



Instituto Politécnico de Tomar

**Escola Superior de Tecnologia de Tomar**

# **Implementação de um Sistema de Business Intelligence**

Relatório de Estágio

**Tatiana Marina Gaspar Martins Duarte**

Mestrado em Analítica e Inteligência Organizacional

**Tomar / junho / 2018**



Instituto Politécnico de Tomar

[www.ipt.pt](http://www.ipt.pt)



Instituto Politécnico de Tomar

**Escola Superior de Tecnologia de Tomar**

**Tatiana Marina Gaspar Martins Duarte**

# **Implementação de um Sistema de Business Intelligence**

Relatório de Estágio

Orientado por:

IPT: Professor Doutor Ricardo Campos

Entidade de Estágio: Doutor Nuno Miquelina

Relatório de Estágio apresentado ao Instituto Politécnico de Tomar  
para cumprimento dos requisitos necessários à obtenção do grau de Mestre  
em Análítica e Inteligência Organizacional

“Para vencer - material ou imaterialmente - três coisas definíveis são precisas: saber trabalhar, aproveitar oportunidades, e criar relações. O resto pertence ao elemento indefinível, mas real, a que, à falta de melhor nome, se chama sorte.”

Fernando pessoa



## RESUMO

---

Atualmente os dados são um dos mais importantes e críticos ativos de uma empresa. A sua exploração e análise é uma mais valia no auxílio e no suporte à tomada de decisão. Numa era em que o volume de dados cresce exponencialmente, o acesso a esta informação em tempo útil pode fazer toda a diferença no contexto organizacional. Conhecedores desta nova realidade, as empresas recorrem à implementação de sistemas de *Business Intelligence* como forma de extrair e analisar grandes quantidades de dados, possibilitando assim uma tomada de decisão mais consciente e informada. Este processo, conhecido como gestão do conhecimento, caracteriza-se por ser um conjunto de processos e ferramentas que organizam e sistematizam os dados com o objetivo de os transformar em conhecimento. Com vista a este fim, são utilizadas ferramentas baseadas em Tecnologias de Informação (TI) com o propósito de automatizar os processos inerentes à transformação dos dados em conhecimento. Nesta tese, desenvolvida no âmbito do Mestrado em Analítica e Inteligência Organizacional, e que decorre do estágio protocolado com a Compta Emerging Business (CEB), descrevemos a implementação de um sistema de *Business Intelligence* para uma organização ligada à recolha de resíduos. Primeiramente, analisaram-se os dados provenientes de uma fonte de dados interna de forma a identificar os dados que se enquadravam nas necessidades da organização. No passo seguinte, foi criado um modelo de dados multidimensional para a construção de um *data warehouse (DW)*, para onde são carregados os dados extraídos e transformados no processo ETL (*Extraction, Transformation and Loading*). Por fim, foram elaborados relatórios e *dashboards*, como meio de apoio e suporte à tomada de decisão.

**Palavras-chave:** *Business Intelligence; Data Warehouse; Extraction, Transformation and Loading;* Tecnologias de Informação



## ABSTRACT

---

Data is currently considered one of the most important and critical assets for a company. It's exploration and analysis is without any doubt an added value in the process of decision making. In an era, where the volume of data grows exponentially, accessing useful data at the right time is at the utmost importance for any organization. Knowing this, organizations are doing investments in Business Intelligent systems to extract and analyze huge quantities of data thus enabling more conscious and informed decisions. This process, known as knowledge management, is characterized by a set of processes and tools that organize and systematize the data in order to transform them into knowledge. Based on this, Information Technology (IT) tools are used with the purpose of automating the processes inherent to the transformation of the data into knowledge. In this thesis, developed within the scope of the master's Degree in Analytical and Organizational Intelligence, which stems from the protocol stage with Compta Emerging Business (CEB), we describe the implementation of a Business Intelligence System for a company of waste management. Firstly, the data from an internal source was analyzed in order to understand the needs of the organization. This supports the second stage, where an Extraction, Transformation and Loading process is carried on, into the data warehouse (DW) multidimensional data model, Based on this, reports and dashboard were developed in the final stage, as a means to support the decision-making process.

**Keywords:** Business Intelligence; Data Warehouse; Extraction, Transformation and Loading; Information Technology



## **AGRADECIMENTOS**

---

Gostaria de agradecer à Compta Emerging Business que aceitou este estágio curricular. Aos meus orientadores Professor Doutor Ricardo Campos, do Instituto Politécnico de Tomar e ao Doutor Nuno Miquelina, da Compta Emerging Business por todo o apoio prestado no decorrer do estágio.

Não poderia de deixar de agradecer ao Doutor Hélder Quintela e ao Paulo Costal, da Compta pelo apoio e suporte.

À minha família.



# Índice

<b>Índice</b>	<b>XI</b>
<b>Índice de Figuras</b>	<b>XIII</b>
<b>Glossário de Abreviaturas e Siglas</b>	<b>XV</b>
<b>1. Introdução</b>	<b>17</b>
1.1. Enquadramento .....	17
1.2. Objetivos do Estágio e Funções Desempenhadas .....	17
1.3. Estrutura do Relatório .....	18
<b>2. Descrição da Entidade Acolhedora</b>	<b>19</b>
2.1. Conta Emerging Business – (CEB) .....	19
2.2. TAGUSVALLEY   Tecnopolo do Vale do Tejo .....	21
<b>3. Fundamentação Teórica</b>	<b>23</b>
3.1. Dados, Informação e Conhecimento.....	23
3.2. Arquitetura de um Sistema de Business Intelligence.....	24
3.2.1. Fonte de Dados .....	24
3.2.2. ETL (Extract, Transformation and Loading).....	25
3.2.3. Repositório de Dados .....	26
3.2.4. Apresentação, Análise e Exploração dos Resultados .....	29
<b>4. Estágio</b>	<b>31</b>
4.1. Fonte de Dados .....	32
4.2. ETL (Extract, Transformation and Loading) .....	32
4.3. Repositório de Dados .....	37
4.4. Apresentação, Análise e Exploração dos Resultados .....	39
<b>5. Conclusões</b>	<b>41</b>
<b>Referências</b>	<b>43</b>



## Índice de Figuras

Figura 1: Certificações Grupo Compta.....	20
Figura 2: Valores Grupo Compta .....	20
Figura 3: Parceiros Grupo Compta.....	20
Figura 4: Componentes de um sistema de Business Intelligence.....	24
Figura 5: Processo ETL .....	26
Figura 6: Arquitetura Data Marts .....	27
Figura 7: Esquema Estrela (Star Schema) .....	28
Figura 8: Esquema Constelação (Constellation Schema).....	28
Figura 9: Esquema Floco de Neve (Snowflake Schema) .....	29
Figura 10: Cubo OLAP .....	29
Figura 11: Workflow do Projeto.....	31
Figura 12: Ligação Fonte de Dados.....	32
Figura 13: Interface Gráfica - Spoon.....	33
Figura 14: ETL – Conexões Data Base (input; output).....	33
Figura 15: ETL – Conexão/Consulta com a Base de Dados Operacional.....	34
Figura 16: ETL - Select Values .....	34
Figura 17: ETL - Value Mapper .....	35
Figura 18: ETL - Conexão com Data Warehouse .....	36
Figura 19: ETL - Transformation Tabela de Factos .....	36
Figura 20: ETL - Criação de um Job .....	37
Figura 21: Modelo de Dados .....	38
Figura 22: Criação das Tabelas Dimensão .....	38
Figura 23: Criação das Tabelas de Factos .....	39
Figura 24: Análise Excel - Power Pivot .....	40
Figura 25: Conexão Data Warehouse e Consulta Dados.....	40



## Glossário de Abreviaturas e Siglas

BI	<i>Business Intelligence</i>
CEB	Compta Emerging Business
DW	<i>Data Warehouse</i>
ETL	<i>Extract, Transforming and Loading</i>
IA	Inteligência Artificial
IPS	Instituto Politécnico de Santarém
IPT	Instituto Politécnico de Tomar
IoT	<i>Internet of Things</i>
NERSANT	Associação Empresarial da Região de Santarém
OLPT	<i>On-line Transaction Processing</i>
OLAP	<i>On-line Analytical Processing</i>
TI	Tecnologias de Informação
SGBDOR	Sistema de Gestão de Base de Dados Objeto-Relacional
SWOT	<i>Strengths, Weaknesses, Opportunities and Threats</i>



# Capítulo 1

## Introdução

Neste primeiro capítulo que dá início ao relatório começamos por fazer uma introdução ao estágio. Na Seção 1.1. é feito um primeiro enquadramento sobre a empresa de acolhimento. Na Seção 1.2 são descritos e listados os objetivos do estágio bem como as funções desempenhadas no seu decurso. Por fim, a Seção 1.3 descreve a estrutura do relatório e dos capítulos que o compõem.

### 1.1. Enquadramento

O presente relatório resulta do estágio curricular realizado no âmbito do Mestrado em Analítica e Inteligência Organizacional do Instituto Politécnico de Tomar que decorreu no período compreendido entre 27 de setembro de 2017 a 16 de fevereiro de 2018, na Compta Emerging Business (CEB)<sup>1</sup>, mais concretamente no Centro de Inovação Tecnológica & Desenvolvimento, sediado em Abrantes na TagusValley – Tecnopolo do Vale do Tejo. A CEB é uma empresa do Grupo Compta<sup>2</sup>, um dos mais antigos Grupos Tecnológicos Nacionais a atuar nas áreas de Telecomunicações e Sistemas de Informação. Com a Lusídeias<sup>3</sup> e três outros projetos (ProdFarme, Sou Cidadão e um projeto de investigação), o Centro de Abrantes conta com uma equipa de seis colaboradores. Um maior detalhe sobre a estrutura da entidade acolhedora, assim como do TagusValley será facultado no Capítulo 2.

### 1.2. Objetivos do Estágio e Funções Desempenhadas

Com o meio envolvente (interno e externo) em constante mutação, as organizações procuram responder de forma rápida e eficaz às ameaças encontradas, com base em informação útil e fiável que lhes permita tomar a melhor decisão. Os sistemas de *Business Intelligence* (BI) proporcionam-lhes isso mesmo através da extração, manipulação e recolha de dados. O contexto deste relatório centra-se na descrição da implementação de

---

<sup>1</sup> Site: [www.ceb-solutions.com](http://www.ceb-solutions.com) (15-05-2018)

<sup>2</sup> Site: [www.compta.pt](http://www.compta.pt) (15-05-2018)

<sup>3</sup> Plataforma Nacional de Inovação onde é possível partilhar ideias, criações ou projetos no âmbito das aplicações móveis.

um sistema de *Business Intelligence* numa organização de recolha de resíduos. Primeiramente foram analisados os dados provenientes do sistema de informação da empresa com o intuito de os consolidar numa *data warehouse (DW)*. Com base nestes dados tornou-se, posteriormente, possível proceder à elaboração de relatórios de suporte à tomada de decisão. Para tal foram utilizadas três ferramentas: *PostgreSQL*; *Pentaho - Data Integration* e *Microsoft Office Excel - PowerPivot*.

Durante o estágio procurou-se compreender a organização do cliente e desenvolver competências técnicas no âmbito das ferramentas utilizadas. A execução destes objetivos, permitiu não só aprofundar os conhecimentos adquiridos no decorrer da formação académica, mas também a obtenção de novos conhecimentos.

### **1.3. Estrutura do Relatório**

O presente documento foi escrito tendo como base o novo acordo ortográfico em vigor desde janeiro de 2009, em Portugal. A estrutura adotada respeita as normas para apresentação e elaboração das Teses de Mestrado, disponibilizado na página *web* do Instituto Politécnico de Tomar. Assim, o relatório encontra-se organizado em cinco capítulos distintos. No capítulo atual - Capítulo 1 - é feita a introdução ao presente relatório. No Capítulo 2, é feita uma apresentação da empresa e das instalações acolhedoras. No Capítulo 3, é feita uma fundamentação teórica sobre os conceitos base de suporte à arquitetura de um sistema de *Business Intelligence*. No Capítulo 4, são apresentadas as tarefas efetuadas na implementação do sistema de BI ao longo do estágio. Por razões de confidencialidade, não é possível, para efeitos do relatório, fazer uso dos dados utilizados no decorrer do projeto. Para contornar este problema recorreremos a dados fictícios. O presente relatório termina com o Capítulo 5, onde são apresentadas as conclusões deste trabalho.

## Capítulo 2

### Descrição da Entidade Acolhedora

O Grupo Compta possui quatro Centros de Inovação Tecnológica & Desenvolvimento (Lisboa, Porto, Abrantes e Évora) e é considerado uma referência no mercado das Tecnologias de Informação (TI), pois os serviços e soluções que tem como oferta para os seus clientes permite-lhes acrescentar mais valor à sua atividade e vantagem competitiva. Neste capítulo, é feita uma descrição da empresa acolhedora do estágio bem como das instalações do Centro de Abrantes (TagusValley - Tecnopolo do Vale do Tejo) onde teve lugar este estágio. Assim, a Secção 2.1 introduz a Compta Emerging Business. Por sua vez, a Secção 2.2 descreve o TagusValley.

#### 2.1. Compta Emerging Business – (CEB)

Fundada em 1972 em Lisboa, a **Compta - Equipamentos e Serviços de Informática**, é um dos Grupos Tecnológicos Nacionais mais antigos, tendo iniciado a sua atividade em 1972 como um Centro de Processamento de Dados<sup>4</sup>.

Em 1987, a empresa foi alvo de uma reorganização com a criação do Grupo Compta. Mais tarde, em 2005, juntou-se a Compta Emerging Business – (CEB), empresa especializada em soluções para Cidades Inteligentes e Indústria nos setores do Meio Ambiente; Agricultura; Logística; Energia; e Mar. Para colmatar as necessidades dos setores acima referidos, a CEB concebeu um conjunto de produtos, Bee2Energy; Bee2Waste; Bee2Lighting; Bee2Green; Bee2Crop; Bee2FireDetetion; Bee2Citizen; e Cargo e-Business que permitem a otimização dos serviços, reduzindo custos e rentabilizando as operações ao mesmo tempo. Hoje, o Grupo Compta, conta com cerca de 400 certificações, ver Figura 1 (as cinco principais certificações), uma equipa com mais de 240 colaboradores especializados, e com uma larga experiência no desenvolvimento de soluções e produtos tecnológicos.

---

<sup>4</sup> Centro de Processamento de Dados ou *Data Center* - local onde estão concentrados os sistemas computacionais de uma organização.



Figura 1: Certificações Grupo Compta

Fonte: Compta

A sua missão passa por “selecionar no mercado mundial as melhores tecnologias adaptando-as aos mercados onde atua” empregando valores, descritos na Figura 2, como o Aprender Sempre; o Espírito de Equipa; o Profissionalismo; e a Competição.



Figura 2: Valores Grupo Compta

Fonte: Compta

A Figura 3 apresenta os parceiros do **Grupo Compta**.

**Temos as alianças certas para construir o sucesso do seu negócio!**



Figura 3: Parceiros Grupo Compta

Fonte: Compta

## **2.2. TAGUSVALLEY | Tecnopolo do Vale do Tejo**

Em 2016, o Grupo Compta abriu um novo Centro de Inovação e Desenvolvimento em Abrantes, com instalações na TagusValley – Tecnopolo do Vale do Tejo. O TagusValley, outrora fábrica de adubos, é um Parque de Ciência e Tecnologia sediado em Abrantes, desde 7 de novembro de 2003, com início de atividade em 2006. A sua conceção é fruto “da aposta da Câmara Municipal de Abrantes (em 2000) em estimular o empreendedorismo e a competitividade na Região, tendo por base a Inovação e a Tecnologia” (TAGUSVALLEY, 2017). Inicialmente, contou com a parceria da Associação Empresarial da Região de Santarém (NERSANT) e o Instituto Politécnico de Tomar (IPT) aos quais mais tarde se juntaram como associados o Instituto Politécnico de Santarém (IPS) e a Tejo Energia.

Voltado para o desenvolvimento regional, através de uma estratégia de implementação de um sistema potenciador de inovação e empreendedorismo, o TagusValley procura, em conjunto com as empresas regionais, identificar novas oportunidades e sinergias, de forma a atrair e estimular recursos humanos qualificados.



## Capítulo 3

### Fundamentação Teórica

Neste capítulo, serão abordados os componentes que fazem parte da arquitetura de um sistema de BI. Assim, a Seção 3.1 faz a distinção entre os conceitos de Dados, Informação e Conhecimento. A Seção 3.2 apresenta a arquitetura de um sistema de BI. Em particular, fazemos referência às diferentes fontes de dados, processo de ETL (*Extract, Transformations and Loading*), repositório de dados e apresentação, análise e exploração dos resultados.

#### 3.1. Dados, Informação e Conhecimento

No contexto de um sistema de BI e, não obstante serem o elemento mais básico nesta cadeia de criação de valor, os dados assumem uma importância vital. Segundo (Alter, 1999), os dados são “fatos, imagens ou sons que podem ou não ser pertinentes ou úteis para uma tarefa particular”. Habitualmente correlacionados com as operações do dia a dia de uma organização (podendo assim ser armazenados na forma de bases de dados transacionais), os dados por si só, não representam, geralmente, relevância, propósito ou significado para a tomada de decisões. Porém, não deixam de ser um fator importante para a criação de informação, uma vez que, de uma análise cuidada sobre os dados se obtêm padrões e tendências, que por sua vez se traduzem em informações relevantes que servem de ponte para a extração do conhecimento. Conhecimento pode definir-se como sendo “um conjunto completo de informações, dados e relações que auxiliam os indivíduos na tomada de decisão, à realização de tarefas e à geração de novas informações e conhecimentos.” (Céci, 2012). Nos últimos anos, e devido ao aumento exponencial do volume de dados armazenados, os sistemas e competências de gestão das organizações têm sido alvo de novos desafios relacionados com a criação de conhecimento específico, tornando-se imprescindível o recurso às TI no processo de transformação dos dados em conhecimento. O *Business Intelligence* surge neste contexto como um importante contributo na recolha, análise e transformação de dados (Antonelli, 2009) ao nível empresarial, proporcionando às organizações, conhecimento útil e oportuno para que seja tomada a decisão mais acertada. A próxima seção detalha a arquitetura de um sistema de BI.

## 3.2. Arquitetura de um Sistema de Business Intelligence

Antes de se implementar um sistema de BI é fundamental que numa organização, se faça, internamente, uma análise SWOT<sup>5</sup> (*Strengths, Weaknesses, Opportunities and Threats*), isto é, uma análise que permita fazer um diagnóstico da organização. Desta forma, será possível identificar as necessidades existentes que irão permitir saber que informação procurar e, posteriormente, recolher entre os dados armazenados.

Um sistema de BI possui quatro grandes componentes: (1) fonte de dados; (2) processo de extração, transformação e carregamento de dados (ETL - *Extraction, Transformation and Loading*); (3) repositório de dados; e (4) área de exploração dos resultados. A Figura 4 ilustra a arquitetura de um sistema de BI. As próximas subsecções apresentam em maior detalhe cada um destes componentes.



Figura 4: Componentes de um sistema de *Business Intelligence*

Fonte: Agência Digital VM2<sup>6</sup>

### 3.2.1. Fonte de Dados

Fonte de dados (*data sources*) é o nome que se dá ao local onde são armazenados os dados produzidos diariamente, dados esses que podem ser caracterizados como estruturados ou não estruturados. Inicialmente, as fontes de dados que alimentavam os sistemas de BI eram baseadas em dados estruturados, ou seja, eram alimentadas por sistemas OLTP (*On-line Transaction Processing*) os quais têm por objetivo registar as tarefas operacionais do dia a dia (registo de clientes, registo de encomendas, etc.). Habitualmente, estas fontes, recorrem a técnicas de relacionamento e normalização dos dados para desta forma

<sup>5</sup> Análise SWOT – Análise que caracteriza o ambiente interno (pontos fortes e fracos) e externo (oportunidades e ameaças)

<sup>6</sup> <http://www.vm2.com.br/bi-business-intelligence> (19-03-2018)

garantir a eficiência do sistema. Por via deste fato, limitam-se a guardar o estado mais atual das transações não mantendo por isso o histórico dos dados. Normalmente, os dados estruturados encontram-se armazenados em base de dados relacionais (*SQL Server*; *PostgreSQL*; *Oracle*; entre outras).

Já os dados não estruturados (e.g., textos, vídeos, imagens, etc) são dados caracterizados pela ausência de uma estrutura explicitamente definida dificultando assim o acesso e a recuperação desses mesmos conteúdos. A massificação no uso da Internet, a proliferação das redes sociais e o surgimento da Internet das Coisas – IoT (do inglês, *Internet of Things*), levou, no entanto, a que nos últimos anos, se assistisse a um aumento no volume de dados criando novos desafios aqueles que de uma forma rápida e eficaz pretendem aceder a uma infinidade de novas fontes, incluindo social media, posts, tweets, rede de sensores, dispositivos móveis, etc. O aumento exponencial do volume de dados, levou ao aparecimento do conceito de *Big data* e dos três v's: volume, velocidade e variedade. Volume, porque tem a capacidade de armazenar grandes volumes de dados semiestruturados e não estruturados; Velocidade, porque permite analisar esses mesmos dados em tempo real; e Variedade porque, os dados armazenados podem ser de vários tipos e formatos. Outros autores consideram também um outro “v” decorrente do “valor” que se obtém a partir da análise dos dados, uma característica frequentemente associada ao espaço temporal, isto é, à obtenção de informação/conhecimento em tempo real. Os sistemas de BI surgem neste contexto.

### **3.2.2. ETL (Extract, Transformation and Loading)**

Uma vez estabelecida a comunicação com as fontes de dados, segue-se o processo ETL (*Extract, Transformation and Loading*) que tem por objetivo garantir a qualidade das informações transferidas para o *data warehouse*. De uma forma genérica, o processo ETL é usado para extrair dados de uma ou mais fontes, transformá-los e carregá-los num repositório de dados. A etapa da transformação é a mais complexa, uma vez que a transformação dos dados tem que ser feita de acordo com as especificidades do negócio. Este processo é por vezes antecedido de uma limpeza dos dados com o objetivo de garantir que a informação se apresenta de forma correta, para não existir a possibilidade de dados incorretos levarem a organização a cometer erros críticos aquando a tomada de decisão. A Figura 5 ilustra o processo de ETL.

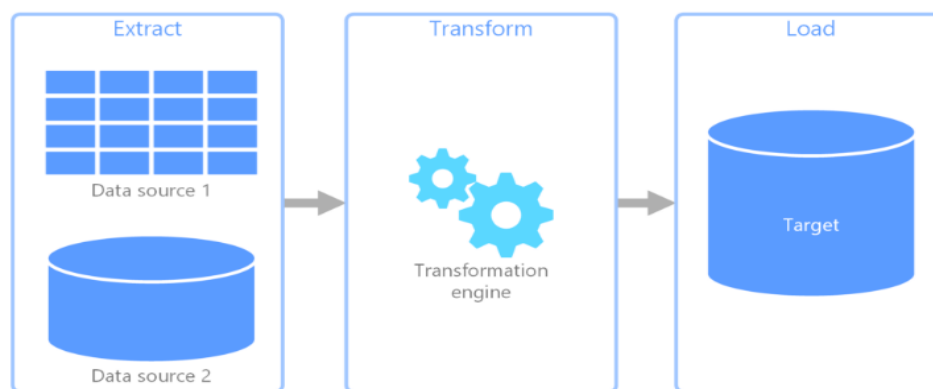


Figura 5: Processo ETL

Fonte: Microsoft Azure

Na próxima subsecção falaremos em maior detalhe sobre o repositório de dados. Posteriormente no Capítulo 4 demonstraremos alguns dos passos de um processo de ETL com recurso à ferramenta *Pentaho – Data Integration*.

### 3.2.3. Repositório de Dados

Depois do processo de ETL os dados são guardados em *Data Warehouses*. Segundo (Ramos da Costa, 2012), Inmon define um *data warehouse* “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”. Contrariamente aos OLTPs, os *Data Warehouses* são, segundo (Céci, 2012), orientados ao processo de negócio (por oposição ao registo de transações), limitados a poucos utilizadores, utilizam muito espaço, guardam dados históricos (por oposição a dados atuais), assentam numa estrutura de dados não normalizada e executam a sua atualização em lote, ou seja, de tempos a tempos (por oposição a uma atualização contínua).

Ao invés de guardar os dados num único *Data Warehouse*, as organizações optam muitas das vezes por recorrer aos *data marts*, um subconjunto de um *data warehouse* em que os dados se referem a um ou mais assuntos específicos, oferecendo assim informação mais detalhada. Existem duas maneiras de implementar um *data mart*, conforme Figura 6: a primeira através de *data marts* dependentes em que os dados são carregados diretamente do *data warehouse* para os *data marts* correspondentes; a segunda através de *data marts* independentes que não proporcionam uma visão integra dos dados por estes serem carregados diretamente dos sistemas OLTP. Este tipo de *data marts* apenas integra a informação relevante para uma área específica do negócio, ao contrário, dos *data marts* dependentes.

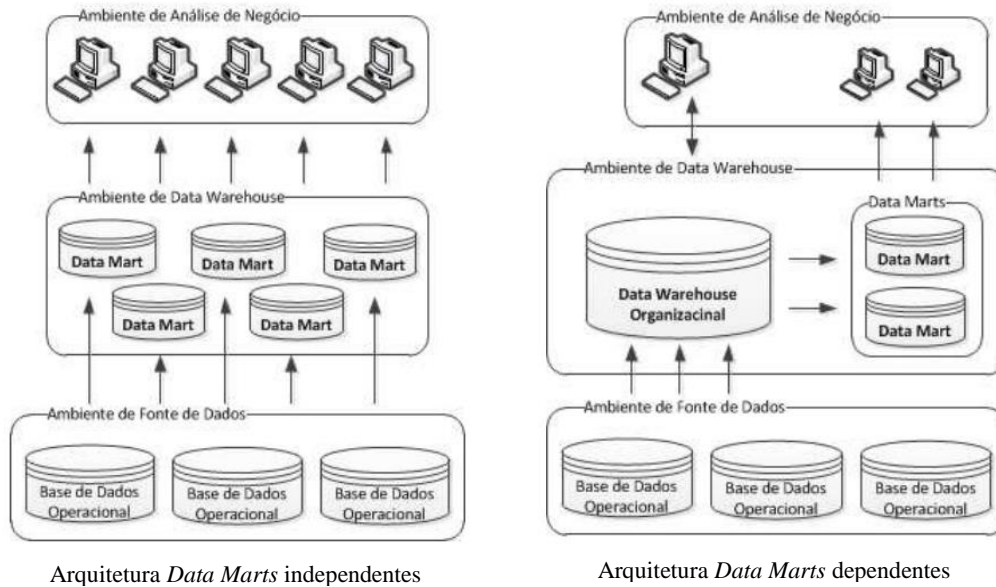


Figura 6: Arquitetura *Data Marts*

Fonte: (Ramos da Costa, 2012)

A implementação de *data warehouse* obedece normalmente ao esquema Estrela (*Star Schema*) ou ao esquema Floco de Neve (*Snowflake Schema*). O esquema Estrela é o mais utilizado para modelar um *data warehouse*, uma vez que os dados são modelados em tabelas dimensionais que estão ligadas a uma única tabela de Factos, o que torna simples a compreensão da relação entre os dados. Uma particularidade, deste tipo de esquemas, é o facto de se poderem converter num esquema designado como esquema Constelação (*Constellation Star*) onde várias tabelas de factos partilham uma ou mais tabelas dimensão. A tabela de Factos representa um determinado assunto, o que se pretende analisar (e.g. vendas, compras). Nela, “cada registo integra um conjunto de colunas (chaves estrangeiras) que a relacionam com as tabelas de dimensão e colunas que retratam as medidas e/ou indicadores de análise” (Ramos da Costa, 2012). A tabela Dimensão, por sua vez, representa os registos descritivos pelos quais a tabela de factos pode ser analisada (e.g., dimensão tempo, a dimensão produto, a dimensão clientes). Para cada registo em ambas as tabelas é gerada uma chave primária. A Figura 7 exemplifica o formato de um esquema Estrela e a Figura 8 o formato de um esquema Constelação.

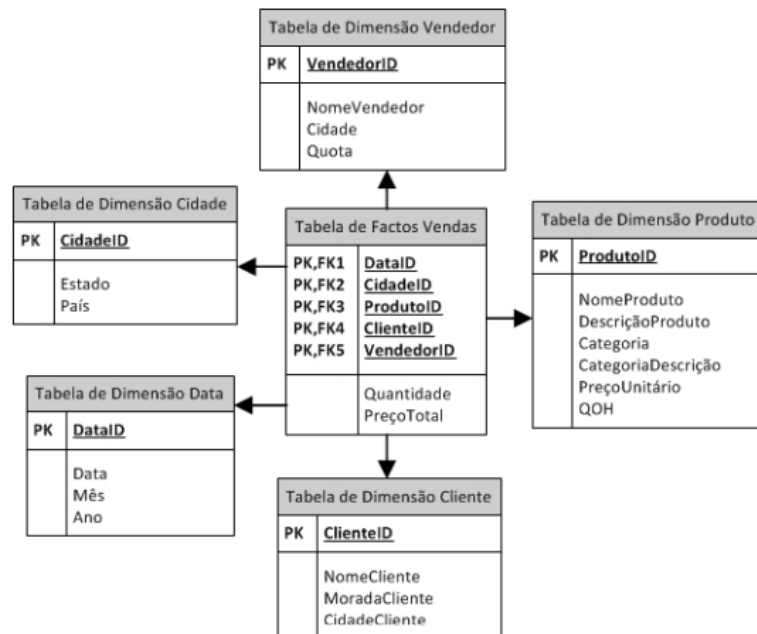


Figura 7: Esquema Estrela (*Star Schema*)

Fonte: (Ramos da Costa, 2012)

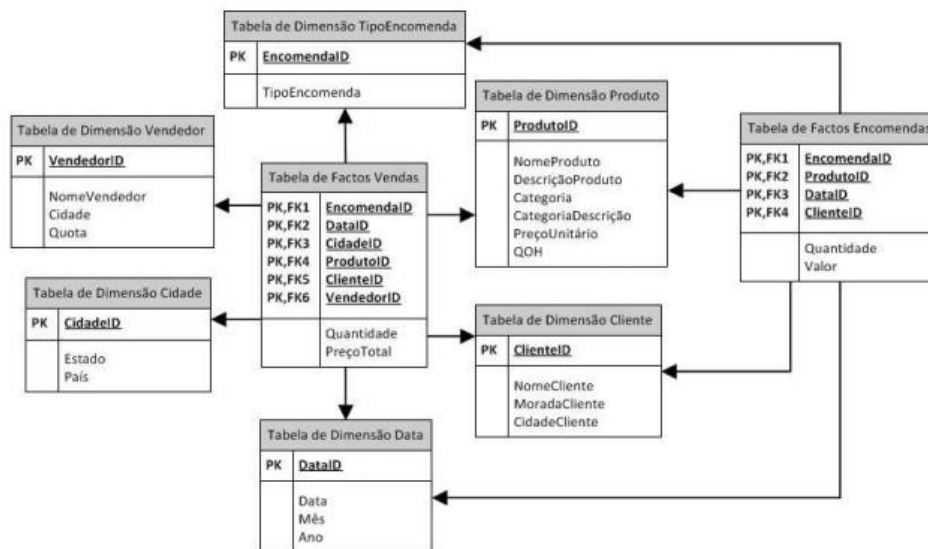


Figura 8: Esquema Constelação (*Constellation Schema*)

Fonte: (Ramos da Costa, 2012)

Já o esquema Floco de Neve caracteriza-se por ser uma extensão do esquema estrela no qual as tabelas dimensão são normalizadas. A Figura 9 ilustra um exemplo de um esquema Floco de Neve.

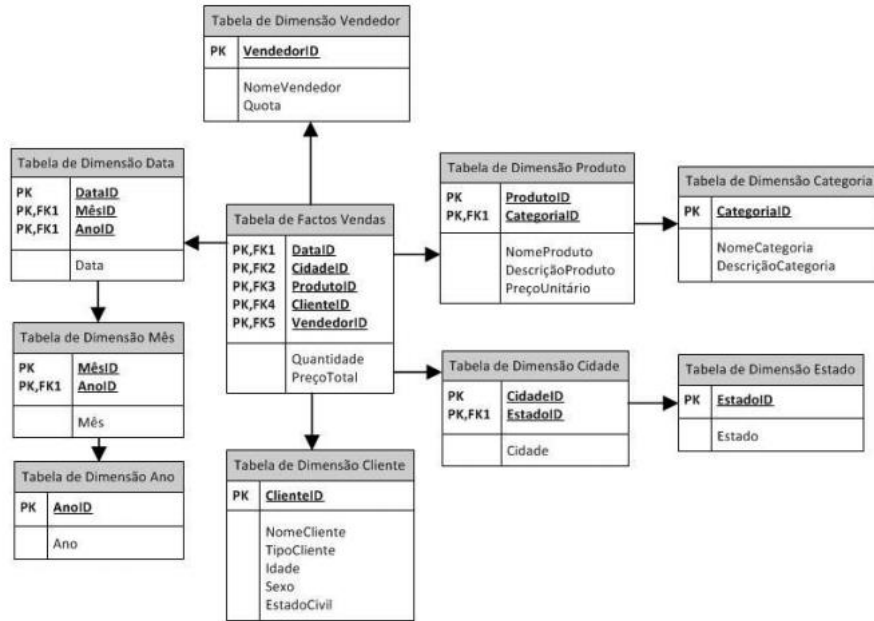


Figura 9: Esquema Floco de Neve (*Snowflake Schema*)  
 Fonte: (Ramos da Costa, 2012)

### 3.2.4. Apresentação, Análise e Exploração dos Resultados

Com vista a armazenar e organizar as informações provenientes das bases de dados transacionais, os sistemas de *data warehouse* recorrem ao modelo multidimensional, que consiste numa técnica de modelação de bases de dados, onde os dados são analisados e observados sob a perspetiva de várias dimensões (cubo de dados ou cubo OLAP, ver Figura 10).

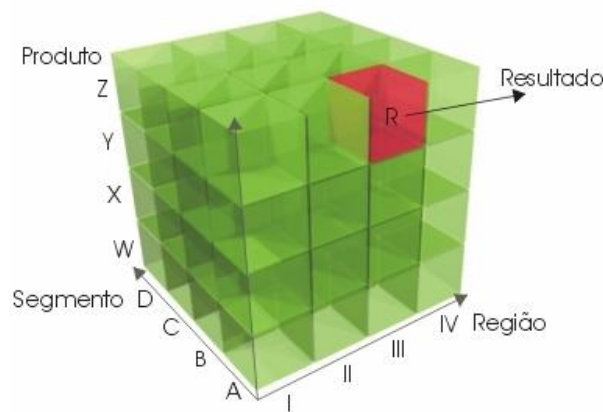


Figura 10: Cubo OLAP  
 Fonte: TMKT<sup>7</sup>

Os sistemas OLAP (*On-Line Analytical Processing*) requerem que o histórico dos dados se mantenha. Com vista a cumprir este objetivo, recorrem a dados não

<sup>7</sup> [http://www.tmktbrasil.com.br/htmls/servicos/servicos\\_crm.asp?idm=pt](http://www.tmktbrasil.com.br/htmls/servicos/servicos_crm.asp?idm=pt) (04-04-2018)

normalizados, para desta forma permitirem a realização de consultas mais complexas (Kimball & Ross, 2013). Tais pesquisas/consultas, são efetuadas por meio de *queries* (e.g., **Select \* from** (*table\_name*)), um pedaço de código que oferece aos analistas a possibilidade de pesquisar os dados armazenados na *data warehouse*. Alternativamente é oferecido ao utilizador o recurso a ferramentas como o *PowerPivot* do *Microsoft Excel* e o *PowerBI* da *Microsoft*, com o intuito de os resultados obtidos serem apresentados através de relatórios e *dashboards*, de forma a que a sua leitura seja mais clara.

## Capítulo 4

### Estágio

Neste capítulo são descritas as atividades desenvolvidas ao longo do estágio curricular que se realizou na Compta Emerging Business (CEB), nas instalações do Centro de I&D de Abrantes, com início a 27 de setembro de 2017 e término a 16 de fevereiro de 2018. No âmbito da minha colaboração com a CEB fui integrada num projeto de implementação de um sistema de BI, numa organização cuja atividade está ligada à recolha de resíduos. Um *workflow* do projeto pode ser visto na Figura 11.

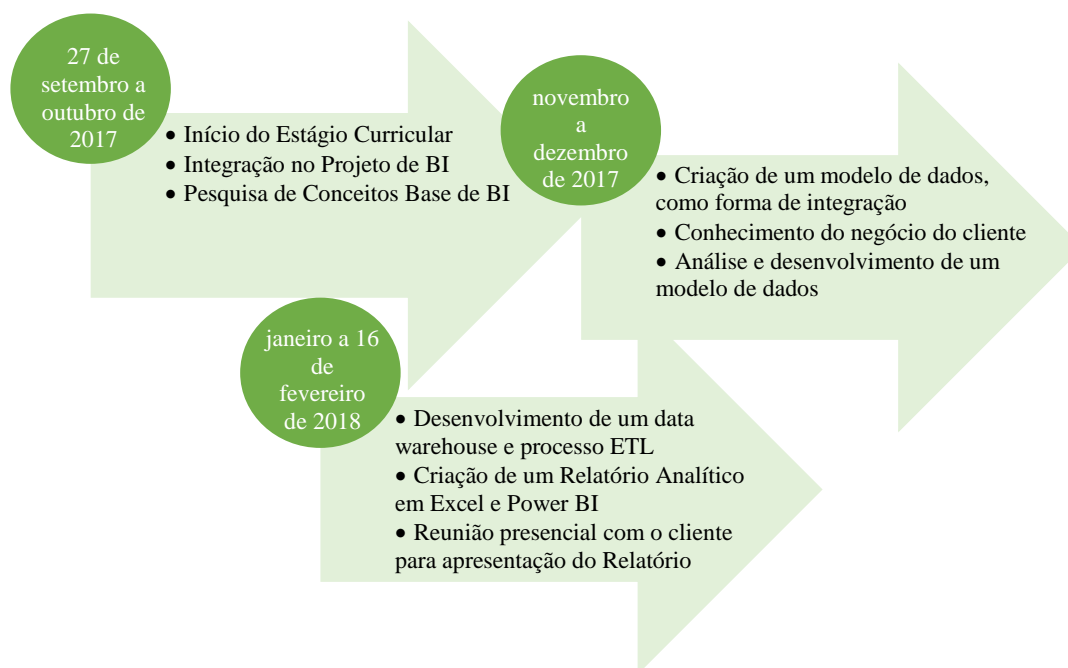


Figura 11: *Workflow* do Projeto

O principal objetivo do projeto passa por auxiliar a organização a ter mais controlo e perceção do negócio, através da apresentação de relatórios e *dashboards*. Por motivos contratuais, não é possível, detalhar mais o projeto. Para o sistema de BI implementado, durante o estágio, foram utilizadas ferramentas para (1) a obtenção dos dados; (2) extração de dados do sistema de informação operacional da organização em causa; (3) armazenamento de dados segundo um modelo adequado, nomeadamente exploração multidimensional; e (4) aplicação analítica para visualização e análise dos dados. Na Seção 4.1 são descritas as atividades efetuadas.

## 4.1. Fonte de Dados

Num primeiro momento, analisaram-se os dados provenientes da base de dados do cliente (base de dados OLTP), de forma a identificar os dados que se enquadravam nas suas necessidades. A Figura 12 demonstra a ligação à fonte de dados com recurso ao *SQL Server*.

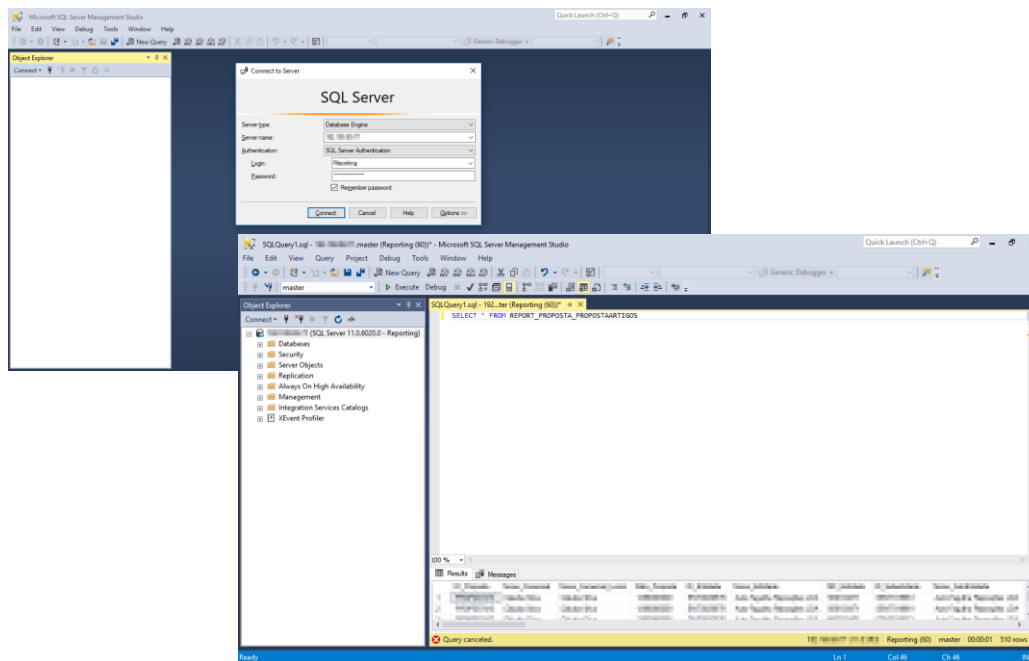


Figura 12:Ligação Fonte de Dados

## 4.2. ETL (Extract, Transformation and Loading)

Seguiu-se todo o processo de extração, transformação e carregamento de dados conhecido por ETL. Para o desenvolvimento deste processo, recorreu-se ao *Pentaho – Data Integration*. Também conhecido como *Kettle*, o *Data Integration*, é uma, de entre as várias soluções que o *Suite Pentaho* disponibiliza. Uma ferramenta *open source*, que permite extrair os dados das mais diversas fontes, transformá-los num formato compreensível e relevante para os utilizadores finais, e carregá-los de forma consistente numa base de dados (*data warehouse*). O *Kettle* é formado por dois componentes, *Transformation* e *Job*, que são construídos por meio da interface gráfica, ver Figura 13, o *Spoon*.

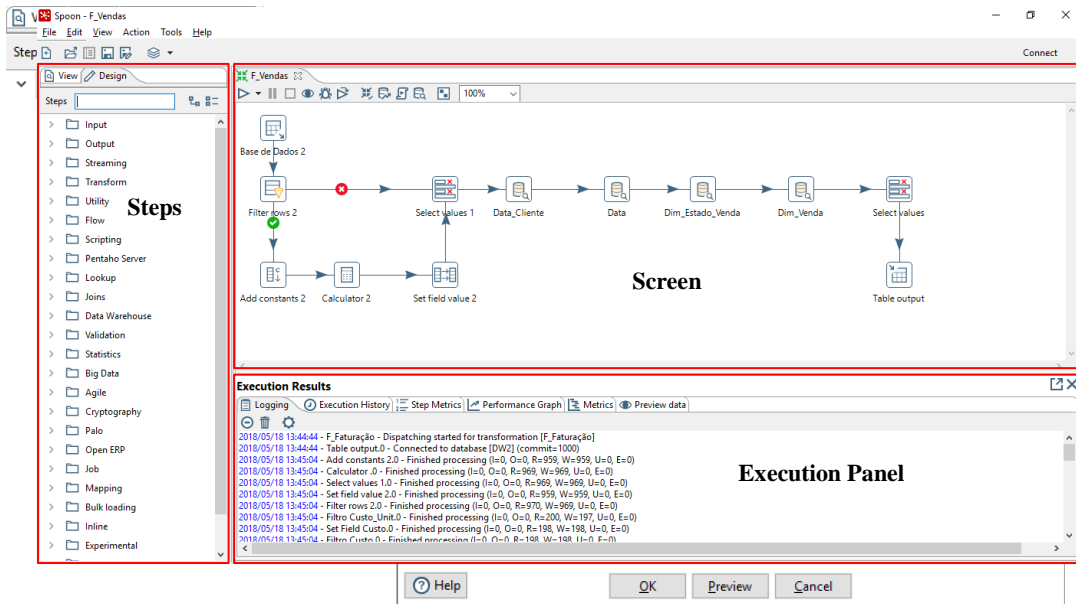


Figura 13: Interface Gráfica - Spoon

No *Spoon*, inicialmente, criaram-se as *Transformations*<sup>8</sup>. Conforme a Figura 14, estabelecem-se as conexões de *input* e *output* e depois para cada *Transformations*, ver Figura 15, na aba *Design*, campo *Input* arrasta-se para a tela o ícone mais apropriado com a fonte de origem dos dados. Neste caso, o ícone mais adequado foi o *Table Input*, por a fonte de origem dos dados ser uma base de dados operacional. Faz-se duplo clique sobre o ícone, de modo a aparecer uma nova janela onde se seleciona a conexão correspondente à base de dados de onde se vão extrair os dados e escreve-se a *query* (consulta) que se pretende.

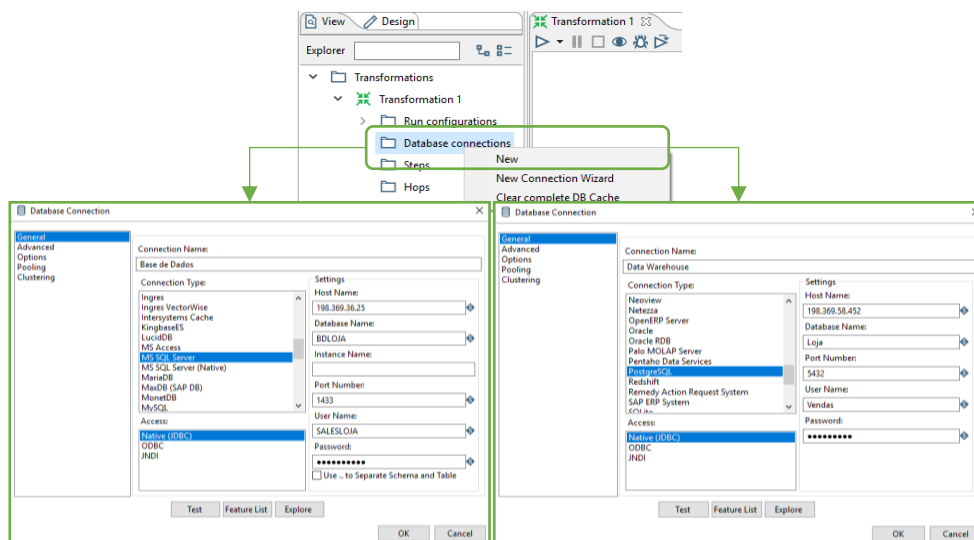


Figura 14: ETL – Conexões Data Base (*input*; *output*)

<sup>8</sup>As *Transformations* das tabelas de dimensão são as primeiras a serem executadas.

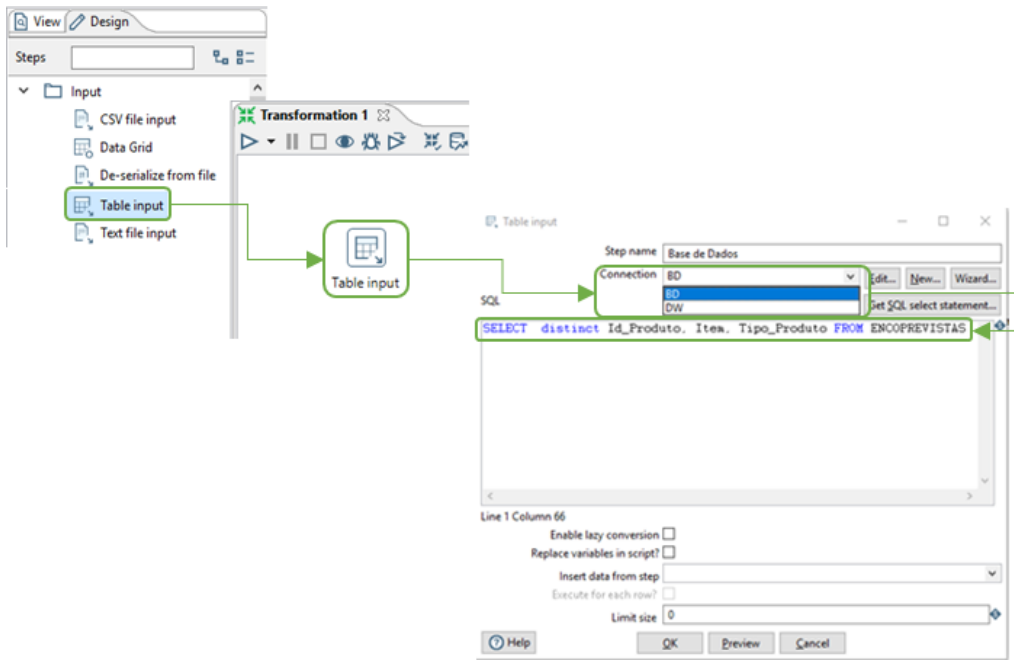


Figura 15: ETL – Conexão/Consulta com a Base de Dados Operacional

Após a conexão com a base de dados, ilustrada na Figura 15, pode-se começar a modificar e limpar os dados conforme o que for mais pertinente. Exemplo disso é o que a Figura 16 demonstra. Pois ao utilizar o ícone *Select Values* (aba *Design*, campo *Transform*), da consulta feita, podem-se selecionar os campos que para aquela *Transformation* forem necessários e até mesmo alterar o *type* do campo, isto é, o formato dos dados de determinado campo.

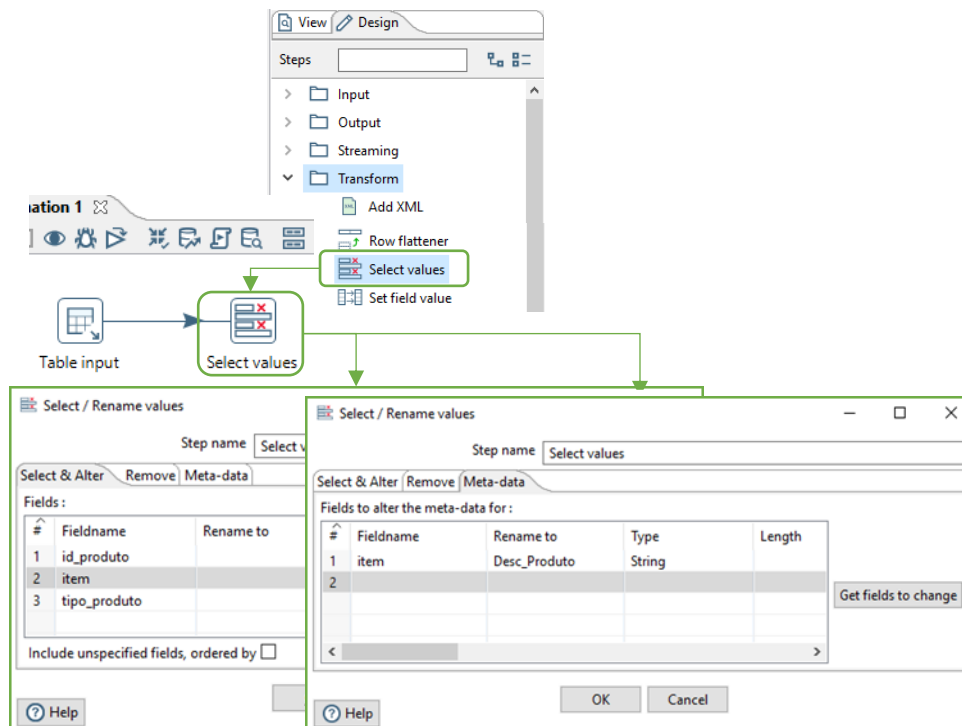


Figura 16: ETL - *Select Values*

A Figura 17, demonstra um outro exemplo ainda dentro da área da transformação dos dados utilizando o ícone *Value Mapper* (aba *Design*, campo *Transform*). Este ícone, normalmente, é utilizado quando o valor dos dados de origem representa, por exemplo, um *ID* e no *DW* pretende-se que apareça a descrição desse mesmo *ID*.

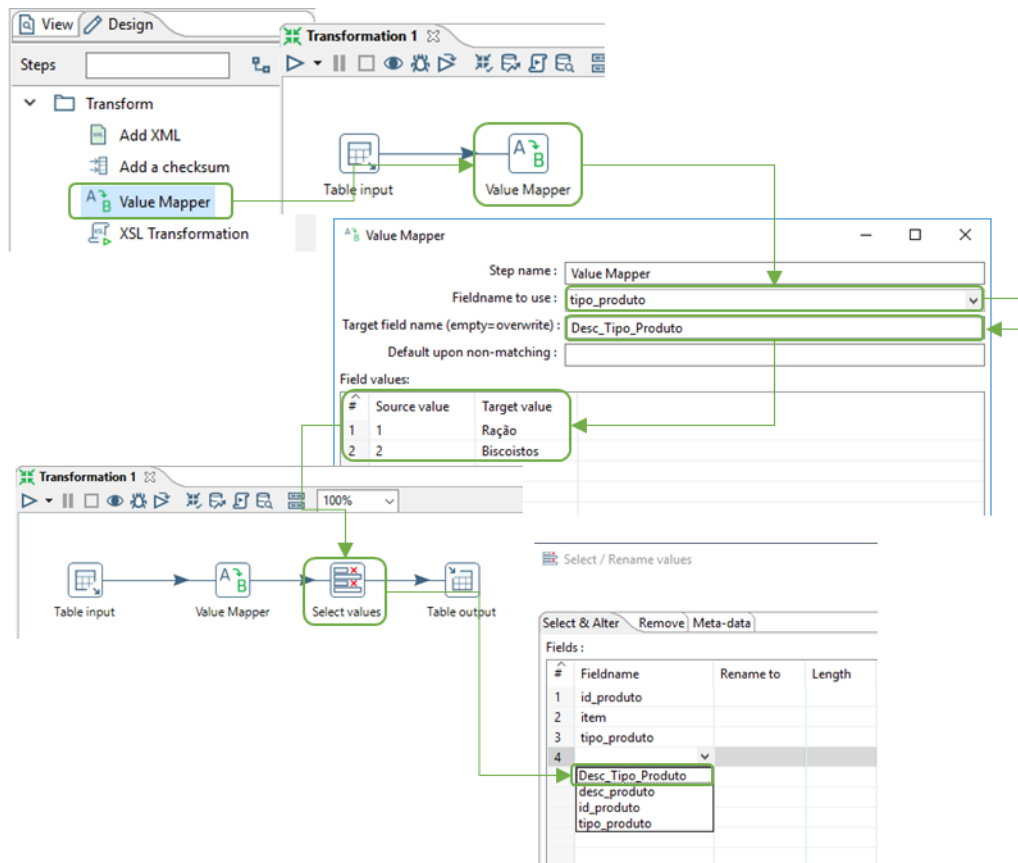


Figura 17: ETL - Value Mapper

Por fim, para concluir uma *Transformation* estabelece-se a conexão com o *data warehouse*, onde vão ser armazenados os dados já tratados. Na aba *Design*, campo *Output* arrasta-se para a tela o ícone *Table Output* e liga-se o *Select Values* ao *Table Output*. Nesta fase, para além da conexão com o *DW*, pode-se observar na Figura 18, que ainda se estabelece uma relação entre os campos da tabela de destino (*target table*) com os campos de fluxo (*stream field*), isto é, com os campos selecionados anteriormente no *Select Values*.

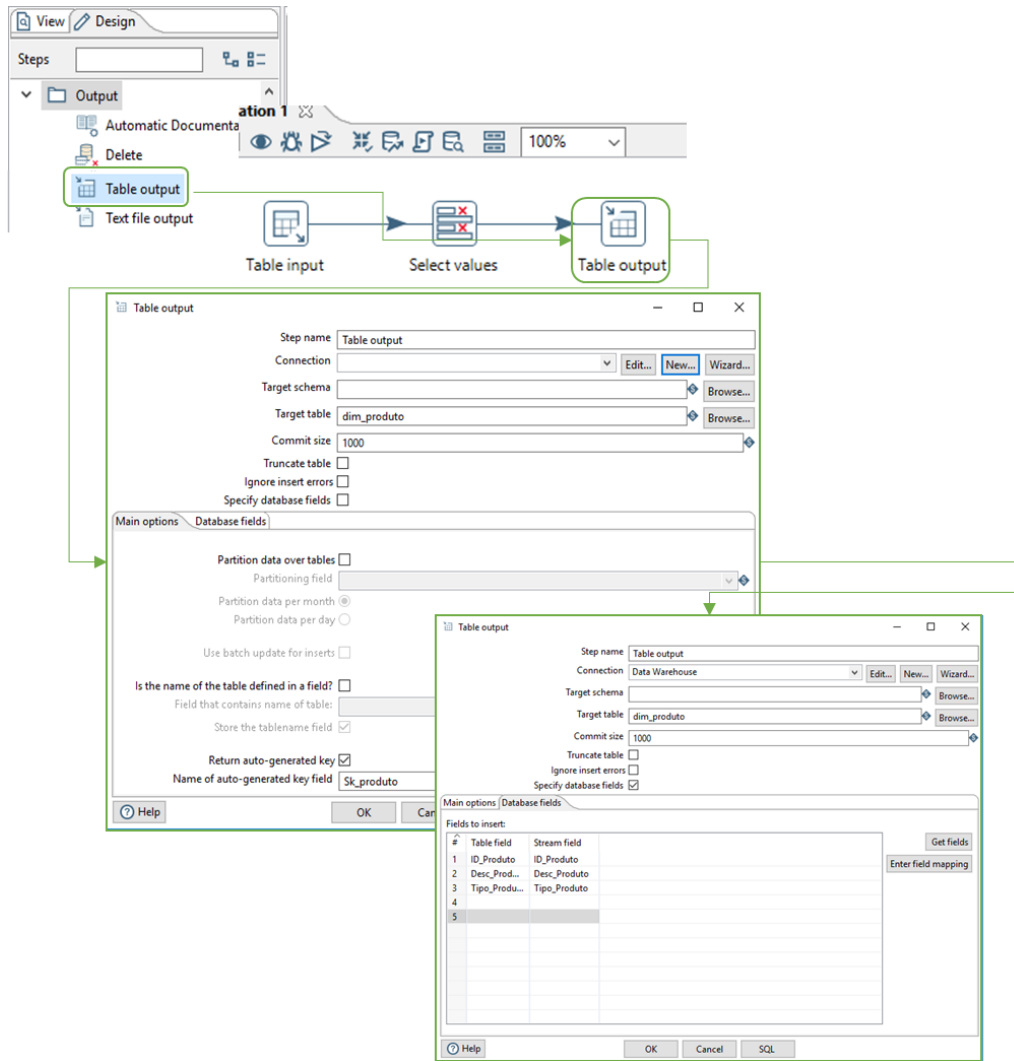


Figura 18:ETL - Conexão com *Data Warehouse*

Quando já estiverem todas as *Transformations* das tabelas de dimensão criadas, crie-se a *Transformation* da(s) tabela(s) de factos, que podem ter um aspeto idêntico ao da Figura 19.

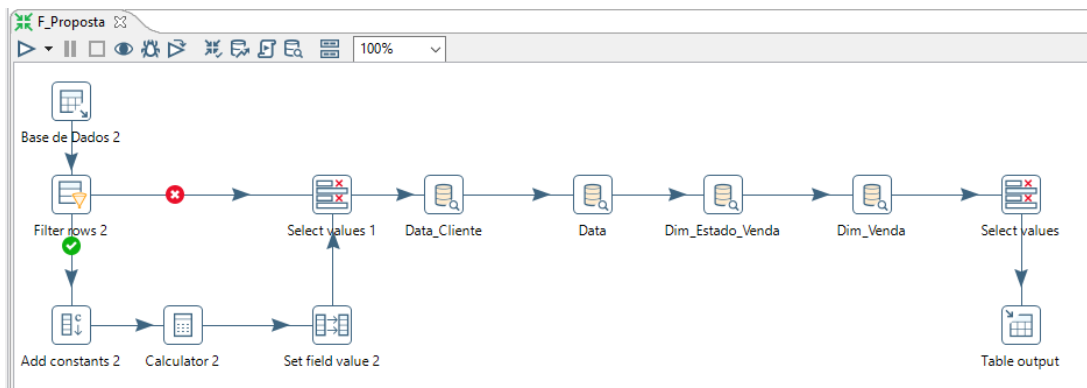


Figura 19: ETL - *Transformation* Tabela de Factos

No entanto, para que a integração dos dados no DW seja feita de forma mais rápida e eficaz é necessária a criação de um *Job*, idêntico ao da Figura 20.

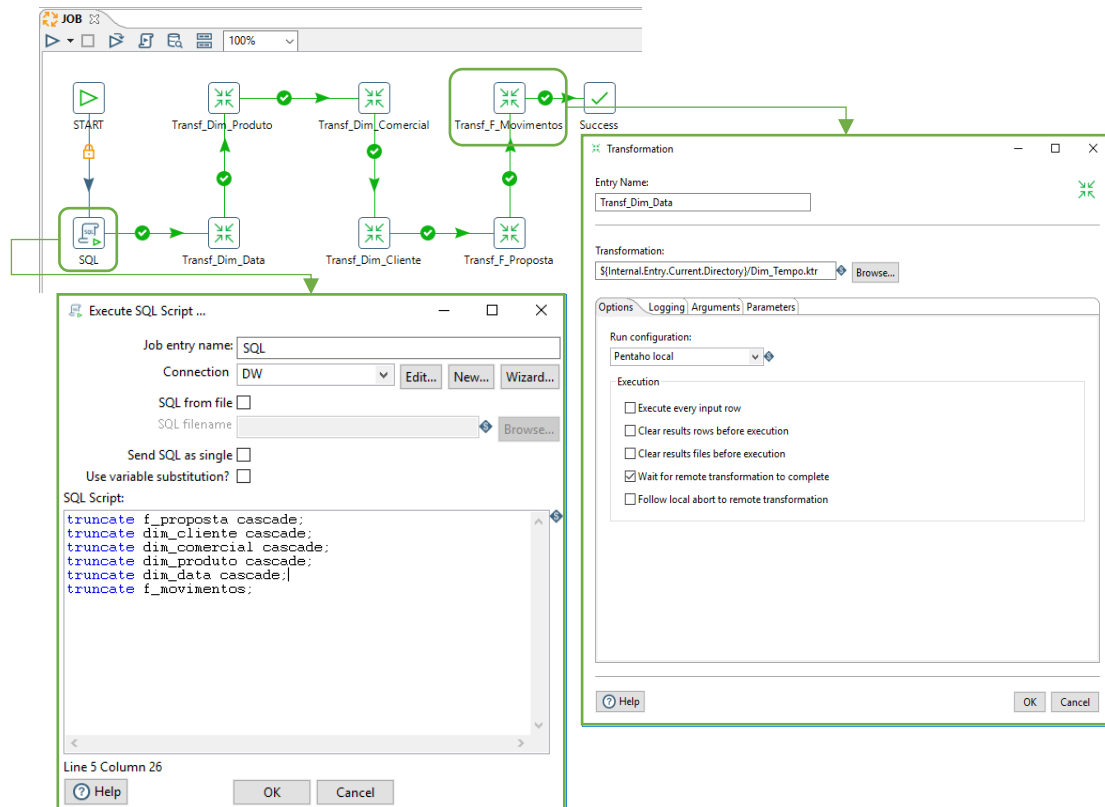


Figura 20: ETL - Criação de um *Job*

### 4.3. Repositório de Dados

Para a criação do *data warehouse* foi utilizada a ferramenta *PostgreSQL*, um sistema de gestão de base de dados objeto-relacional (SGBD) de código aberto capaz de poder ser executado nos principais sistemas operativos (*Linux, Mac OS X, Windows*). A estabilidade, robustez, e fácil administração, tornam o *PostgreSQL* numa opção viável para todo o tipo de projetos (pequeno, médio ou grande) e adequado a qualquer ramo de negócio. De salientar, que por motivos de confidencialidade os dados apresentados ao longo desta seção são fictícios.

Inicialmente optou-se por criar um modelo de dados multidimensional baseado no **esquema Star**, no entanto com o desenvolver dos trabalhos e devido aos requerimentos do cliente, teve que se optar por um **esquema Constelação** (de lembrar que este esquema pode possuir mais de duas tabelas de factos). A Figura 21 ilustra o modelo de dados do *data warehouse*, que consiste em três tabelas de dimensão, cada uma, constituída pela respetiva chave primária e atributos e duas tabelas de factos em que, cada

uma, também é constituída pela respetiva chave primária, por chaves estrangeiras e por atributos. A relação entre as tabelas é de um para muitos.

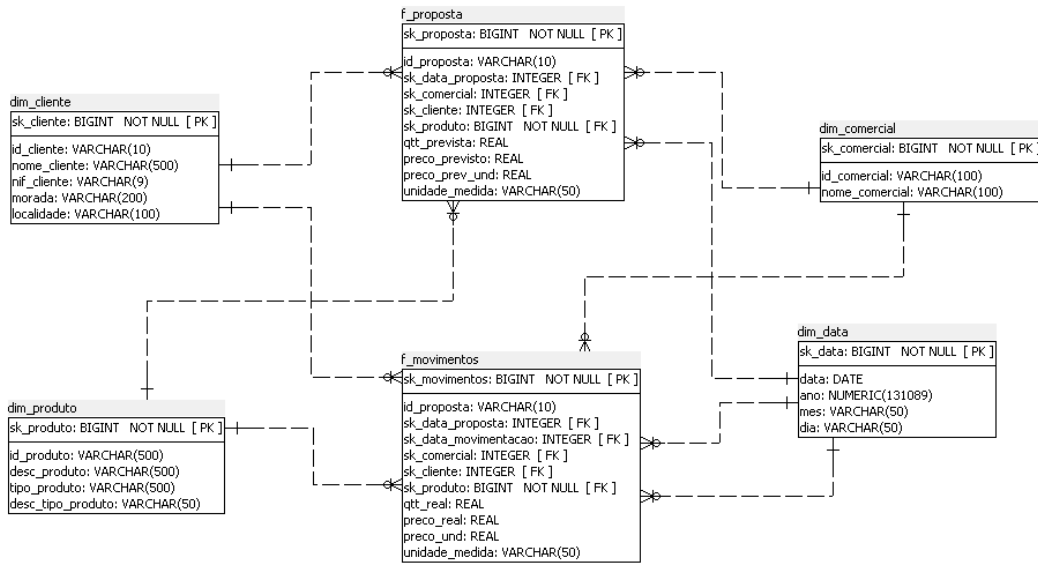


Figura 21: Modelo de Dados

Do exemplo em questão, pode-se observar, nas Figura 22 e Figura 23, uma pequena demonstração de como foram criadas as tabelas de dimensão e as tabelas de facto.

```

SQL Editor | Graphical Query Builder
Previous queries
CREATE SEQUENCE sq_dim_comercial_pkey START 1;

CREATE TABLE Dim_Comercial (
  Sk_Comercial BIGINT NOT NULL DEFAULT nextval('sq_Dim_Comercial_pkey'),
  ID_Comercial VARCHAR (100),
  Nome_Comercial VARCHAR (100),
  CONSTRAINT Dim_Comercial_pkey PRIMARY KEY (Sk_Comercial)
);

CREATE SEQUENCE sq_dim_produto_pkey START 1;

CREATE TABLE Dim_Produto (
  Sk_Produto BIGINT NOT NULL DEFAULT nextval('sq_Dim_Produto_pkey'),
  ID_Produto VARCHAR (500),
  Desc_Produto VARCHAR (500),
  Tipo_Produto VARCHAR (500),
  Desc_Tipo_Produto VARCHAR (50),
  CONSTRAINT Dim_Produto_pkey PRIMARY KEY (Sk_Produto)
);

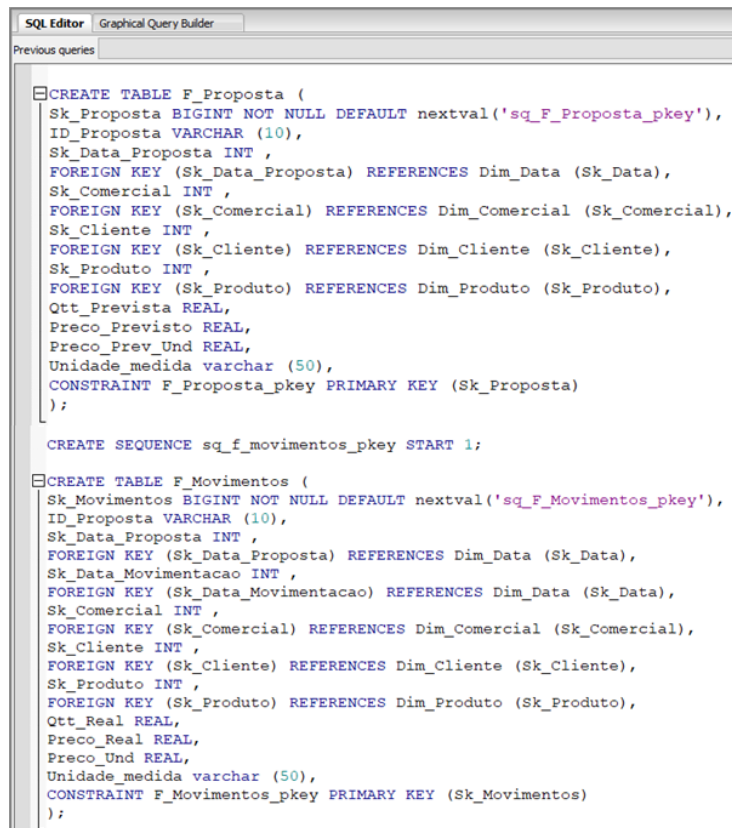
CREATE SEQUENCE sq_dim_cliente_pkey START 1;

CREATE TABLE Dim_Cliente (
  Sk_Cliente BIGINT NOT NULL DEFAULT nextval('sq_Dim_Cliente_pkey'),
  ID_Cliente VARCHAR (10),
  Nome_Cliente VARCHAR (500),
  NIF_Cliente VARCHAR (9),
  Morada VARCHAR (200),
  Localidade VARCHAR (100),
  CONSTRAINT Dim_Cliente_pkey PRIMARY KEY (Sk_Cliente)
);

CREATE SEQUENCE sq_dim_data_pkey START 1;

CREATE TABLE Dim_Data (
  Sk_Data BIGINT NOT NULL DEFAULT nextval('sq_Dim_Data_pkey'),
  data DATE,
  ano NUMERIC,
  mes VARCHAR (50),
  dia VARCHAR (50),
  CONSTRAINT Dim_Data_pkey PRIMARY KEY (Sk_Data)
);
    
```

Figura 22: Criação das Tabelas Dimensão



```

SQL Editor  Graphical Query Builder
Previous queries

CREATE TABLE F_Proposta (
  Sk_Proposta BIGINT NOT NULL DEFAULT nextval('sq_F_Proposta_pkey'),
  ID_Proposta VARCHAR (10),
  Sk_Data_Proposta INT ,
  FOREIGN KEY (Sk_Data_Proposta) REFERENCES Dim_Data (Sk_Data),
  Sk_Comercial INT ,
  FOREIGN KEY (Sk_Comercial) REFERENCES Dim_Comercial (Sk_Comercial),
  Sk_Cliente INT ,
  FOREIGN KEY (Sk_Cliente) REFERENCES Dim_Cliente (Sk_Cliente),
  Sk_Produto INT ,
  FOREIGN KEY (Sk_Produto) REFERENCES Dim_Produto (Sk_Produto),
  Qtt_Prevista REAL,
  Preco_Previsto REAL,
  Preco_Prev_Und REAL,
  Unidade_medida varchar (50),
  CONSTRAINT F_Proposta_pkey PRIMARY KEY (Sk_Proposta)
);

CREATE SEQUENCE sq_f_movimentos_pkey START 1;

CREATE TABLE F_Movimentos (
  Sk_Movimentos BIGINT NOT NULL DEFAULT nextval('sq_F_Movimentos_pkey'),
  ID_Proposta VARCHAR (10),
  Sk_Data_Proposta INT ,
  FOREIGN KEY (Sk_Data_Proposta) REFERENCES Dim_Data (Sk_Data),
  Sk_Data_Movimentacao INT ,
  FOREIGN KEY (Sk_Data_Movimentacao) REFERENCES Dim_Data (Sk_Data),
  Sk_Comercial INT ,
  FOREIGN KEY (Sk_Comercial) REFERENCES Dim_Comercial (Sk_Comercial),
  Sk_Cliente INT ,
  FOREIGN KEY (Sk_Cliente) REFERENCES Dim_Cliente (Sk_Cliente),
  Sk_Produto INT ,
  FOREIGN KEY (Sk_Produto) REFERENCES Dim_Produto (Sk_Produto),
  Qtt_Real REAL,
  Preco_Real REAL,
  Preco_Und REAL,
  Unidade_medida varchar (50),
  CONSTRAINT F_Movimentos_pkey PRIMARY KEY (Sk_Movimentos)
);

```

Figura 23: Criação das Tabelas de Factos

#### 4.4. Apresentação, Análise e Exploração dos Resultados

Por fim, seguiu-se a criação de relatórios no *Microsoft Excel* através do *PowerPivot*, suplemento que permite efetuar análises rápidas a grandes volumes de dados, bem como a criação de tabelas e gráficos dinâmicos. A execução deste projeto envolveu a realização de várias reuniões com o cliente. O acompanhamento permanente por parte do cliente, permitiu efetivar alterações pontuais ao que inicialmente estava previsto. A Figura 24 demonstra um exemplo das possíveis análises que podem ser feitas no *Excel* com o apoio do *Power Pivot* (dados são fictícios). Onde é necessária a ligação com o *data warehouse*, para podermos relacionar os dados de diversas maneiras possíveis, através da execução de *queries* (consultas) que melhor se enquadram na análise pretendida, ver Figura 25.

Análise Comercial

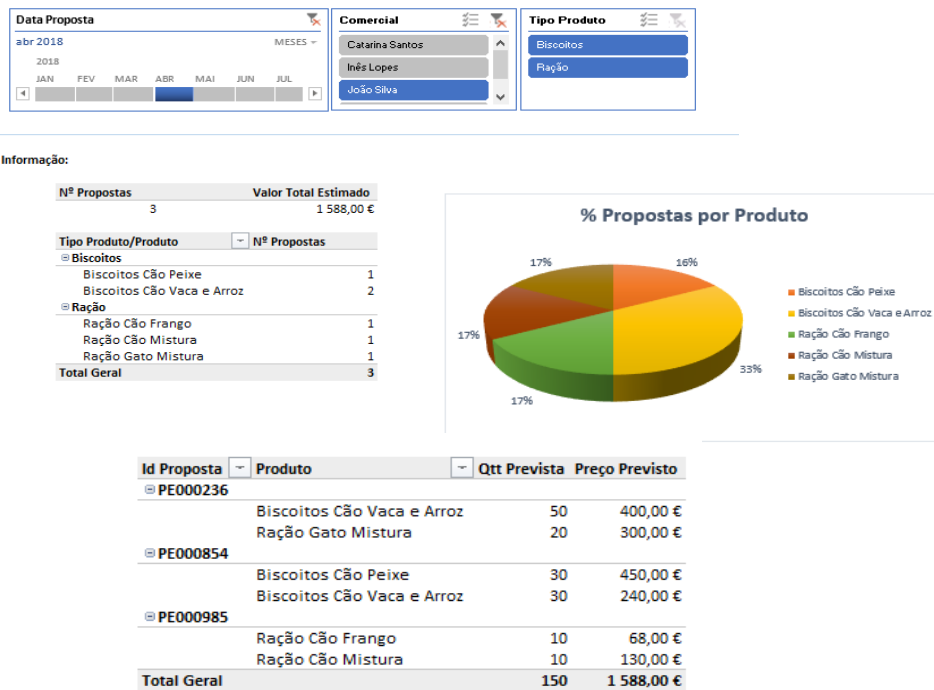


Figura 24: Análise Excel - Power Pivot

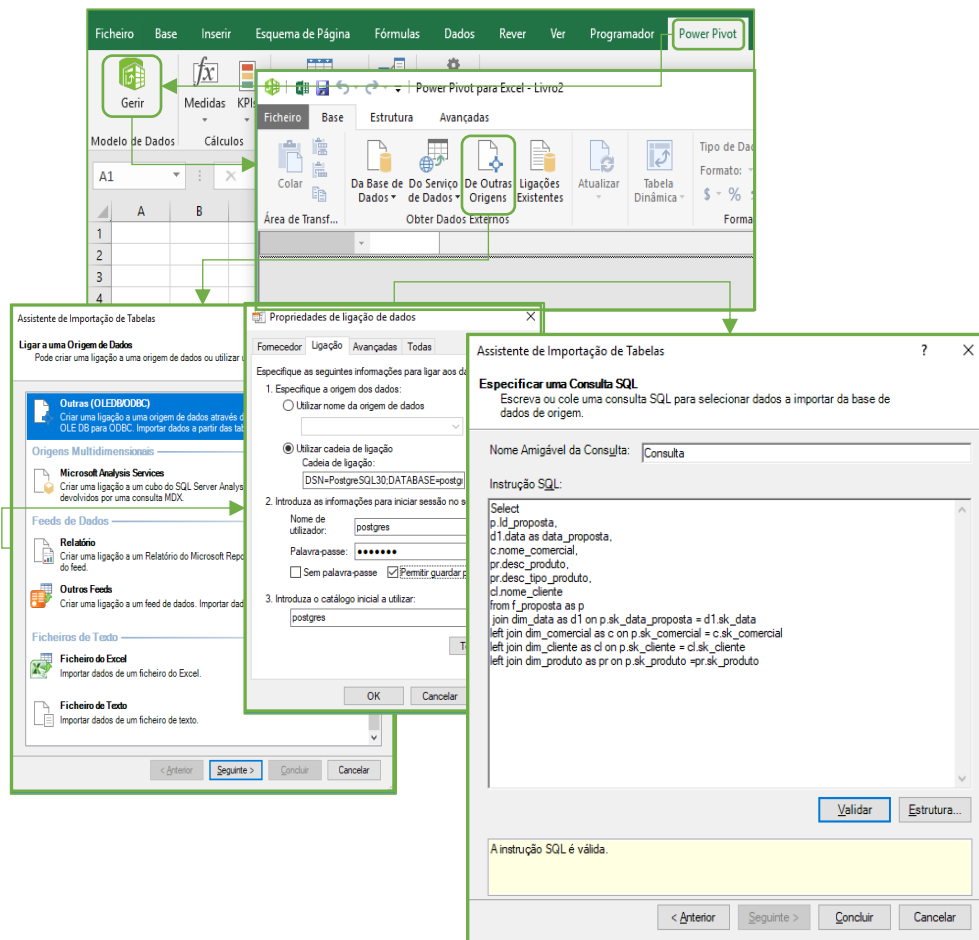


Figura 25: Conexão Data Warehouse e Consulta Dados

## Capítulo 5

### Conclusões

Numa era cada vez mais competitiva, as organizações tendem a adaptar-se às alterações, internas e externas, do meio ambiente envolvente. Desenvolver estratégias eficazes e tomar as melhores decisões, de forma rápida, passaram a ser os principais objetivos de uma organização. No entanto, é preciso que lhe seja disponibilizada informação útil e coerente. Com o volume de dados cada vez mais a aumentar, torna-se imprescindível o recurso a tecnologias de informação que facilitem o processo de gestão do conhecimento. Processo que consiste em organizar e sistematizar os dados, dados que são transformados na informação que cria o conhecimento. Os sistemas de *Business Intelligence* servem de apoio à tomada de decisões de uma organização, por meio de um processo de recolha de dados, em que os dados são adquiridos a partir de uma diversidade de fontes de dados e transformados com vista a serem armazenados numa base de dados especificamente, modelada para o negócio - *data warehouse*. Nos últimos anos tem-se assistido, no entanto ao aparecimento de novas fontes de dados capazes de alimentar as *data warehouses*. Tecnologias relacionadas com a Internet das Coisas, nomeadamente sensores, passaram a estar na base da geração de grandes quantidades de dados, proporcionando o aparecimento do conceito de *Big Data*. Por outro lado, técnicas como o *machine learning* e conceitos como o *data science*, passaram a estar também eles na órbita das organizações ao permitir a construção de modelos mais precisos capazes de tomar decisões sem a intervenção humana.

O presente relatório teve como principal objetivo descrever as atividades desenvolvidas durante o período de estágio. A realização deste estágio proporcionou a aquisição de valências tanto a nível profissional como pessoal, num enquadramento diferente ao da formação académica. O fato de estagiar numa empresa de renome e dinâmica como a Compta, tornou o estágio numa experiência deverás enriquecedora. Atualmente e após cinco meses de estágio, fui convidada a integrar a equipa da CEB, o que realça a importância do trabalho desenvolvido.



## Referências

- Alter, S. (1999). *Information systems: a management perspective* (3º ed.). Addison-Wesley Educational Publishers.
- Antonelli, R. A. (2009). Conhecendo o Bussiness Inteligence (BI). *Revista TECAP, Vol. 3*.
- Cardoso, O. N., & Machado, R. T. (2008). Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. 496 - 505.
- Céci, F. (2012). *Business Intelligence*. Livro Digital. Retrieved 2018
- Goldschmid, R., Passos, E., & Bezerra, E. (20015). *Data Mining* (2ª ed.). Elsevier Editora, Ltda.
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (Vol. Third Edition). Wiley. Retrieved 2018
- Ramos da Costa, S. A. (2012). *Sistema de Business Intelligence como*. Tese de Mestrado, Universidade do Minho. Retrieved 2018
- Sferra, H. H., & Corrêa , Â. M. (2003). Conceitos e Aplicações de Data Mining. *Revista de Ciência & Tecnologia*.