.)*. USA: Wiley Publishing, Inc.

Bing, H., & Ye-bai, L. (2009). Research and Application of Association Rules Methods in Data Mining For Commercial Sales Analysis. *2009 International Conference on Networking and Digital Society*, 183-185.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., et al. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. USA: SPSS Inc.

Chaudhuri, S., & Narasayya, V. (2011). New Frontiers in Business Intelligence. *The 37th International Conference on Very Large* (pp. 1052-1053). Seattle, Washington: Proceedings of the VLDB Endowment.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An Overview of Business Intelligence Technology. *Communications of the ACM*, 88-98.

Chee, T., Chan, L.-K., Chuah, M.-H., Tan, C.-S., Wong, S.-F., & Yeoh, W. (2009). Business Intelligence Systems: State-of-art review and contemporary applications. *Symposium on Progress in Information & Communication Technology*, 96-91.

Crnkovic, G. D. (2010). Constructive Research and Info-Computational Knowledge Generation. (M. Lorenzo, Ed.) *Springer Berlin / Heidelberg, 314*, 359-380.

Dayal, U., & Chaudhuri, S. (1997). An overview of Data Warehousing and OLAP technology. *ACM SIGMOD Record, Volume 26*, 65-74.

Dayal, U., Castellanos, M., Simitsis, A., & Wilkinson, K. (2009). Data Integration Flows for Business Intelligence. *In Proceedings of the International Conference on Extending DatabaseTechnology (EDBT)*, Saint Petersburg, Russia.

Fang, X., & Junwu, X. (2011). The Practice and Application of BI. *2011 International Conference on Internet Computing and Information Services*, 246-248.

Fayyad, U., Piatetsky-Shapiro, G., S. P., & Uthurasamy, R. (1996). Advances in Knowledge Discovery and Data Mining. *AAAI Press/MIT Press*.

Fitriana, R., Eriyatno, & Djatna, T. (2011). Progress in Business Intelligence System research: A literature Review. *International Journal of Basic & Applied Sciences IJBAS-IJENS Vol: 11 No: 03*, 96-105.

Fouché, G., & Langit, L. (2011). *Foundations of SQL Server 2008 R2 Business Intelligence*. Apress.

Golfarelli, M., Maio, D., & Rizzi, S. (1998). Conceptual Design of Data Warehouses from E/R Schemes. *Published in the Proceedings of the Hawaii International Conference On System Sciences*, IEEE.

Habul, A., & Pilav-Velic, A. (2010). Business Intelligence and Customer Relationship Management. *Proceedings of the ITI 2010 32nd Int. Conf. on Information Technology Interfaces (pp. 169-174)*. Cavtat, Croatia: IEEE.

Han, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques (3 ed.)*. USA: Morgan Kaufmann Publishers.

Inmon, W. H. (2005). *Building the Data Warehouse*. New York: Wiley.

Jourdan, Z., Raine, R. K., & E. Marshall, T. (2008). Business Intelligence: An Analysis of the Literature. *Information Systems Management*, 121-131.

Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit : the complete guide to dimensional modeling (2 ed.)*. John Wiley and Sons, Inc.

Kimball, R., Reeves, L., & Thornthwaite, W. (1998). T*he Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses*. John Wiley & Sons.

Larson, B. (2009). *Microsoft SQL Server 2008: Delivering Business Intelligence*. United States: The McGraw-Hill Companies.

Laursen, G., & Thorlund, J. (2010). *Business Analytics for Managers: Taking Business Intelligence Beyond Reporting*. Canada: JohnWiley & Sons.

Levene, M., & Loizou, G. (2002). Why is the snowflake schema a good data warehouse design? *Information Systems 28. Elsevier Science*, 225-240.

Maimon, O. Z., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook (2 ed.)*. New York, USA: Springer Science Business Media, Inc.

Nguyen, T. M., Tjoa, A. M., Nemec, J., & Windisch, M. (2006). An approach towards an event-fed solution for slowly changing dimensions in data warehouses with a detailed case study. *Data & Knowledge Engineering 63 (2007) Elsevier*, 26-43.

Olszak, C. M., & Ziemba, E. (2006). Business Intelligence Systems in the Holistic Infrastructure Development Supporting Decision-Making in Organisations. *Interdisciplinary Journal of Information, Knowledge, and Management*, 46-58.

Olszak, C., & Ziemba, &. E. (2007). Approach to Building and Implementing Business Intelligence Systems. *Interdisciplinary Journal of Information, Knowledge, and Management 2*, 135-148.

Osadník, P., & Landryová, L. (2011). Principles of Key Performance Indicators for Small and Medium Enterprise in European Union. *12th International Carpathian Control Conference (ICCC)*, 275-279.

Panian, Z. (2007). Just-in-Time Business Intelligence and Real-Time Decisioning . *International Journal of Applied Mathematics and Informatics*, 28-35.

Polsku. (2009). *OLTP vs. OLAP.* Obtido de Datawarehouse4u: Http:// datawarehouse4u.info/OLTP-vs-OLAP.html. Último acesso: Setembro de 2012.

QlikView. (2012). *Sales Management and Customer Analysis.* Obtido de http://eu.demo.qlikview.com/detail.aspx?appName=Sales%2 0Management%20and%20Customer%20Analysis.qvw. Último acesso: Outubro de 2012.

Rahm, E., & Do, H. H. (2000). Data Cleaning: Problems and Current Approaches. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 23(4).

Rainardi, V. (2008). *Building a Data Warehouse: With Examples in SQL Server.* United States of America: Apress.

Sá, J. V. (2009). *Metodologia de Sistemas de Data Warehouse.* Tese de Doutoramento. Guimarães: Universidade do Minho.

Santos, A. J. (2008). *Gestão Estratégica: Conceitos, Modelos e Instrumentos.* Lisboa, Portugal: Escolar Editora.

Santos, M. F., & Azevedo, C. (2005). *Data Mining e Descoberta de Conhecimento em Base de Dados.* Lisboa, Portugal: FCA.

Santos, M. Y., & Ramos, I. (2009). *Business Intelligence - Tecnologias da Informação na Gestão de Conhecimento (2 ed.).* Lisboa, Portugal: FCA.

Teresa, M., & Bernardino, J. (2011). Open Source Business Intelligence Tools for SMEs. *2011 6th Iberian Conference on IEEE Information Systems and Technologies (CISTI)*, 1-4.

Turban, E., Sharda, R., & Delen, D. (2010). *Decision support and business intelligence systems (9 ed.).* Upper Saddle River, NJ, USA: Prentice Hall Press.

Vaishnavi, V., & Kuechler, W. (Dezembro de 2011). *Design Research in Information Systems.* Obtido de Desrist: Http://desrist.org/ design-research-in-information-systems.

Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2002). Conceptual Modeling for ETL Processes. *Paper presented at the 5th ACM international workshop on Data Warehousing and OLAP. Virginia, USA.*

Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making.* Politecnico di Milano, Italy: A John Wiley and Sons, Ltd., Publication.

Zhao, L., & Huang, X. (2009). Research on the Application of Business Intelligence in Logistics Management. *International Conference on Management and Service Science*, 1-4.

Bibliografia

Crónica, I. d. (2011). *Implementação de um Sistema de Business Intelligence para análise da Doença Pulmonar Obstrutiva Crónica*. Universidade do Minho: Tese de Mestrado. Engenharia e Gestão de Sistemas de Informação.

Eckerson, W. (2003). Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligence Solutions. *TDWI Report Series*.

Farooq, F., & Sarwar, S. M. (2010). Real-Time Data Warehousing For Business Intelligence. *Proceedings of the 8th International Conference on Frontiers of Information Technology ACM*.

Fries, J. (2006). *The Contribution of Business Intelligence to Strategic*. Economische Sociale en Politeke Wetenschappen en Solvay Business School.

Gardner, S. R. (1998). Building the Data Warehouse. *Communications of the ACM*, 52-60.

Guerreiro, E. A. (2008). *Canalização e Visualização de dados em Data Warehouse para Call Centers*. Universidade Técnica de Lisboa: Tese de Mestrado. Instituto Superior Técnico.

Gupta, B., Moorthy, M. N., & Babu, M. R. (2011). Analyzing Data Mining Algorithms in SQL Server. I*nternational Journal of Research and Reviews in Computer Science (IJRRCS)*, 670-675.

Imhoff, C., Galemmo, N., & Geiger, J. G. (2003). *Mastering Data Warehouse Design Relational and Dimensional Techniques*. Canada: Wiley Publishing, Inc., Indianapolis, Indiana.

Krneta, D., Radosav, D., & Radulovic, B. (2008). Realization Business Intelligence in Commerce using Microsoft Business Intelligence. *6th International Symposium on Intelligent Systems and Informatics,*

2008.

Lirong Wang, Q. G., & Ranran-Fu. (2010). Application research of Data warehouse technology in decision-making of drug distribution enterprise . *2010 2nd International Conference on Industrial Mechatronics and Automation*, 202-205.

Mintzberg, H. (1994). *The Rise and Fall of Strategic Planning*. United Kingdom: Prentice Hall International (5-34).

Monteiro, R. A. (2009). *Business Intelligence Aplicação Prática no Sector Financeiro*. Universidade Nova de Lisboa: Tese de Mestrado. Instituto Superior de Estatística e Gestão da Informação.

Mundy, J., Thornthwaite, W., & Kimball, R. (2011). *The Microsoft Data Warehouse Toolkit: With SQL Server 2008 R2 and the Microsoft Business Intelligence Toolset, Second Edition*. Canada: Wiley Publishing, Inc.

Ranjan, J. (2006). Business Intelligence: Concepts, Components, Techniques and Benefits . *Journal of Theoretical and Applied Information Technology*, 60-70.

Ren, Z. (2009). Study on Building Data Mining Application. *5th International Conference on Wireless Communications, Networking and Mobile Computing, 2009*, 1-4.

Santos, M. Y., & Ramos, I. (2006). Como tornar o seu negócio realmente competitivo – Desafios tecnológicos e de gestão. CXO: *Tecnologias de Informação para Executivos*, 56-61.

Seify, M. (2010). Importance of KPI in BI system, Case study: Iranian industries. *2010 Seventh International Conference on Information Technology*, 1245-1246.

Sezões, C., Oliveira, J., & Baptista, M. (2006). Business Intelligence. *SPI – Sociedade Portuguesa de Inovação*.

Silva, C. I. (2011). *Protótipo de Data Warehouse a aplicar numa Escola Secundária/3 de Ensino*. Instituto Politécnico de Castelo Branco: Tese de Mestrado. Escola Superior de Tecnologia.

Singh, M., Singh, P., & Suman. (2007). Conceptual Multidimensional Model. *World Academy of Science, Engineering and Technology 36*, 232-237.

Tennick, A., & Sarka, D. (2011). *Practical SQL Queries for Microsoft SQL Server 2008 R2*. United States: McGraw-Hill Companies.

Wang, Q., Xi, L., & Gao, K. (2009). Application of Business Intelligence in the information development of Construction. *2009 Fifth International Conference on Natural Computation*, 212-215.

Watson, H. J., & Wixom, B. H. (2007). The Current State of Business Intelligence. *IEEE*, 96-99.

Zhai, D., & He, W. (2010). An Application of Business Intelligence Based on Patent in Data Integration and Analysis. *2010 International Conference on Web Information Systems and Mining*, 288-292.

Zhenyuan, W., & Haiyan, H. (2010). OLAP Technology and Its Business Application. *2010 Second WRI Global Congress on Intelligent Systems*, 92-95.

Anexos

Anexo A - Especificação dos Atributos da Tabela Produtos

Atributo	Descrição	Tipo	Keys	Domínio	Exemplo
STstamp	Identificador único de uma linha	Char (25)	--	--	"ADMA21911436EBF414BADD539"
CodProduto	Identificador único de um produto - código	Char (40)	PK	--	"02070401000055"
Designação	Descrição do Produto	Char (60)	--	--	~~Confidencial~~
SuperGrupo	Descrição do super grupo do produto	Varchar (81)	--	--	~~Confidencial~~
Grupo	Descrição do grupo do produto	Varchar (81)	--	--	~~Confidencial~~
Familia	Descrição da família do produto	Varchar (81)	--	--	~~Confidencial~~
SubFamilia	Descrição da sub família do produto	Varchar (81)	--	--	~~Confidencial~~
Serviços	Se o produto é ou não um serviço	Bit	--	[0;1]	"1"
Stock	*Stock* do produto em armazém	Numeric (13,3)	--	[-496,00;486998,0]	"9120,00"
PVR	Preço revenda (preço para o cliente)	Numeric (19,6)	--	[0;12556,10]	"5565,30"
PVD	Preço de distribuidor	Numeric (19,6)	--	[0;10918,40]	"4565.90"
PVA	Preço de agente/mordomos	Numeric (19,6)	--	[0,1937,90]	"613,40"
PVSD	Preço de sub distribuidor	Numeric (19,6)	--	[-3,00;2007,00]	"612,900"
PVP	Preço de venda ao público (aconselhado ao publico)	Numeric (19,6)	--	[0;117620,00]	"501.80"
Tabiva	Identificador único da taxa de iva	Numeric (2,0)	FK	[0;9]	"3"
Ecopilhas	Valor do eco pilhas do produto	Numeric (19,6)	--	[0;3,3390]	"1,4310"
Ecoreee	Valor do ecoreee do produto	Numeric(19,6)	--	[0;6,050]	"6,050"
Ecovalor	Eco valor do produto (Ecopilhas +	Numeric(19,6)	--	[0;6,050]	"6,050"

Figura 6.1: Especificação dos atributos da tabela BI_Produtos

Anexo B - Caracterização dos Erros da Tabela Produtos

Designação Atributo	Dados em Falta / Erros nos dados	Correções a Efetuar
STstamp	--	--
CodProduto	36 Registos em branco (serviços)	(Eliminar ou não considerar, visto que os produtos em questão não tem grande relevância para o negócio)
Designação	--	--
SuperGrupo	100 Registos "null" (não existe supergrupo associado ao produto (94-> serviços))	Criar SuperGrupo "00-Desconhecido"
Grupo	100 Registos "null" (não existe grupo associado ao produto (94-> serviços))	Substituir por "0000-Desconhecido"
Família	100 Registos "null" (não existe família associado ao produto (94-> serviços))	Substituir por "000000-Desconhecido"
Subfamília	100 Registos "null" (não existe subfamília associado ao produto (94-> serviços))	Substituir por "00000000-Desconhecido"
Serviços	--	--
Stock	Apresenta alguns valores negativos	--
CodFornecedor	175 Registos sem código de fornecedor	
PVR	--	--
PVD	--	--
PVA	--	--
PVSD	2 Registos com valores negativos ("-3,00 e -1,00")	
PVP	--	--
Tabiva	--	--
Ecopilhas	--	--
Ecoreee	--	--
Ecovalor	--	--
Inactivo	--	--
Pentrega	1 Registo com valor "600"	Substituir 600 -> 60
Epcusto	--	--
Epcpond	--	--

Figura 6.2: Caracterização dos erros da tabela BI_Produtos

Anexo C - Criação da Tabela DimHora

```
DECLARE @hora TIME

SET NOCOUNT ON
SET @hora = '00:00:00'

CREATE TABLE dbo.DIM_HORA
    (
    DIM_HORA_PK [int] IDENTITY(1,1) PRIMARY KEY,
    TEMPO time(7) NOT NULL,
    HORA tinyint NOT NULL,
    MINUTO tinyint NOT NULL,
    SEGUNDO tinyint NOT NULL,
    HORÁRIO VARCHAR(10) NOT NULL)

WHILE @hora <> '23:59:59'
BEGIN
    INSERT INTO DIM_HORA
    VALUES(
            @hora,
            DATEPART(HOUR,@hora),
            DATEPART(MINUTE,@hora),
            DATEPART(SECOND,@hora),
            CASE
                WHEN Datepart(HOUR,@hora)>=8 AND Datepart(HOUR,@hora)<12 THEN 'MANHÃ'
                WHEN Datepart(HOUR,@hora)>=12 AND Datepart(HOUR,@hora)<20 THEN 'TARDE'
                WHEN Datepart(HOUR,@hora)>=20 AND Datepart(HOUR,@hora)<23 THEN 'NOITE'
                WHEN Datepart(HOUR,@hora)=23 THEN 'NOITE'
                WHEN Datepart(HOUR,@hora)=0 THEN 'NOITE'
                WHEN Datepart(HOUR,@hora)>=1 AND Datepart(HOUR,@hora)<8 THEN 'NOITE'
            ELSE 'ERRO'
            END
            )
    SELECT @hora = DATEADD(SECOND,1,@hora) END
```

Figura 6.3: Criação da DimHora

Anexo D - Criação da Tabela DimTempo

```
DECLARE @OrderDate DATETIME
DECLARE @OrderFim DATETIME
DECLARE @Mes as INT

SET @OrderDate = '2001001/01'
SET @OrderFim = '2020/12/31' -- '2010/01/01' -- '2010/11/02'


WHILE @Orderdate <= @Orderfim
BEGIN
Set @Mes = DATEPART(MONTH,@OrderDate)
 INSERT INTO dbo.DIM_TEMPO (
    DATA,
    ANO,
    SEMESTRE,
    TRIMESTRE,
    MES,
    QUINZENA,
    SEMANA,
    DIA,
    ESTACAO_ANO
)
VALUES
(
@Orderdate,
convert(int,DATEPART(yyyy,@Orderdate)),
Case
  When Datepart(mm,@Orderdate)<7 THEN '1º Semestre'
  ELSE '2º Semestre'
END,
Case
     When Datepart(mm,@Orderdate)<4 THEN '1º Trimestre'
     When Datepart(mm,@Orderdate)>3 and Datepart(mm,@Orderdate)<7 THEN '2º Trimestre'
     When Datepart(mm,@Orderdate)>6 and Datepart(mm,@Orderdate)<10 THEN '3º Trimestre'
     When Datepart(mm,@Orderdate)>9 THEN '4º Trimestre'
     ELSE 'ERRO'
END,
```

Figura 6.4: Criação da DimTempo (1)

```
END, Case
        when @Mes = 1 then '01-Janeiro'
        when @Mes = 2 then '02-Fevereiro'
        when @Mes = 3 then '03-Março'
        when @Mes = 4 then '04-Abril'
        when @Mes = 5 then '05-Maio'
        when @Mes = 6 then '06-Junho'
        when @Mes = 7 then '07-Julho'
        when @Mes = 8 then '08-Agosto'
        when @Mes = 9 then '09-Setembro'
        when @Mes = 10 then '10-Outubro'
        when @Mes = 11 then '11-Novembro'
        when @Mes = 12 then '12-Dezembro'
END, Case
     When DATEPART(dd,@Orderdate) Between 1 and 15 Then '1ª Quinzena'
     Else '2ª Quinzena'
END,
DATEPART(ww,@Orderdate),
Left(CONVERT(VarChar, @Orderdate, 103), 2),
Case
     When Datepart(mm,@Orderdate)=12 and Datepart(dd,@Orderdate)>=21 Then 'Inverno'
     When Datepart(mm,@Orderdate)<3 Then 'Inverno'
     When Datepart(mm,@Orderdate)=3 and Datepart(dd,@Orderdate)<21  Then 'Inverno'

     When Datepart(mm,@Orderdate)=3 and Datepart(dd,@Orderdate)>=21  Then 'Primavera'
     When Datepart(mm,@Orderdate)>3 and Datepart(mm,@Orderdate)<6 Then 'Primavera'
     When Datepart(mm,@Orderdate)=6 and Datepart(dd,@Orderdate)<21  Then 'Primavera'

     When Datepart(mm,@Orderdate)=6 and Datepart(dd,@Orderdate)>=21  Then 'Verao'
     When Datepart(mm,@Orderdate)>6 and Datepart(mm,@Orderdate)<9 Then 'Verao'
     When Datepart(mm,@Orderdate)=9 and Datepart(dd,@Orderdate)<21  Then 'Verao'

     When Datepart(mm,@Orderdate)=9 and Datepart(dd,@Orderdate)>=21  Then 'Outono'
     When Datepart(mm,@Orderdate)>9 and Datepart(mm,@Orderdate)<12 Then 'Outono'
     When Datepart(mm,@Orderdate)=12 and Datepart(dd,@Orderdate)<21  Then 'Outono'
     Else 'ERRO'
END
) SET @Orderdate = @Orderdate + 1 END
```

Figura 6.5: Criação da DimTempo (2)