

**FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO**



**FEUP**

# **Solução de Business Intelligence utilizando tecnologias Open Source**

**Carlos Jorge Lemos Nunes**

VERSÃO FINAL

Relatório de Projecto/Dissertação  
Mestrado Integrado em Engenharia Informática

Orientador: Luís Paulo Reis

Março de 2010

Solução de Business Intelligence utilizando tecnologias Open Source

**Carlos Jorge Lemos Nunes**

Relatório de Projecto/Dissertação  
Mestrado Integrado em Engenharia Informática

Aprovado em provas públicas pelo Júri:

Presidente: António Augusto Sousa

---

Arguente: José Manuel Torres

Vogal: Luís Paulo Reis

# Resumo

O presente documento apresenta o enquadramento e introdução à proposta de dissertação intitulada, “Solução de Business Intelligence utilizando Open Source”, realizada na PT Sistemas de Informação.

O tema a ser apresentado corresponde a uma solução de Business Intelligence End-To-End que foi implementada na PT-SI, para colmatar a necessidade de diversos tipos de Indicadores, em várias áreas da Empresa. Essa solução teve como requisito principal ser construída em tecnologias Open Source.

Para isso fez-se um levantamento de requisitos com o objectivo de saber quais as funcionalidades pretendidas por parte do cliente.

Seguidamente houve um estudo sobre a informação disponibilizada, o formato das fontes, o horário que estas estariam disponíveis para a sua extracção e também sobre o formato final pretendido. Suportado pelo estado da arte, e pelas pretensões da empresa, as tecnologias adoptadas para a resolução do problema foram: PostgreSQL como SGBD e o Pentaho Suite para as funções de Extracção, Transformação e Carregamento da informação, cubos Olap e Relatórios, e, finalmente, Dashboards.

A solução final apresentada é constituída por um datawarehouse alimentado diariamente por três sistemas de informação diferentes, o CRM e o EPM da PT-SI e o CRM da PT-PRIME. A parte de reporting conta com 20 relatórios estáticos, 2 cubos OLAP, que permitem ao utilizador a criação dos seus próprios relatórios, e um OnePage Dashboard que facilita a navegação entre vários níveis de informação.

Apesar da volumetria da informação não ser elevada, as opções tomadas a nível tecnológico para a solução mostraram-se ser as acertadas, tanto a nível de performance e fiabilidade, como de funcionalidades, salvo a rara excepção da ferramenta de construção de relatórios, cujas algumas lacunas podem ser melhoradas futuramente.

# Abstract

The following document serves both as an overview and as an induction to the dissertation proposal entitled “Solução de Business Intelligence utilizando Open Source”, to be fulfilled at PT Sistemas de Informação.

The subject to be presented is related with an End-To-End Business Intelligence solution which was implemented at PT-SI, and will fill the need for distinct indicators, throughout several company business areas. The main requirement/constraint for this solution was that it had to be developed using Open Source technologies.

A requirement definition document was built together with PT-SI in order to define all required functionality for this solution.

The next step was to analyse all available information, and to know all data sources format, the schedule at which these would be available for extraction, and to define the format for End-State.

Based on state of the art design, and on the company desire, the adopted technologies were as follows: PostgreSQL as SGBD and Pentaho Suite for information extraction, handling, and loading, OLAP cubes, Reports, and Dashboards.

The End-State solution contains a data warehouse that is supplied daily by 3 different information systems: CRM and EPM from PT-SI and CRM from PT-PRIME. The reporting section handles 20 static reports, 2 OLAP cubes, which allow the user to create its own reports, and one OnePage Dashboard that allows the user to easily access several layers of information.

Although the volume of information was not that high, every technology decision made has shown to be a right bet, not only at performance level, but has also shown to be reliable and in possession of almost all required functionality: the exception being the report builder tool, that has some gaps that will be fixed in the future.

# Agradecimentos

Na certeza de que não mencionarei todos quantos contribuíram para que eu fosse capaz de chegar a esta fase, não poderia deixar de referir a minha gratidão pelos que me auxiliaram e acolheram e que, de diversas formas, me ajudaram e apoiaram nesta fase final do curso e da realização do meu projecto.

Aos meus pais, um agradecimento especial, pelo apoio e amizade que sempre me dispensaram, nos bons e nos maus momentos, acreditando sempre que eu era capaz.

À PT-SI, pela oportunidade de efectuar o projecto numa empresa jovem, dinâmica e ambiciosa, proporcionando-me as melhores condições.

Agradeço ao Pedro Cruz pela orientação e apoio que me deu em todos os momentos que compuseram este projecto.

Não posso, também, esquecer de agradecer a todos os meus colegas de trabalho pela amizade e boa disposição com que fui recebido. Uma palavra especial de agradecimento para o Alexandre Melo, Tiago Miranda, Hugo Fonseca, João Marques e João Oliveira.

Um agradecimento especial também ao meu orientador da FEUP, professor Luís Paulo Reis, por todo o apoio e pelas sugestões dadas.

Agradeço, ainda, a todos os meus amigos, que sempre me apoiaram e que me fizeram esquecer que a vida nem sempre é fácil!

# Índice

<b>1. Introdução.....</b>	<b>1</b>
1.1 Contexto/Enquadramento .....	1
1.2 Projecto .....	1
1.3 Motivação e Objectivos.....	2
1.4 Estrutura da Dissertação.....	3
<b>2. Revisão Bibliográfica.....</b>	<b>5</b>
2.1 Open Source .....	5
2.2 Business Intelligence .....	6
2.2.1 Datawarehouse .....	7
2.2.2 ETL .....	8
2.2.3 Data Analysis – Cubos OLAP .....	10
2.3 Plataformas de BI – Open Source .....	10
2.3.1 Pentaho Open BI Suite .....	10
2.3.2 Talend Open Studio.....	11
2.3.3 Avaliação .....	12
2.4 Conclusões .....	12
<b>3. Caracterização do Problema.....</b>	<b>15</b>
3.1 Funcionalidades Pretendidas .....	15
3.1.1 Repositório único – Datawarehouse (PT-SI) .....	15
3.1.2 Cubo Analítico para exploração AD-HOC .....	16
3.1.3 One Page Dashboard.....	16
3.1.4 Indicadores.....	16
3.2 Solução Apresentada.....	18
3.2.1 Modelação da Base de Dados e Processos ETL .....	18
3.2.2 Reporting .....	21
3.3 Conclusões .....	23
<b>4. Implementação .....</b>	<b>25</b>
4.1 Datawarehouse .....	25
4.1.1 Staging Area .....	25
4.2 ETL .....	27
4.2.1 – Extracção para DSA’s .....	28
4.2.2 – Carregamento das tabelas de Referência .....	30
4.2.3 – Transformação e carregamento das tabelas Factuais.....	32
4.3 Reporting.....	33
4.3.1 Relatórios estáticos.....	33
4.3.2 Cubos Olap .....	36
4.3.2 OnePage Dashboard.....	37
4.4 Conclusões .....	38

<b>5. Conclusões e Trabalho Futuro.....</b>	<b>39</b>
5.1 Satisfação dos Objectivos.....	39
5.2 Trabalho Futuro .....	39





# Lista de Figuras

Figura 1 - Modelo Típico de uma solução de BI.....	6
Figura 2 – Arquitectura Pentaho BI Suite.....	11
Figura 3 – Talend Open Studio.....	12
Figura 4 - Arquitectura da Solução Proposta.....	18
Figura 5 - Modelo Snow Flake.....	19
Figura 6 – Transformer Kettle.....	20
Figura 7 - Schema Workbench.....	21
Figura 8 - Report Designer.....	22
Figura 9 - Design Studio.....	22
Figura 10 - Dashboard PT-SI.....	23
Figura 11 – Transformação para carregamento de DSA.....	28
Figura 12 – Extracção de Stored Procedure.....	28
Figura 13 – Extracção de Excel.....	29
Figura 14 – Mapeamento dos campos.....	29
Figura 15 – Carregamento Referências.....	30
Figura 16 – Operador de Mapeamento.....	30
Figura 17 – Operador Lookup.....	31
Figura 18 – Slowly Changing Dimensions.....	32
Figura 19 – Job Global de Carregamento.....	32
Figura 20 - Evolução de Propostas.....	33
Figura 21 - Nº de Propostas.....	34
Figura 22 - Custo de Elaboração de Propostas.....	35
Figura 23 - Propostas por PVP.....	35
Figura 24 - Cubo PT-PRIME.....	36
Figura 25 - Top 5 Clientes.....	37
Figura 26 - Detalhe Top 5.....	37
Figura 27 - Dashboard.....	38

# Lista de Tabelas

Tabela 1 – Fases do Projecto .....	2
Tabela 2 – Os quatro tipos de Liberdade .....	5
Tabela 3 – Requisitos do Datawarehouse .....	15
Tabela 4 – Tipologias dos Indicadores .....	16
Tabela 5 – Indicadores Financeiros .....	16
Tabela 6 – Indicadores Operacionais .....	17
Tabela 7 – Indicadores de Clientes/Parceiros .....	17
Tabela 8 – Características do Kettle .....	20
Tabela 9 – 1ª Solução .....	26
Tabela 10 – 2ª Solução .....	26
Tabela 11 – Tabela de Regras.....	26
Tabela 12 – Tabela de Referência .....	26



# Abreviaturas e Símbolos

BI	Business Intelligence
DW	Datawarehouse
ETL	Extract Transform and Load
PT	Portugal Telecom
PT-SI	Portugal Telecom – Sistemas de Informação
CRM	Costumer Relationship Management
EPM	Enterprise Project Management
ERP	Enterprise Resource Planning
ETL	Extract Transform and Load
PME	Pequenas e Médias Empresas
SOHO	Small Office, Home Office
SI	Sistemas de Informação
OLAP	On-line Analytical Processing
GPL	General Public License
OLTP	Online Transaction Processing
SGBD	Sistema de Gestão de Base de Dados
SQL	Structured Query Language
CDF	Community Dashboard Framework
HTML	Hypertext Markup Language
CSS	Cascading Style Sheets
DSA	Data Staging Area
KPI	Key Performance Indicator

# 1. Introdução

## 1.1 Contexto/Enquadramento

No Mundo actual, onde as empresas têm de tomar decisões bastante importantes em curtos espaços de tempo, toda e qualquer informação que as ajude a tomar essas decisões de uma forma acertada tem um valor incalculável para elas.

A PT-SI, criada a 22 de Dezembro de 1999, é a empresa do Grupo Portugal Telecom responsável pelo fornecimento de Soluções de Tecnologias e Sistemas de Informação ao mercado empresarial.

Beneficiando da vasta experiência e know-how do Grupo PT, procura contribuir para a competitividade dos seus Clientes, criando valor através da integração de Tecnologias e Sistemas de Informação com Telecomunicações, tendo sempre em conta as melhores práticas existentes no mercado e a inovação, assegurando sempre a melhoria contínua e a natural evolução tecnológica.

As parcerias estabelecidas com empresas reconhecidas pelo mercado, entre as quais se destacam a Microsoft, SAP, Oracle e Primavera, entre outros, representam uma grande mais-valia da PT-SI [1].

## 1.2 Projecto

Os SI são de extrema importância para o sucesso ou insucesso de muitas empresas. CRM's, EPM's, ERP's são cruciais para o seu funcionamento e desenvolvimento.

No entanto, diariamente, há decisões estratégicas para se tomarem: mercados onde se deve investir; sectores de mercado que se vieram a revelar uma má aposta; clientes mais rentáveis que convém cativar; etc. Estas decisões, baseadas em relatórios de sistemas operacionais, eram por si só bastante difíceis de tomar, uma vez que as variáveis que as condicionam são imensas e de extrema complexidade.

Este projecto enquadra-se, portanto, na necessidade de implementar uma solução de Business Intelligence na área de Reporting para a PT-SI que permita a produção automática de indicadores e a sua disponibilização pelas diferentes áreas da empresa.

O projecto contempla a produção de diferentes deliverables divididos em três fases, como se apresenta detalhadamente na tabela 1.

1ª Fase	Especificação de Requisitos (Análise Funcional e Técnica).
2ª Fase	Repositório Único (Datawarehouse PT-SI) Processos de ETL
3ª Fase	Cubo Analítico para Exploração. Relatórios com Indicadores Operacionais e Financeiros. OnePage Dashboard

**Tabela 1 – Fases do Projecto**

Como descrito na tabela 1, na 1ª fase será feita uma especificação de Requisitos, através da elaboração de uma Análise Funcional e de uma Análise Técnica. A análise funcional contempla todas as funcionalidades desejadas, assim como as metodologias para as conseguir desenvolver. A análise técnica expõe o modo de como implementar, chegando a ir ao pormenor de cada acção a desenvolver.

A 2ª fase passa pelo início da implementação, a construção do Datawarehouse único, assim como todos os processos de ETL, que tiveram de ser escrupulosamente desenhados para garantir a integridade e a coerência da informação. Esta fase foi dificultada pelas diversas fontes que convergiam numa só.

O projecto foi concluído numa 3ª fase, a fase do Reporting, onde foram criados 2 cubos analíticos para exploração (OLAP), que servirão de suporte a relatórios dinâmicos modelados pelos utilizadores do sistema. Foram construídos relatórios estáticos com indicadores operacionais e financeiros. Estes relatórios também irão ser disponibilizados em servidor, podendo ser exportados em vários formatos. Finalmente, foi criado um Dashboard onde será possível a navegação entre vários níveis de informação de uma forma mais intuitiva e perceptível a todos os utilizadores, mesmo os menos familiarizados com o negócio.

### **1.3 Motivação e Objectivos**

O negócio das empresas de Sistemas de Informação é muito complexo e alargado. Desde a chegada de uma oportunidade até à sua adjudicação, a oportunidade vai passando por muitas fases, muitas pessoas são envolvidas e muitos recursos são gastos.

A possibilidade de ajudar uma empresa desta dimensão a criar indicadores para poderem filtrar de antemão as propostas que querem analisar, as empresas que querem trabalhar, estratégias de acção que querem definir (os melhores colaboradores para as melhores empresas), possibilitando assim maximizar os ganhos, minimizando os gastos, aumentar o lucro da empresa e a satisfação do cliente, deu-me uma enorme motivação, uma vontade imensa de obtenção de resultados relevantes e úteis para todos os Key Users.

A tecnologia a utilizar também foi bastante aliciante uma vez que era algo pioneiro no domínio de BI. O sucesso do projecto poderia significar o sucesso e a aposta da PT-SI (Sistemas de BI) nas tecnologias de Open Source

## **1.4 Estrutura da Dissertação**

Para além da introdução, esta dissertação contém 4 capítulos. No capítulo 2 é apresentado o Estado da Arte. No capítulo 3 são apresentadas as funcionalidades base do sistema, os requisitos detalhados e a solução encontrada, que é depois detalhada no capítulo no capítulo 4, tanto a nível funcional como técnico. No 5º e último capítulo, são apresentadas as conclusões do trabalho efectuado, e é abordado o tema de trabalho futuro.



## 2. Revisão Bibliográfica

### 2.1 Open Source

A filosofia do Software Livre encontra as suas raízes na livre troca de conhecimentos e de pensamentos que podem tradicionalmente ser encontrados no campo científico. Tal como as ideias, os programas de computador não são tangíveis e podem ser copiados sem perda. A sua distribuição é a base de um processo de evolução que alimenta o desenvolvimento do pensamento.

No início dos anos 80, Richard M. Stallman [2] foi o primeiro a formalizar esta maneira de pensar para o software sobre a forma de quatro liberdades:

<b>1ª Liberdade:</b>	A liberdade de executar o software, para qualquer uso.
<b>2ª Liberdade:</b>	A liberdade de estudar o funcionamento de um programa e de adaptá-lo às suas necessidades.
<b>3ª Liberdade:</b>	A liberdade de redistribuir cópias.
<b>4ª Liberdade:</b>	A liberdade de melhorar o programa e de tornar as modificações públicas de modo que a comunidade inteira beneficie da melhoria.

**Tabela 2 – Os quatro tipos de Liberdade**

O software que siga os quatro princípios apresentados na Tabela 2 é chamado "Software Livre" (ou Free Software).

Para suportar essa ideia e fazer com que tudo isso se realize, Richard M. Stallman criou a "Free Software Foundation" em 1984 e lançou o projecto GNU [3]. A Licença Pública Geral GNU não somente concede as quatro liberdades descritas acima, mas também as protege. Graças a essa protecção, a GPL é, hoje em dia, a licença mais utilizada para o Software Livre.

Ao lado da GPL existem outras licenças que concedem essas liberdades, o que as qualifica de licenças de Software Livre. Uma delas, a licença FreeBSD [4], merece

uma menção particular. A principal diferença com a GPL é que ela não procura proteger a liberdade.

Quando se fala de Software Livre, uma confusão frequente é de pensar que um tal software deve ser grátis (principalmente porque em inglês Free significa "livre", mas também significa "grátis"). Na realidade, uma grande parte dos protagonistas do Software Livre, que trabalham no campo do Software Livre comercial.

Em 1998, a "Definição do Open Source" (Open Source Definition) foi escrita, tendo o cidadão dos E.U.A. Bruce Perens [4] como autor principal. O seu objectivo era descrever as propriedades técnicas do Software Livre e ser utilizada como texto fundador do movimento "Open Source" (Open Source Movement).

A "Definição do Open Source" é ela mesma derivada das "Linhas Directoras do Software Livre Debian", que resultam das quatro liberdades mencionadas acima. Consequentemente, as três definições descrevem as mesmas licenças; a "Licença Pública Geral GNU" (GPL) é a licença de base de todas as definições.

O movimento "Open Source" tem por objectivo ser um programa de marketing do Software Livre. Esse objectivo deliberadamente ignora todos os aspectos filosóficos ou políticos; que são considerados prejudiciais à comercialização.

Por outro lado, o movimento Software Livre considera o ambiente filosófico/ético e político como uma parte essencial do movimento e um dos seus pilares fundamentais.

## 2.2 Business Intelligence

O termo Business Intelligence (BI) foi criado pelo Gartner Group [6] nos anos 80 e tem como objectivo não só transformar dados resultantes de várias actividades de uma determinada organização em informação legível, mas principalmente transformar esses dados em conhecimento que possibilite a tomada de melhores decisões críticas com vista ao desenvolvimento de uma empresa. As soluções de BI facultam às organizações um conhecimento mais aperfeiçoado dos seus clientes e potenciam uma gestão táctica e estratégica dos negócios que lhes possibilita aprender com o passado, antecipar oportunidades e prever potenciais problemas. Permite ainda às empresas orientar as estratégias de abordagem aos seus clientes de forma consolidada, reunindo o maior número de informações díspares, normalmente existentes em diversos sistemas.

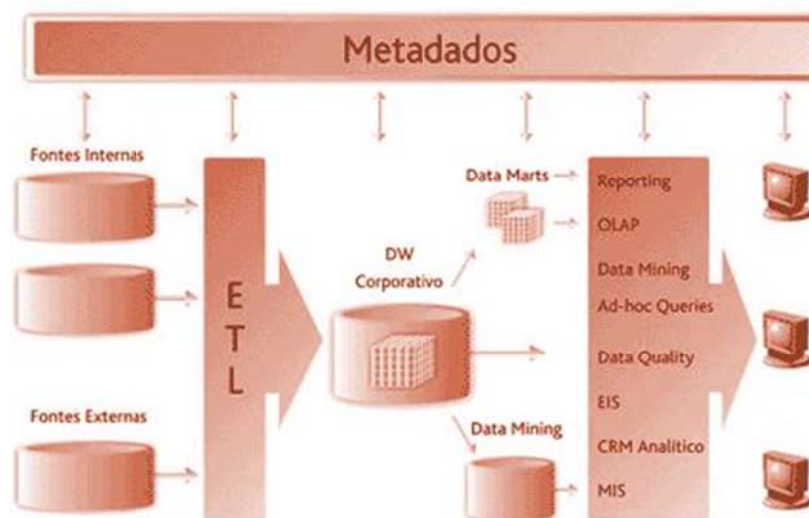


Figura 1 - Modelo Típico de uma solução de BI

Numa solução de BI é essencial a existência de um Datawarehouse, dado ser o componente que permite armazenar de uma forma consistente e segura um elevado volume de informação relativa às actividades da organização. A sua arquitectura visa otimizar o processo de análise aos dados armazenados suportando e simplificando o processo de elaboração de relatórios. Hoje, qualquer organização necessita, a todos os níveis de gestão de informação e meios de análise, de ferramentas que ajudem à tomada de decisões críticas e importantes. Para isso, esta informação tem de ser actual e de disponibilidade imediata, suportada em métodos de acesso rápidos e intuitivos acompanhados de processos de análise e de reporting poderosos.

Actualmente, a maioria das empresas já investiram ou estão a investir em aplicações de BI que garantam a correcta manutenção e qualidade dos dados para futuras utilizações, nomeadamente na produção de relatórios que permitam uma análise eficiente de toda a informação armazenada. Este tipo de aplicações é, hoje em dia, indispensável para garantir uma competitividade permanente da empresa com todas as outras do mesmo ramo. Os estudos e as conclusões provenientes destas aplicações permitem uma correcta aplicação de capital, um aperfeiçoamento nos métodos de obtenção de novos clientes, uma eficaz análise por parte da administração, uma compreensão generalizada da situação da empresa, entre muitas outras vantagens.

## **2.2.1 Datawarehouse**

Um datawarehouse é um sistema de computação utilizado para armazenar informações relativas às actividades de uma organização em bases de dados, de forma consolidada. O desenho da base de dados favorece os relatórios, a análise de grandes volumes de dados e a obtenção de informações estratégicas que podem facilitar a tomada de decisão.

O datawarehouse possibilita a análise de grandes volumes de dados, colectados dos sistemas operacionais. É o histórico que possibilita uma melhor análise de eventos passados, oferecendo suporte às tomadas de decisões presentes e a previsão de eventos futuros. Por definição, os dados num datawarehouse não são voláteis, ou seja, eles não mudam, salvo quando é necessário fazer correcções de dados previamente carregados. Os dados estão disponíveis somente para leitura e não podem ser alterados.

Os datawarehouse surgiram como conceito académico na década de 80. Com o amadurecimento dos sistemas de informação empresariais, as necessidades de análise dos dados cresceram paralelamente. Os sistemas OLTP não conseguiam cumprir a tarefa de análise com a simples geração de relatórios. Nesse contexto, a implementação do datawarehouse passou a tornar-se realidade nas grandes corporações. O mercado de ferramentas de datawarehouse, que faz parte do mercado de BI, cresceu então, e ferramentas melhores e mais sofisticadas foram desenvolvidas para apoiar a estrutura do datawarehouse e sua utilização.

Actualmente, por sua capacidade de sumarizar e analisar grandes volumes de dados, o datawarehouse é o núcleo dos sistemas de informações gerências e apoio à decisão das principais soluções de business intelligence do mercado.

As duas tecnologias Open Source de SGBD mais usadas mundialmente são o MySQL e o PostgreSQL.

### **2.2.1.1 – MySQL**

O MySQL é um SGBD veloz, ideal para aplicações onde velocidade seja algo mais crucial do que segurança. Apesar de fornecer garantia de integridade e

consistência dos dados, ele deixa a desejar em vários aspectos, como, por exemplo, integridade referencial [7].

### 2.2.1.2 – PostgreSQL

O PostgreSQL é mais lento do que o MySQL, sobretudo quando o quesito em questão é a velocidade observada durante o armazenamento e recuperação de informações. Por esses motivos esse SGBD é recomendado em aplicações nas quais a segurança dos dados é algo crucial, mas a performance não é prioritária[7].

É também o mais indicado para a construção de ferramentas mais complexas devido ao seu maior suporte de funções SQL, tais como, views, triggers, constraints, procedures, que diminuem algum trabalho a nível de programação. Todavia, este maior suporte tem grandes custos a nível de performance [7].

### 2.2.2 ETL

Esta etapa do processo deve-se basear na busca das informações mais importantes em sistemas fontes ou externos e que estejam em conformidade com a modelagem do DW. Tal busca de dados pode ser obstruída por problemas como a distribuição das origens dos dados, que podem estar em bases distintas com plataformas diferentes, exigindo a utilização de formas de extracção diferentes para cada local[9].

No momento de criação do DW é comum um carregamento de dados inicial que faça com que a extracção procure todos os dados dos sistemas fontes; mas com o decorrer do tempo a extracção deve estar preparada apenas para fazer carregamentos incrementais. O carregamento incremental, que carrega apenas os registos que foram alterados ou inseridos desde o carregamento inicial, é muito mais eficiente[9].

A transformação dos dados é a fase subsequente à sua extracção. Esta fase não só transforma os dados, mas também efectua a limpeza dos mesmos. A correcção de erros de digitação, a descoberta de violações de integridade, a substituição de caracteres desconhecidos, a padronização de abreviações podem ser exemplos desta limpeza. Segundo Kimball [8], as características mais relevantes para garantir a qualidade dos dados são:

- Unicidade, evitando assim duplicações de informação;
- Precisão. Os dados não podem perder suas características originais assim que são carregados para o DW;
- Completude, não gerando dados parciais de todo o conjunto relevante às análises;
- Consistência, ou seja, os factos devem apresentar consistência com as dimensões que o compõem.
- É necessário que os dados fiquem em uma forma homogénea para serem carregados no DW. Durante o processo de homogeneização, são encontrados muitos conflitos de modelagem. Estes conflitos podem ser divididos em semânticos e estruturais.

Os conflitos semânticos são todos aqueles que envolvem o nome ou a palavra associada às estruturas de modelagem, por exemplo, mesmo nome para diferentes entidades ou diferentes nomes para a mesma entidade. Já os conflitos estruturais englobam os conflitos relativos às estruturas de modelagem escolhidas, tanto no nível de estrutura propriamente dita como no nível de domínios. Os principais tipos de conflitos estruturais são aqueles de domínio de atributo que se caracterizam pelo uso de diferentes tipos de dados para os mesmos campos.

Os conflitos típicos de domínio de atributo são:

- Diferenças de unidades: quando as unidades utilizadas diferem, embora forneçam a mesma informação (exemplo: distância em centímetros ou polegadas);
- Diferenças de precisão: quando a precisão escolhida varia de um ambiente para outro (exemplo: o custo do produto é armazenado com duas posições '0,12' ou com seis posições decimais '0,123456');
- Diferenças em códigos ou expressões: quando o código utilizado difere um do outro (exemplo: sexo representado por M ou F e por 0 ou 1);
- Diferenças de granularidade: quando os critérios associados a uma informação, embora utilizando uma mesma unidade, são distintos (exemplo: quando horas trabalhadas correspondem às horas trabalhadas na semana ou às horas trabalhadas no mês);
- Diferenças de abstracção: quando a forma de estruturar uma mesma informação segue critérios diferentes (exemplo: endereço armazenado em um único atributo, ou subdividido em rua e complemento).

Depois de identificados os conflitos de modelagem, devem-se criar as regras de conversão para os padrões estabelecidos pelo Datawarehouse. Essas regras podem ser criadas com o auxílio de ferramentas de integração utilizadas para o processo de extracção e carregamento de dados. Após a criação das regras, a etapa de carregamento dos dados pode ser planeada.

Basicamente são carregadas as dimensões estáticas, de modificação lenta ou remanescente e factos integrantes ao modelo do DW. Este processo pode ter alto custo de processamento além de implicar em tempo de carga que na maioria das vezes não pode ser extenso devido à utilização contínua do DW. Assim, algumas precauções podem ser tomadas antes de se iniciar o carregamento dos dados, como:

- Desligamento de índices e referências de integridade (isso pode prejudicar na qualidade dos dados pois apesar de diminuir o processamento, os dados não são validados no momento da inserção);
- Utilização de comandos do tipo truncate ao invés de delete pois nos SGBD's mais actuais este recurso não gera armazenamento de informações em áreas de recuperação de dados;

Ter a consciência de que no momento do carregamento alguns dados não serão carregados e deste modo os mecanismos do processo devem dar suporte a auditorias de carregamento para que o mesmo possa ser reiniciado no momento em que foi parado e a possibilidade de manter logs com os dados rejeitados para a avaliação dos motivos pelo qual não foram carregados e assim ajustados para integrarem o conjunto a ser carregado.

Dimensões estáticas normalmente não oferecem problemas, pois estas mantêm dados que não sofrem alteração na sua origem e serão carregados uma única vez, assim como as remanescentes que normalmente são originadas de esforço manual na sua confecção, por exemplo, através de folhas de cálculo. Já as dimensões de modificação lenta necessitam da verificação nas suas fontes e nas auditorias dos carregamentos para que se possa identificar qual o momento seguinte depois do último carregamento que deve iniciar o processo, gerando processamento na leitura de logs de sistemas operacionais e comparação de atributos, podendo então ser necessário subscrever todo o conteúdo de um registo, gerar um novo registo na dimensão ou criar um atributo a mais para armazenar o valor antigo [8].

Após as dimensões estarem correctamente carregadas, já é possível iniciar o carregamento dos factos, que depois de modelados para conter apenas os dados de importância para a organização, direccionam quais regras serão utilizadas como, por exemplo, filtros do que será inserido ou somas a serem realizadas, provocando o aparecimento de regras que passaram despercebidas no início da modelagem.

No entanto, os factos precisam de cuidados no seu carregamento como o uso das chaves artificiais das dimensões para que se tenha uma integridade referencial, controlo de valores nulos obtidos no momento da transacção para que não gerem a falta de integridade referencial como datas que, estando nulas, invalidarão o histórico do facto.

Técnicas para amenizar o processo devido ao grande volume de dados podem ser usadas, como o carregamento incremental dos factos, que irá carregar apenas dados novos ou alterados, execução do processo em paralelo e em momentos de pouco ou nenhum uso do SGBD e a utilização de tabelas auxiliares que serão renomeadas como definitivas no fim do carregamento[8] [9].

### **2.2.3 Data Analysis – Cubos OLAP**

O termo Olap foi citado pela primeira vez por E.F.Codd [10], quando ele definiu doze regras que estas aplicações deveriam atender. A visão conceitual multidimensional dos negócios de uma empresa foi umas das regras citadas, a qual se tornou a característica fundamental no desenvolvimento destas aplicações. A visão multidimensional consiste de consultas que fornecem dados a respeito de medidas de desempenho, decompostas por uma ou mais dimensões dessas medidas. Podendo também serem filtradas pela dimensão e/ou pelo valor da Medida. As visões multidimensionais fornecem as técnicas básicas para cálculo e análise requeridos pelas aplicações de BI. Para se obter a visão multidimensional é necessário compreender outras características:

- Cubo é uma estrutura que armazena os dados de negócio em formato multidimensional, tornando-os mais fácil de analisar.
- Dimensão é uma unidade de análise que agrupa dados de negócio relacionados. As dimensões se tornam cabeçalho de colunas e linhas, como exemplo linhas de produto, regiões de venda ou períodos de tempo.
- Hierarquia é composta por todos os níveis de uma dimensão, podendo ser balanceada ou não. Na hierarquia balanceada os níveis mais baixo são equivalentes, porém, isto não ocorre nas hierarquias não balanceadas onde a equivalência hierárquica não existe. Por exemplo, em uma dimensão geográfica o nível país não possui o sub-nível Estado para um determinado membro e possui para outro. No caso específico pode-se citar o país Liechtenstein que não possui Estado e os EUA, que possui uma série de Estados.
- Membro é um subconjunto de uma dimensão. Cada nível hierárquico tem membros apropriados aquele nível. Por exemplo, em uma dimensão geográfica existe o nível e seus membros.
- Medida é uma dimensão especial utilizada para realizar comparações. Ela inclui membros tais como: custos, lucros ou taxas.

A aplicação OLAP soluciona o problema de síntese, análise e consolidação de dados, pois é o processamento analítico online dos dados. Tem capacidade de visualizações das informações a partir de muitas perspectivas diferentes, enquanto matem uma estrutura de dados adequada e eficiente.

A visualização é realizada em dados agregados, e não em dados operacionais porque a aplicação OLAP tem por finalidade apoiar os utilizadores finais a tomar decisões estratégicas. Os dados são apresentados em termos de medidas e dimensão, a maior parte das dimensões é hierárquica [11].

## **2.3 Plataformas de BI – Open Source**

### **2.3.1 Pentaho Open BI Suite**

A Pentaho Open BI Suite [12] é uma plataforma de BI, com facilidades para o utilizador final.

Como mostra a figura 2, a plataforma disponibiliza uma arquitectura orientada a serviços que inclui auditoria, segurança, escalonamento de tarefas, ETL, serviços de Web, repositório de atributos e regras de negócios.

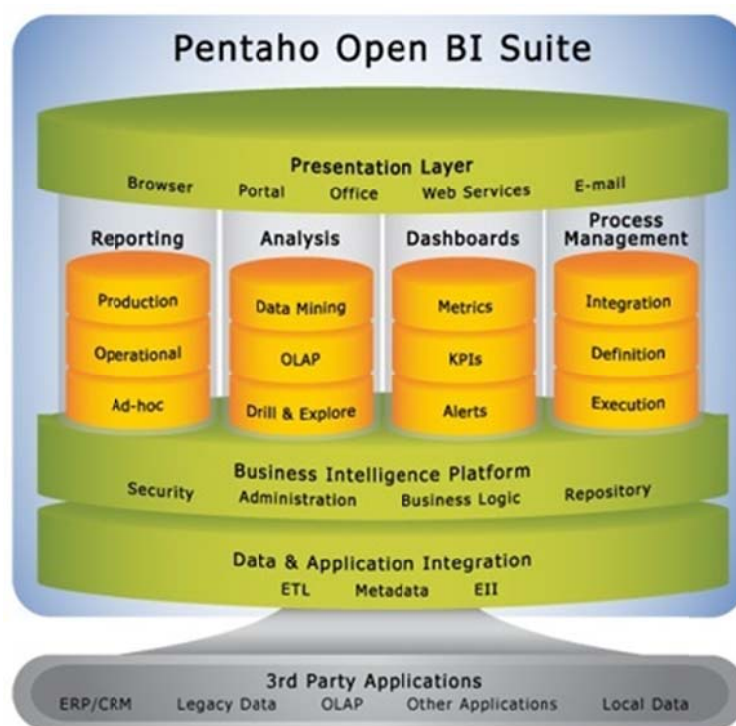


Figura 2 – Arquitectura Pentaho BI Suite

As funcionalidades para o utilizador final incluem relatórios, análise, workflow, dashboards e mineração de dados.

O Pentaho Design Studio é um conjunto de ferramentas para o projecto e gestão, que são integradas ao ambiente Eclipse. Estas ferramentas permitem ao analista de negócios tanto desenvolver como criar relatórios, dashboards, modelos de análise, regras de negócios e processos de BI.

As soluções de BI são projectadas usando o Pentaho Design Studio e desenvolvidas no Servidor Pentaho. O servidor Pentaho é o mecanismo de execução em tempo real, com uma abordagem a workflow, o qual coordena a execução e a comunicação entre os componentes de BI [11].

### 2.3.2 Talend Open Studio

O Talend Open Studio possui uma ferramenta para modelagem que auxilia o relacionamento das actividades de design com todas as etapas de um processo específico. A opção que realiza isto é a “Business Modeler”, e por meio de uma caixa de ferramentas muito intuitiva, o utilizador pode organizar todos os processos, garantindo que estas informações sejam recuperadas no futuro.

A opção “Job Designer” oferece uma interface gráfica e funcional para visualização dos processos integrados. Existem várias opções e componentes conectores e todos os processos podem ser integrados de forma simples: arrastando e soltando um componente para a área de criação, ou seja, permitindo trabalhar e

organizar os diagramas e os relacionamentos entre eles seleccionando e arrastando-os.

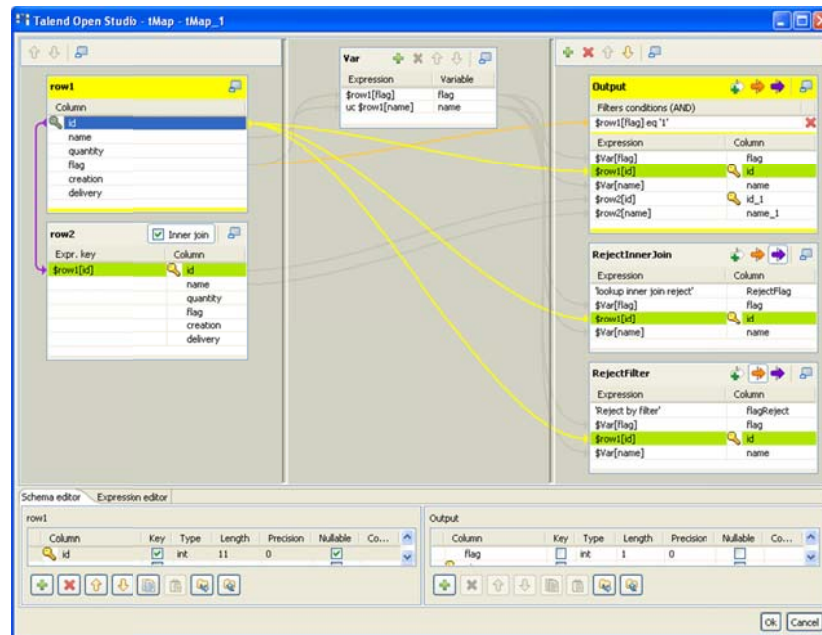


Figura 3 – Talend Open Studio

Como representado na figura 3, o Talend Open Studio oferece o “Component Library”, que é uma biblioteca com mais de 80 componentes e conectores onde se podem desenvolver mapas e funções de integração. Permite também estender as opções dos vários componentes criando padrões nas linguagens de Perl, Java ou SQL.

### 2.3.3 Avaliação

Apesar de ambas as ferramentas serem tecnologicamente bastante evoluídas, o Pentaho tem uma comunidade de utilizadores e comunidades bastante maiores. Estes utilizadores organizam-se em fóruns e partilham o seu conhecimento com o resto da comunidade.

Esse factor pesou muito na escolha da ferramenta Pentaho, uma vez que o suporte por parte da equipa da ferramenta é bastante diminuto, a partilha do conhecimento de todos é de um valor incalculável

## 2.4 Conclusões

O custo necessário para implementação de soluções de Business Intelligence com as ferramentas mais poderosas do mercado torna difícil e por vezes inviável a penetração das grandes empresas de consultoria nos segmentos de mercado mais baixos.

Esta situação leva a uma elevada dependência dos grandes projectos e faz com que as equipas se especializem quase exclusivamente nos produtos topo de gama, deixando de lado alternativas quase tão boas e bem mais baratas, neste caso, as ferramentas open source.

Neste momento já existem soluções bastante estáveis e com uma fiabilidade enorme, que permitem a criação de Sistemas de BI acessíveis a PME's e a SOHO's.

O investimento da PT-SI em tecnologias open source poderia dar a curto prazo uma vantagem competitiva no mercado de consultoria em BI, uma vez alargaria os seus targets podendo retirar valor de um segmento de mercado ainda pouco explorado



## 3. Caracterização do Problema

O Problema em questão enquadra-se na Medida 9 apresentada pela PT-SI. Uma medida que visa aumentar a taxa de êxito e o desenvolvimento de um datawarehouse único para a extracção de Indicadores.

### 3.1 Funcionalidades Pretendidas

Após o levantamento de requisitos junto da Direcção de Planeamento e Controlo e da área de Service Management Office da PT-SI, foi possível identificar as funcionalidades pretendidas e os requisitos a implementar.

Neste contexto pretende-se que sejam implementadas as seguintes funcionalidades:

#### 3.1.1 Repositório único – Datawarehouse (PT-SI)

Desenvolvimento de um repositório único para a PT-SI (DW PT-SI), que contemple a informação necessária para o cálculo de indicadores Financeiros, Operacionais, e de Cliente proveniente dos principais sistemas da PT-SI: EPM e CRM assim como do CRM da PT-PRIME. O objectivo é que o Repositório Único cumpra os requisitos especificados na tabela 3

Armazenamento da informação necessária para o cálculo dos indicadores especificados apenas num único local;
Uniformização dos conceitos existentes na PT-SI, de modo a permitir uma visão única dos indicadores, independentemente da fonte utilizada;
Cálculo de indicadores especificados de acordo com as perspectivas de análise identificadas

Tabela 3 – Requisitos do Datawarehouse

### 3.1.2 Cubo Analítico para exploração AD-HOC

É pretendido a criação de um cubo analítico que permita a extracção autónoma, por diferentes utilizadores, de indicadores existentes no DW PT-SI;

### 3.1.3 One Page Dashboard

Desenvolvimento de um Dashboard com os principais indicadores da PT-SI e da PT-PRIME, que permita obter uma visão global da empresa e a sua evolução ao longo do ano.

Os KPI's presentes no One Page Dashboard estão classificados por cada uma das seguintes Tipologias mencionadas na tabela 4.

Financeiros
Colaboradores
Clientes

**Tabela 4 – Tipologias dos Indicadores**

A nível de indicadores financeiros, considera-se todo o tipo de informação que dê a percepção de como anda a fluir o capital a nível dos negócios. Como por ex. A margem por oportunidade, os custos operacionais, o volume de negócios.

A nível de colaboradores, pretende-se saber as áreas onde cada colaborador vende mais, a margem que obtém em cada área e os custos associados. Dentro dessas áreas, pretende-se também saber quais os clientes principais, tanto a nível de volume de negócios, como os que dão mais benefícios à empresa, os que têm maior número de propostas adjudicadas, e também os que fazem com que a empresa gaste mais dinheiro.

### 3.1.4 Indicadores

#### 3.1.4.1 Indicadores Financeiros

Como está representado na tabela 5, os indicadores financeiros estão divididos em 3 categorias: Volume de Negócios, Margem e Custos, cada indicador terá duas vertentes. A vertente temporal, onde o utilizador poderá escolher o intervalo de tempo que deseja ver o indicador, sendo que o valor padrão é a janela temporal entre o início do Ano e o Mês em análise. A segunda vertente é o nível que o utilizador quer ver associado aos indicadores, poderá ir desde um segmento, a um mercado ou até mesmo uma empresa.

Indicador	Descrição
Volume de Negócios	Volume de Negócios acumulado entre Janeiro e o Mês em análise.
Margem	Margem operacional acumulada entre Janeiro e o Mês em análise.
Custos	Custos operacionais acumulados entre Janeiro e o Mês em análise.

**Tabela 5 – Indicadores Financeiros**

### 3.1.4.2 Indicadores Operacionais

Os indicadores Operacionais representados na tabela 6 dizem respeito ao nº de pedidos que foram feitos e o seu estado. Para cada pedido, pretende-se associar também os esforços realizado.

Nº pedidos	Número de pedidos de orçamento no âmbito do contrato
Pedidos em orçamentação	Número de pedidos no âmbito do contrato em fase de orçamentação, recepcionados no ano em Análise.
Esforço Realizado	Valor do esforço realizado nos pedidos no âmbito dos contratos / extra-contratos.

**Tabela 6 – Indicadores Operacionais**

### 3.1.4.3 Indicadores de Clientes/Parceiros

Os Indicadores de Clientes detalhados na tabela 7, mostram o valor que o cliente tem para a empresa, assim, o WinRatio diz-nos o rácio entre as propostas ganhas e as propostas perdidas, sendo que quanto maior for o WinRatio do Cliente, mais valor ele tem para a empresa. O índice de concretização, mostra o rácio entre as propostas ganhas e todas as propostas, mais uma vez, um índice elevado de concretização, significa um cliente de valor. Todos estes indicadores serão cruciais para uma análise detalhada do cliente em questão e do seu histórico com a PT-SI

Indicador	Descrição
WinRatio	Rácio das propostas ganhas sobre as propostas ganhas e perdidas, entre Janeiro e o mês em análise.
Índice de concretização	Rácio das propostas ganhas sobre as propostas ganhas, perdidas e sem resposta (em análise), entre Janeiro e o mês em análise.
Valor médio das propostas ganhas	Valor médio das propostas ganhas entre Janeiro e o mês em análise.
Valor médio das propostas perdidas	Valor médio das propostas perdidas entre Janeiro e o mês em análise.
Custos comerciais / VN serviços	Peso dos custos comerciais no total de Volume de Negócios de serviços.
Propostas ganhas	Peso das propostas ganhas no total de Volume de Negócios em plano.

**Tabela 7 – Indicadores de Clientes/Parceiros**

### 3.1.4.4 Indicadores PT-PRIME

A PT-PRIME serve de intermediário entre o cliente final e a PT-SI. Os indicadores pretendidos baseiam-se em margens de lucro, proveitos totais (o mesmo conceito que volume de negócios), Win Confidence (percentagem de propostas ganhas sobre o nº total de propostas), e percentagem de vendas nos vários âmbitos (Serviços, Software e Hardware).

## 3.2 Solução Apresentada

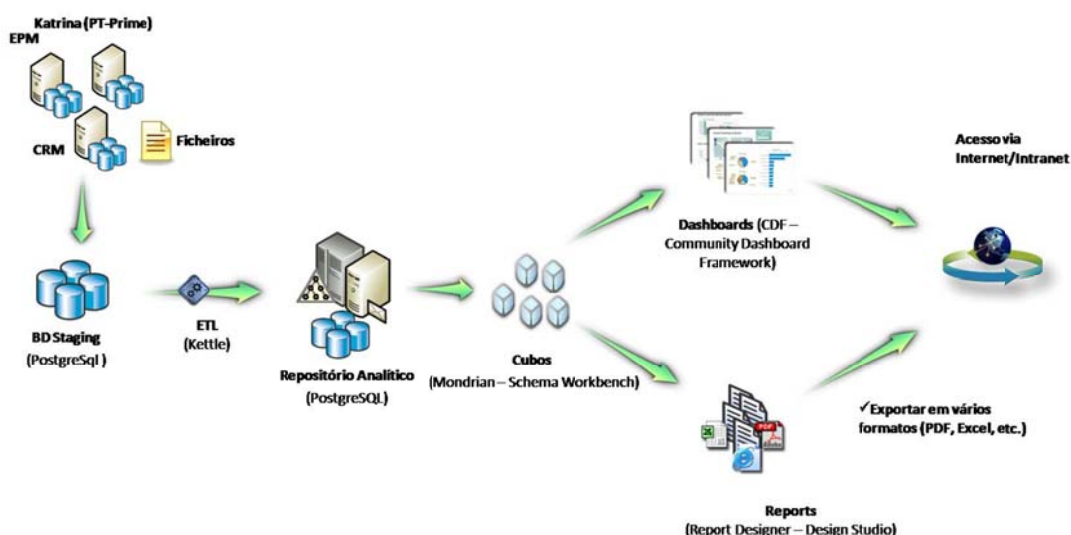


Figura 4 - Arquitectura da Solução Proposta

Depois de escolhidas as Tecnologias a utilizar, a arquitectura apresentada para a solução é a demonstrada na Figura 4, esta solução passa pela extracção da informação de forma integral para uma Staging Area em PostgreSQL, essa staging irá ser uma cópia da informação contida nos sistemas operacionais. Seguidamente serão executados os processos ETL (utilizando a ferramenta Kettle) e aplicadas todas as transformações necessárias ao tratamento, uniformização e convergência da informação pretendida. A fase seguinte é o carregamento do Repositório Analítico (datawarehouse), também este utilizando PostgreSQL.

Depois de alimentado o datawarehouse, passamos à próxima fase, o reporting. A solução adoptada solução passa pela criação de 2 cubos de informação (utilizando Schema-Workbench) que darão suporte a relatórios dinâmicos, pela criação de um dashboard (com recurso à ferramenta CDF) e pela modelação de vários relatórios estáticos (Report Designer e Design Studio) com variáveis de entrada. Finalmente, toda esta informação é disponibilizada via Web utilizando o Pentaho Bi Server.

Como mencionado no capítulo 1.2, depois da fase de análise funcional e técnica o projecto foi dividido em duas fases:

### 3.2.1 Modelação da Base de Dados e Processos ETL

Depois de avaliado o estado da arte, o SGBD escolhido para suportar o datawarehouse foi o PostgreSQL. Uma vez que a volumetria de dados é bastante reduzida e a velocidade de acesso aos dados é muito diminuta, a performance do SGBD deixa de ser prioritária. Também a seu favor, pesou o facto de suportar mais

funcionalidades em SQL, tais como Views e Procedures, que poderiam vir a ser utilizados durante o projecto.

O Modelo dimensional escolhido para a base de dados foi o modelo floco de neve (Snow Flake).

### 3.2.1.1 Snow Flake

No modelo Floco as tabelas dimensionais relacionam-se com a tabela de factos, mas algumas dimensões relacionam-se apenas entre elas, isto ocorre para fins de normalização das tabelas dimensionais, visando diminuir o espaço ocupado por estas tabelas, então informações como Categoria, Departamento e Marca tornaram-se tabelas de dimensões auxiliares[13].

No modelo Floco existem tabelas de dimensões auxiliares que normalizam as tabelas de dimensões principais. Na figura 5 estas tabelas são (Ano, Mês e Dia) que normalizam a Dimensão Tempo, (Categoria, Departamento e Marca) que normalizam a Dimensão Produto e a tabela Meio que normaliza a Dimensão Promoção.

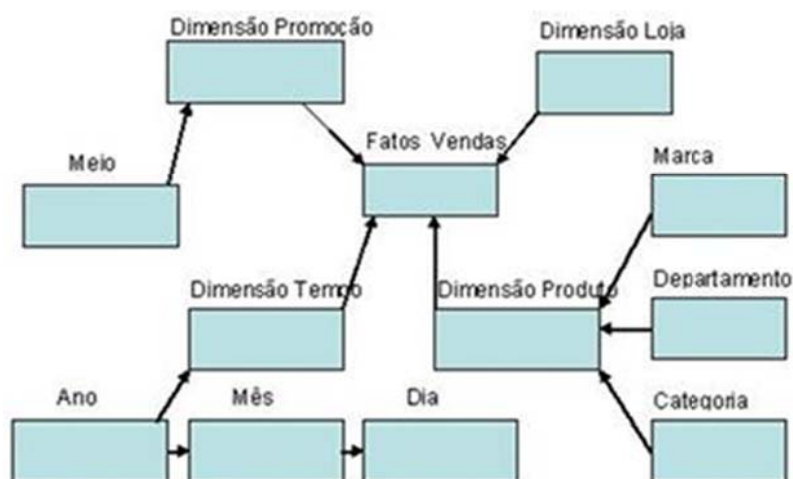


Figura 5 - Modelo Snow Flake

Construindo a base de dados desta forma, passamos a utilizar mais tabelas para representar as mesmas dimensões, mas ocupando um espaço em disco menor do que o modelo estrela. Este modelo chama-se floco de neve, pois cada dimensão se divide em várias outras tabelas, onde, organizadas de certa forma, lembram um floco de neve [13]. O modelo de dados proposto é composto por 3 tabelas factuais e 22 tabelas de referência.

Depois do modelo de dados estar construído, foi necessário construir os processos ETL, como referido neste capítulo existem 3 tipos de Fontes, CRM PT-SI, EPM PT-SI e CRM PT-PRIME. Devido à indisponibilidade de criação e disponibilização de ficheiros, a extracção teve de ser feita na Fonte. A ferramenta de ETL da plataforma escolhida (Pentaho) é o Kettle – Pentaho Data Integration.

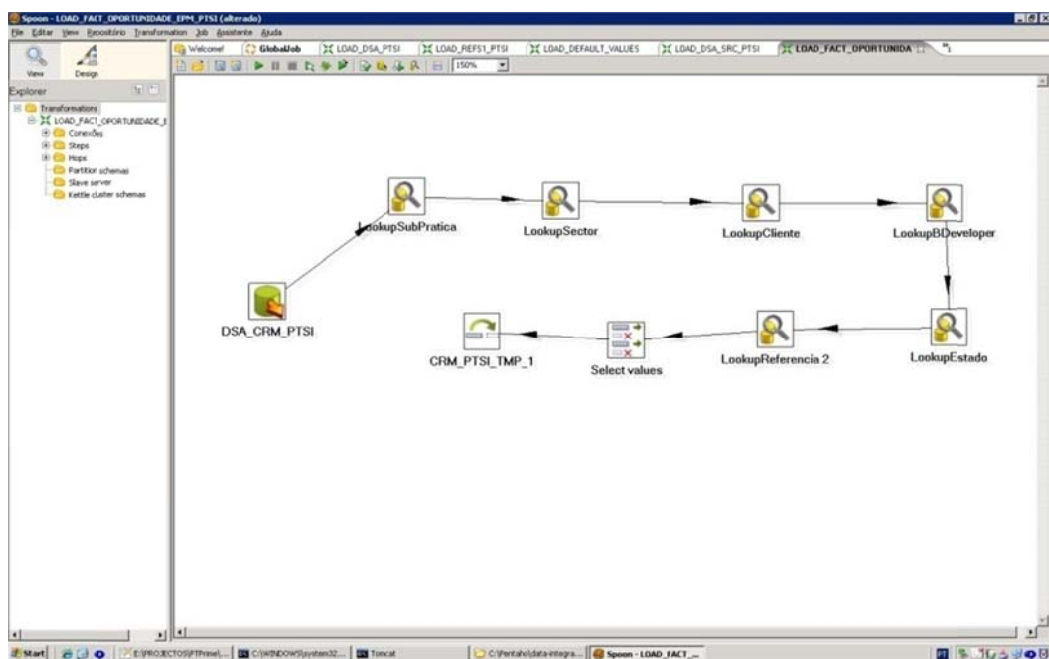
### 3.2.1.2 Kettle – Pentaho Data Integration.

O Kettle é um software desenvolvido pela Pentaho que gere o processo de extracção, transformação e integração dos dados no Datawarehouse. As suas principais características estão detalhadas na Tabela 8

<b>Integração de dados</b>	
<b>Acesso a dados</b>	Aplicações ERP e CRM. Sistemas Legacy. Bases de Dados.
<b>Optimizado para grandes quantidades de dados</b>	Caching sofisticado. Clustering SQL Optimizado.
<b>Simplificação de processos</b>	Completa integração de dados. Interface intuitivo e funcional Templates para as tarefas mais comuns

**Tabela 8 – Características do Kettle**

O Kettle é constituído por Jobs e Transformers. Um job pode ter várias transformações e invocar outros jobs. Com esta funcionalidade conseguimos encadear todo um processo de ETL desde o início até ao fim. Também é possível invocar um Job por linha de comandos, por isso a sua execução pode ser agendada.



**Figura 6 – Transformer Kettle**

Na figura 6 está representado um Transformer com diversos estados. No 1º estado, DSA\_CRM\_PTSI, é extraída da staging area a informação a ser tratada. Seguidamente, são feitos vários lookups para extrair o código bi gerado pelo nosso sistema de informação. Depois disso existe um estado onde são feitas as associações entre os campos origem e os campos destino e, finalmente, são inseridos os registos numa tabela.

## 3.2.2 Reporting

Para solucionar o problema dos Indicadores, foram criados vários tipos de relatórios. Relatórios dinâmicos utilizando a tecnologia OLAP, e relatórios estáticos, podendo o utilizador utilizar as variáveis pré-definidas.

As ferramentas utilizadas para a elaboração dos relatórios foram: Schema – Workbench (modelagem do Cubo Olap), e Report – Designer para criação e modelação dos relatórios estáticos o Design Studio, para adicionar aos relatórios parâmetros de entrada.

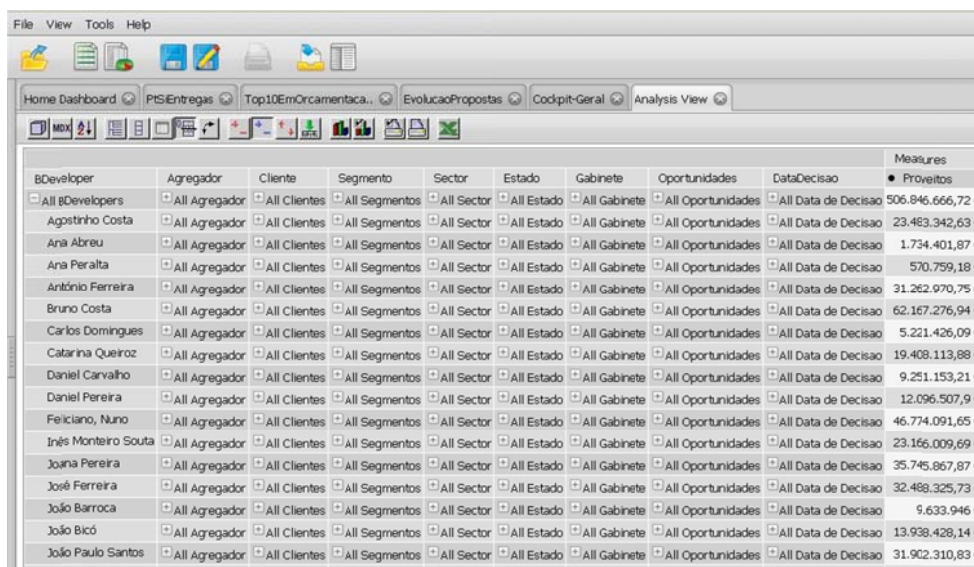
### 3.2.2.1 Schema – Workbench

O Schema Workbench é uma ferramenta da Pentaho para o desenho de cubos Olap. Nele define-se a metadata , as dimensões que se quer mostrar e as métricas.

Devido à limitação da ferramenta (no caso de criação de hierarquias) os ficheiros XML tiveram de ser editados, porém, quando abertos novamente por ela, ela reconhece o que foi alterado e não causa qualquer tipo de incompatibilidade.

A parte da definição do cubo revelou-se bastante importante. Depois de definida a metadata, as ligações entre tabelas factuais e tabelas de referência, e construção de hierarquias, foi possível testar a integridade dos dados e a performance do sistema OLAP.

O servidor OLAP utilizado é o Mondrian, desenvolvido também pela Pentaho e que permite as operações de navegação dos dados.



The screenshot shows the Schema Workbench application interface. The main window displays a data table with the following columns: BDeveloper, Agregador, Cliente, Segmento, Sector, Estado, Gabinete, Oportunidades, DataDecisao, and Measures. The Measures column is set to 'Proveitos'. The table lists 15 rows of data, each representing a developer and their associated metrics across various dimensions.

BDeveloper	Agregador	Cliente	Segmento	Sector	Estado	Gabinete	Oportunidades	DataDecisao	Measures
All BDevelopers	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	506.946.666,72 €
Agostinho Costa	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	23.483.342,63 €
Ana Abreu	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	1.734.401,87 €
Ana Peralta	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	570.759,18 €
António Ferreira	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	31.262.970,75 €
Bruno Costa	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	62.167.276,94 €
Carlos Domingues	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	5.221.426,09 €
Catarina Queiroz	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	19.408.113,88 €
Daniel Carvalho	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	9.251.153,21 €
Daniel Pereira	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	12.096.507,9 €
Feiciano, Nuno	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	46.774.091,65 €
Inês Monteiro Souta	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	23.166.009,69 €
Joana Pereira	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	35.745.867,87 €
José Ferreira	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	32.488.325,73 €
João Barroca	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	9.633.946 €
João Bico	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	13.938.428,14 €
João Paulo Santos	All Agregador	All Clientes	All Segmentos	All Sector	All Estado	All Gabinete	All Oportunidades	All Data de Decisao	31.902.310,83 €

Figura 7 - Schema Workbench

A figura 7 mostra o cubo Olap a ser processado pelo servidor Mondrian, temos várias dimensões e uma métrica, neste caso, proveitos. Todas as dimensões são expansíveis, e podemos alterar a sua ordem podendo assim ter uma visão diferente da informação.

O utilizador pode (de acordo com o cubo que escolheu):

- Gerar múltiplos relatórios.
- Exportá-los em vários formatos (PDF, XLS)
- Gerar gráficos a partir dos relatórios criados.

### 3.2.2.2 Report Designer – Design Studio

O Pentaho Report Designer é um gerador de relatórios da suite. Facilita a criação de relatórios com uma aparência mais profissional e personalizada. Adicionalmente permite a publicação desses relatórios no Pentaho Bi Server, permite inclusive que filtros simples sejam publicados directamente sem a criação de xactions no Pentaho Design Studio. Na figura 8 é visível o layout da ferramenta. Do lado esquerdo situam-se os objectos que se podem colocar no relatório, usando o tradicional, drag and drop. No centro situa-se a tela onde se modelará o relatório. Do lado direito é onde se configuram as conexões à base de dados ou ao cubo, assim como se definem os cabeçalhos e sub-relatórios

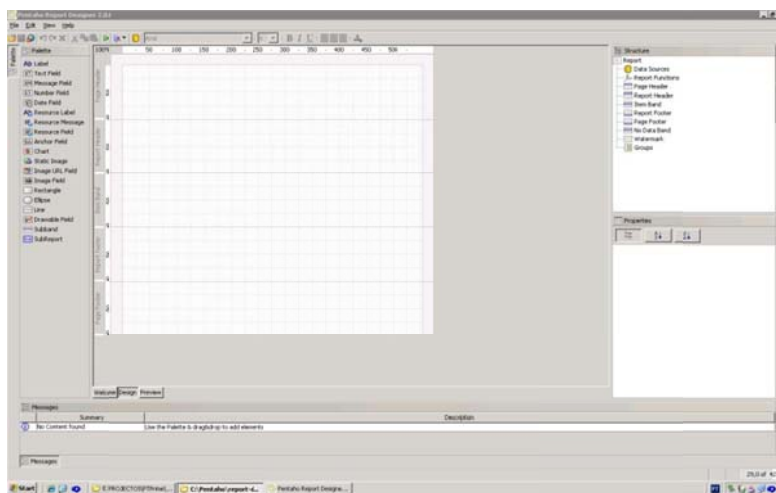


Figura 8 - Report Designer

Pentaho Design Studio – Plugin para a plataforma eclipse que funciona como uma interface gráfica para a criação de xactions, que são arquivos XML interpretados pelo BI Server. São como instruções "passo-a-passo" que dirigem a execução das actividades pelo BI Server, permitem, por exemplo, a definição de filtros avançados para relatórios e o envio de e-mails com os relatórios executados.

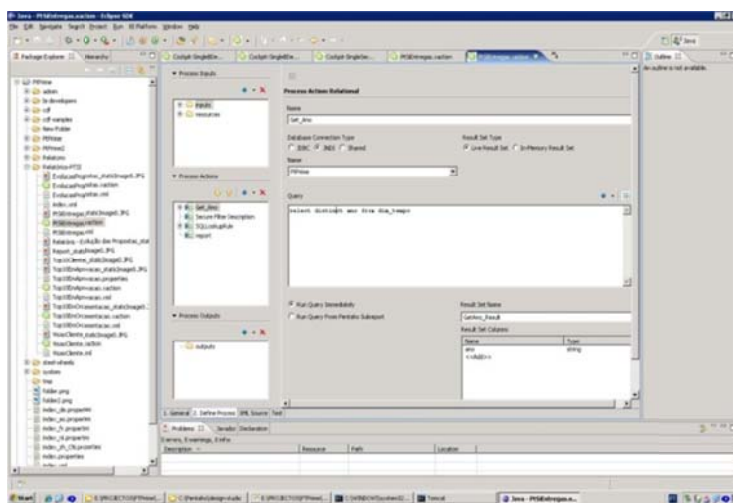


Figura 9 - Design Studio

A figura 9 mostra a área de trabalho da ferramenta Design Studio. Neste caso, mostra a definição de um filtro que pode ser criado de várias maneiras, uma delas, como exemplificado, é feita uma query que devolve todos os valores possíveis para o filtro. Esses valores vão ser passados para uma combo box. Quando o utilizador clicar no valor pretendido, esse é passado ao relatório e o filtro é aplicado.

### 3.2.2.3 One Page Dashboard

Dashboards são colecções de gráficos e/ou relatórios geralmente agrupados por tema para facilitar a visualização e compreensão das informações pelos utilizadores.

O Pentaho Community Edition possui um Framework de criação de dashboards chamado Community Dashboard Framework que facilita bastante a criação de dashboards. O CDF, porém requer de seus desenvolvedores conhecimento de HTML, Javascript, JQuery e CSS, o que pode dificultar a criação de dashboards por desenvolvedores menos experientes ou até mesmo por utilizadores comuns.

Devido às imensas dificuldades encontradas na criação dos Dashboards, os requisitos foram alterados, devendo o dashboard só mostrar dados financeiros, por sector (Figura 10) – dentro de cada sector por business developer (estes 1ºs em gráficos e valores absolutos de proveitos). Os detalhes das propostas dos business developer devem ser mostrados em forma de grid, podendo o utilizador visualizar o máximo detalhe possível.

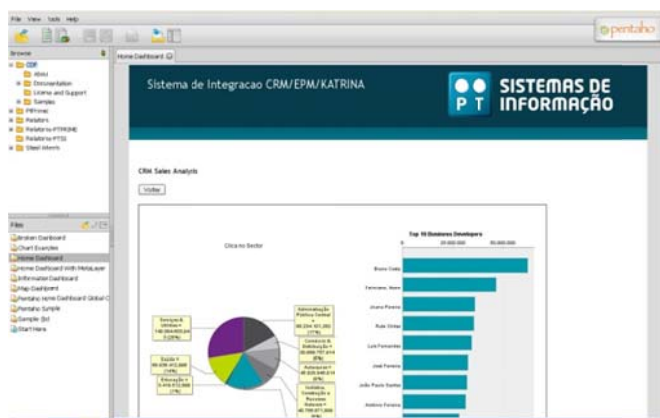


Figura 10 - Dashboard PT-SI

## 3.3 Conclusões

Numa primeira fase questionou-se se seria possível realmente implementar uma solução de BI, baseado unicamente em tecnologia Open Source. O estudo foi feito e a solução encontrada foi a que melhores respostas deu a todos os problemas levantados.

Contudo, o nível de dificuldade do projecto torna-se elevado, uma vez que não existe suporte para estas ferramentas. Todo e qualquer tipo de problema encontrado terá de ser solucionado com recurso a pesquisas ou ajudas em comunidades criadas em torno delas.



# 4. Implementação

## 4.1 Datawarehouse

Durante a fase de implementação do projecto, surgiram vários problemas. Para cada problema, foi identificado um conjunto de soluções, e escolhida a que melhor se enquadrava com o projecto.

Como o datawarehouse vai ser alimentado por 3 bases de dados diferentes, onde cada uma tem um período específico para a extração dos dados (devido à carga imposta pelos processos operacionais), era necessário criar uma área de suporte para onde essa informação será descarregada. Para tal, adoptou-se uma metodologia de Staging Area.

### 4.1.1 Staging Area

Uma Staging Area é um local onde a informação será guardada temporariamente. Essa informação corresponde à que se encontra nos sistemas operacionais, e não lhe é aplicada qualquer tipo de transformação

Normalmente, a staging area é criada por motivos temporais: quando a informação dos sistemas operacionais nos chega em períodos de tempo separados; quando há possibilidade de atrasos na disponibilização da informação; quando é impossível extrair toda a informação ao mesmo tempo do sistema operacional; É também uma maneira de guardarmos um histórico das nossas fontes de informação que se revela bastante útil quando é necessário algum tipo de reproprocessamento.

A staging area também permitirá a realização de testes ao ETL sem que seja necessário recorrer aos sistemas operacionais, sendo assim evita-se o repetido nº de acessos a estes, sendo que todos os aspectos dos processos de ETL se centram no datawarehouse.

A desvantagem desta solução, é o volume de dados guardados no nosso datawarehouse, mas, para isso, foi estipulado que as dsa's só iriam conter os últimos 3 dias de processamento. Caso seja preciso reproprocessar dias anteriores a este, as tabelas de staging devem ser de novo carregadas, e só depois se poderá executar os processos ETL.

O problema seguinte teve a ver com a estrutura das tabelas de referência. Para este caso, duas soluções foram propostas:

<b>Tabela de Referência</b>		
Código da Referência		
Código	Operacional	da Referência
Descrição da Referência		

**Tabela 9 – 1ª Solução**

O método apresentado na tabela 9 seria possível e era a melhor abordagem se os códigos operacionais para o mesmo elemento fossem iguais em todos os sistemas operacionais. Como essa situação não se verifica (Por ex. O business developer 'Carlos Nunes', tem o código operacional 1, na base de dados CRM PTSI, o código operacional 27 no sistema EPM PTSI e o código operacional 173 no CRM Katrina), esta abordagem foi descartada.

<b>Tabela de Referência</b>
Código da Referência
Descrição da Referência

<b>Tabela de Regras</b>
Sistema Fonte
Código Operacional
Código Referência

**Tabela 10 – 2ª Solução**

Esta segunda abordagem, apresentada na tabela 10, é a indicada para quando existem vários dados de vários sistemas diferentes, que têm de convergir numa única tabela. Para a construção da tabela de regras, é necessário fazer um estudo de todos os sistemas fonte, e ver os códigos operacionais para todos os elementos iguais que coabitam nele. Neste caso a tabela de regras estaria parametrizada de acordo com o especificado na tabela 11

<b>Sistema Fonte</b>	<b>Código Operacional</b>	<b>Código Referência</b>
CRM-PTSI	1	4
EPM-PTSI	27	4
CRM-PTPRIME	173	4

**Tabela 11 – Tabela de Regras**

O conteúdo da Tabela de Referência seria o que está representado na tabela 12

Código de Referência	Descrição de Referência
4	Carlos Nunes

**Tabela 12 – Tabela de Referência**

A tabela 10 indica-nos que se a fonte for CRM-PTSI então, quando o código Operacional é 1, o código da referência é o 4. Se o sistema fonte for EPM-PTSI e o código operacional for 27, então, o código de referência também é o 4, e a mesma regra se aplica para quando a fonte for CRM-PTPRIME, caso o código Operacional

seja 173, o Código de Referência também é o 4. Como mostrado na tabela 11, o código 4 corresponde ao Business Developer Carlos Nunes, assim como esperado.

A solução que mais se adequou ao nosso problema foi a 2ª, uma vez que, como demonstrado, era a única que conseguia resolver a questão da convergência de códigos

Ainda na fase de modelação de dados, e no que se refere às referências, foi detectado que algumas delas poderiam sofrer alterações ao longo do tempo, e, se não fossem tratadas de uma maneira diferente, poderiam ser responsáveis por relatórios incorrectos.

Voltando ao caso do Business Developer, todos os elementos desta referência têm uma área de negócio associada, mas, ao longo do tempo, ele pode mudar de área.

Problema: Em 2009 o business developer 'Carlos Nunes' está na área das PME's, e esteve envolvido em 4 projectos. Em 2010, ele mudou de área, e agora a área dele é SOHO. Neste ano, ainda só esteve envolvido num projecto. Se o modelo não for alterado, ao tirar um relatório dos projectos onde Carlos Nunes esteve envolvido nos últimos 2 anos e a que área corresponderam, este iria dar os 6 projectos, mas, a área iria ser sempre a mesma, uma vez que à data do relatório, só existia a área SOHO associada a Carlos Nunes.

Para resolver este problema, foram encontradas 3 soluções:

- Criação de um Histórico

A criação de um histórico, com a utilização do código do dia (ex. 20100101), permitiria que soubesse a que área pertenceu o Carlos Nunes em cada dia.

Esta solução resolvia o problema mas, ela é mais indicada para quando as alterações da referência são constantes.

- Inclusão na tabela factual da Área do Business Developer

Com a inclusão da área no business developer, na tabela factual do projecto, a associação entre um projecto e uma área deixaria de ser possível só através do business developer como passaria a ser directa.

Esta solução resolvia o problema com bastante eficiência, mas a inclusão de mais um campo na tabela factual não mostrou ser a melhor opção

- Utilização da Metodologia Slowly Changing Dimension

A metodologia Slowly Changing Dimension é utilizada para tabelas de referência onde existam poucas alterações. Neste caso, seriam adicionadas duas colunas, uma com a data inicial do estado e outra com a data final. Assim conseguíamos manter um histórico de todas as nossas alterações na dimensão, mas esse histórico só era actualizado quando houvesse alguma alteração nesta.

A solução que melhor se adequou ao problema foi a última, SCD, uma vez que mostrou ser superior às outras a nível de espaço ocupado. A nível de performance não mostrou ser inferior.

## 4.2 ETL

A fase de Extração, Transformação e Carregamento de dados foi dividida em 3 fases distintas. Em primeiro lugar foi feita a extração da informação para carregar as DSA's, seguidamente foram carregadas as tabelas de Referência e, finalmente, a informação foi transformada e carregada nas tabelas factuais.

## 4.2.1 – Extracção para DSA's

A extracção das DSA's é uma extracção integral da informação que está nos sistemas operacionais. A Figura 11 apresenta 2 tipos de extracções diferentes, a 1ª tem como fonte o sistema de CRM da PT-SI, a 2ª extrai informação de um ficheiro Excel, este contém todos os projectos a ser contemplados pelo sistema.

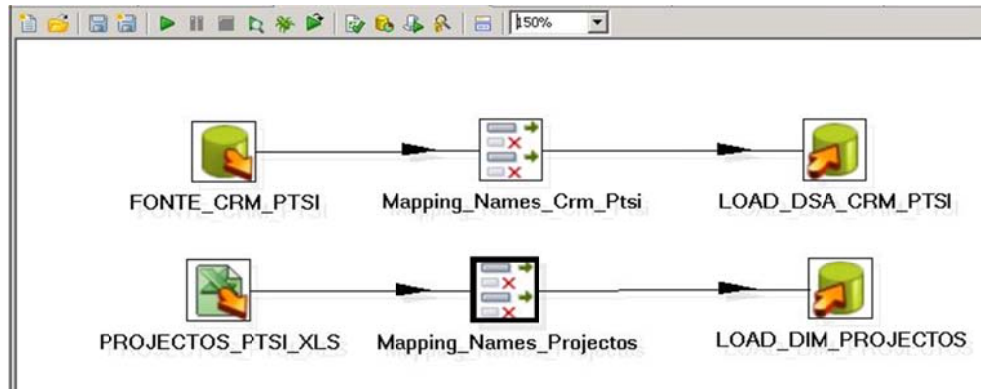


Figura 11 – Transformação para carregamento de DSA

A figura 12 mostra o interior do objecto FONTE\_CRM\_PTISI, a fonte de informação é um stored procedure que está criado na base de dados do CRM da PT-SI. A principal dificuldade na extracção dos dados desta forma deveu-se ao facto dos processos de ETL estarem a correr numa Virtual Machine e a base de dados do CRM estar noutra máquina. Foi preciso configurar a firewall para permitir a ligação à base de dados e criar um utilizador com permissões de execução sobre o stored procedure.

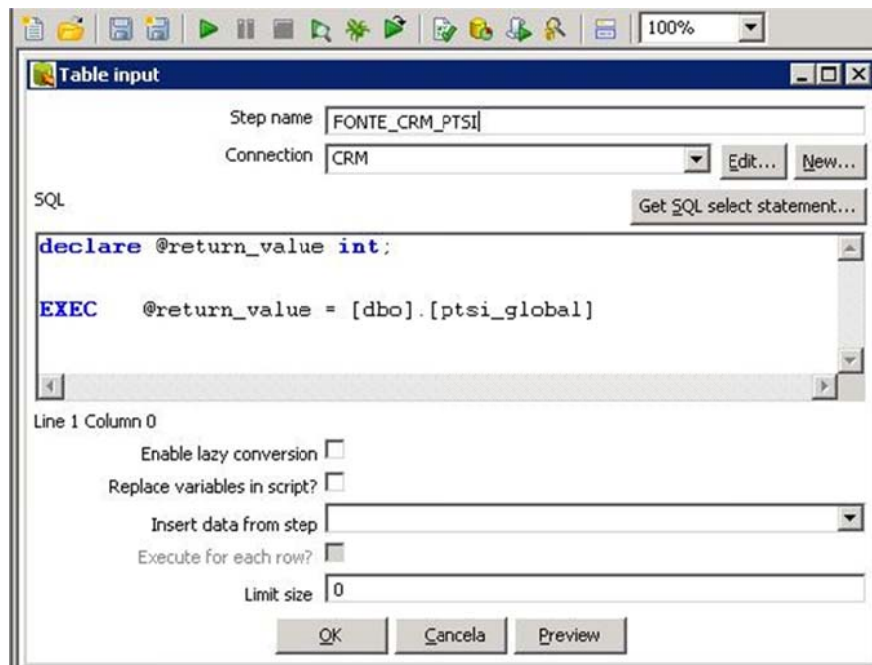


Figura 12 – Extracção de Stored Procedure

Na figura 13 é possível ver a maneira como é feita a extracção de informação de folhas de cálculo, neste caso, um ficheiro excel. Pode-se obter os campos do ficheiro de duas maneiras distintas: a 1ª, e caso o nosso ficheiro contenha uma linha de cabeçalho, carregando no botão “Get fields from header row”, além de extrair o nome das colunas, tenta identificar o tipo de informação contido nelas. A 2ª é escrevendo manualmente o nome das colunas, tipo, tamanho e precisão. Este objecto também permite formatar os campos a serem extraídos o que o torna um bom auxílio à uniformização da informação

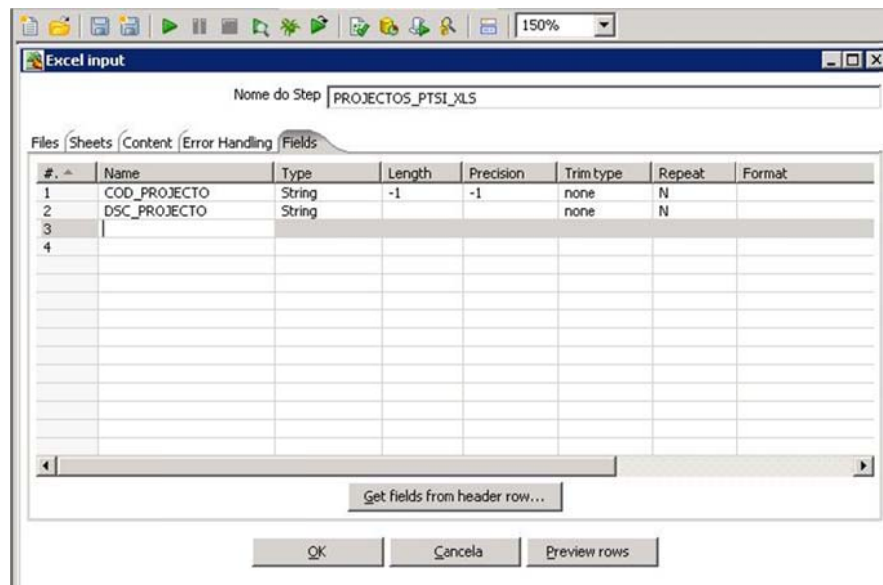


Figura 13 – Extracção de Excel

Depois de extraída a informação, é preciso mapeá-la para a dsa do destino, ou seja, fazer corresponder cada campo na fonte, a um campo no destino, isso é representado na figura 14. Para facilitar o mapeamento, caso seja possível, pode-se fazer um levantamento do nome dos campos no destino e ao extrair a informação, extrair já com esses nomes, desse modo, a ferramenta consegue automaticamente associar o campo destino à ao campo fonte.

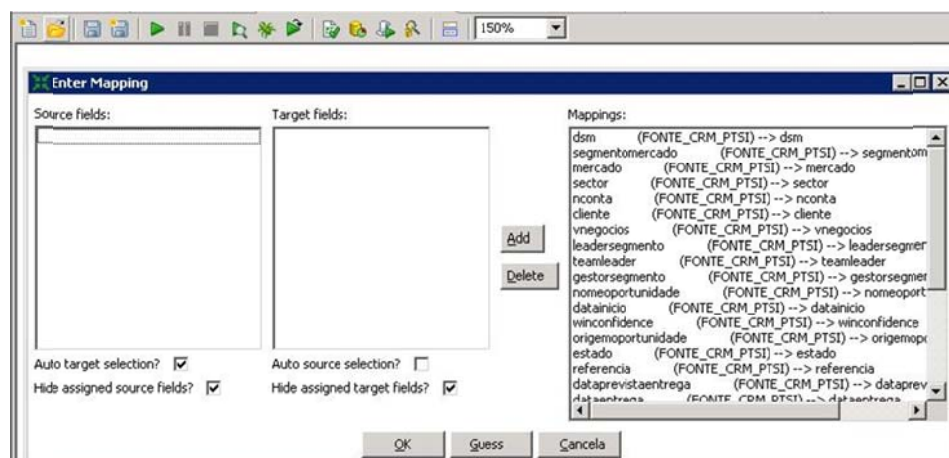


Figura 14 – Mapeamento dos campos

## 4.2.2 – Carregamento das tabelas de Referência

A extração e carregamento das tabelas de referências foram divididos em 3. Tabelas com conceitos únicos que não sofrem alterações. Tabelas com conceitos diversificados nos 3 sistemas que precisam de uma regra de convergência e Tabelas que precisam de guardar histórico uma vez que podem ser alteradas ao longo do tempo (SCD). Estes 3 tipos estão exemplificados na Figura 15.

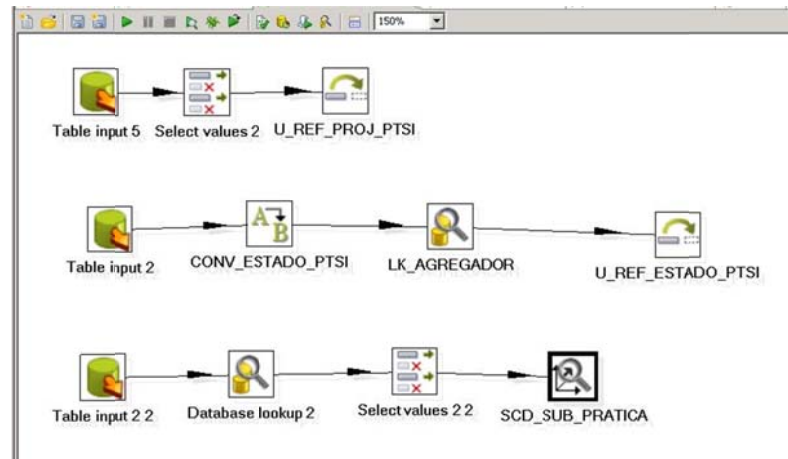


Figura 15 – Carregamento Referências

O 1º caso, é um carregamento simples da tabela de referência. Os dados são tratados no momento de extração, mapeados e é feita a inserção ou update, de acordo com a sua chave. Cada objecto poderá mudar no máximo o seu descritivo.

No 2º caso, temos uma tabela de referência onde é necessária a convergência de dados. Inicialmente foi planeado o uso de uma tabela de regras, como proposto no capítulo 4.1, mas, depois de comprovada a eficácia do operador de mapeamento, presente na figura 16, optou-se pelo uso deste

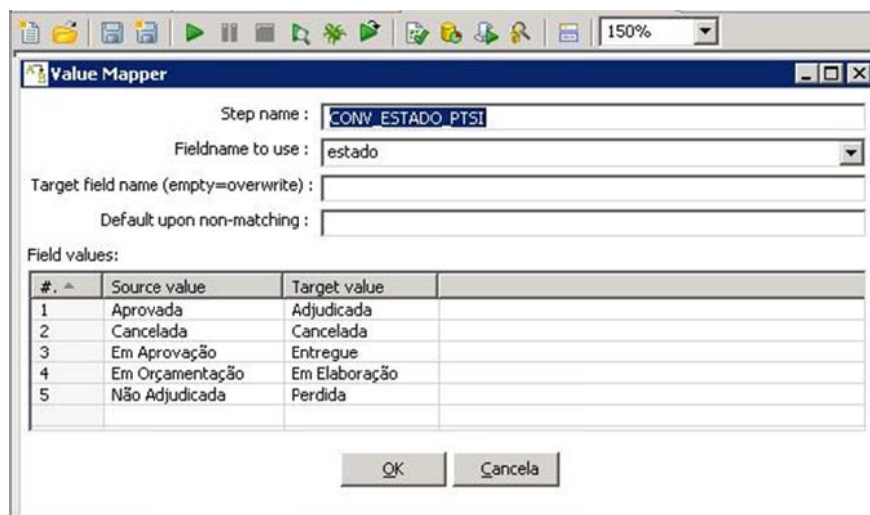


Figura 16 – Operador de Mapeamento

Como mostra a Figura 16, para cada valor de entrada, corresponde o valor de saída, é este valor que, posteriormente, irá fazer lookup na tabela de referência.

Na Figura 17 podemos ver como funciona o operador Lookup. Para cada valor de entrada (estado), ele compara com o tipo\_estado na tabela agregador, e devolve o id\_agregador com o novo nome agregador e do tipo Integer.

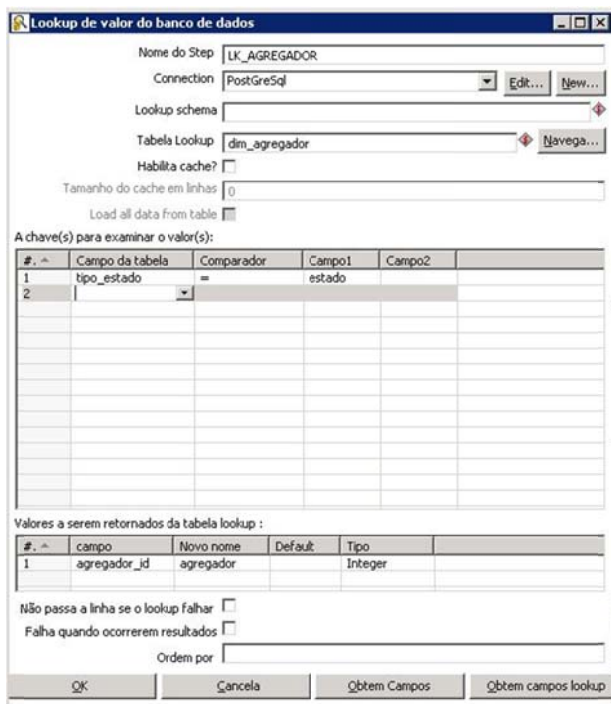


Figura 17 – Operador Lookup

As tabelas de referência em que os seus registos podem sofrer alterações ao longo do tempo (não contemplando alterações na sua descrição), são conhecidas como Slowly Changing Dimensions. Para o carregamento destas foi usado o Tipo 2 (contempla a data inicial e final do estado), sendo suportado pelo operador Dimension Lookup / Update representado na Figura 18. Neste operador, pode-se definir o campo para comparação (lookup), neste caso foi o código operacional da Sub-Prática, COD\_OP\_SUBPRATICA, este campo vem da fonte com o nome SUBPRATICA, também se definiu a chave a ser incrementada caso o COD\_OP inserido seja novo, definiu-se o campo da data, neste caso o COD\_DIA e, em “Fields”, definiu-se os campos que serão actualizados e o tipo de actualização. No caso da sub-pratica, interessa saber se a sua prática associada é a mesma, sempre que mudar, tem de ser alterada e guardada em histórico. É a este campo que se associa a actualização slowly changing dimensions tipo 2.

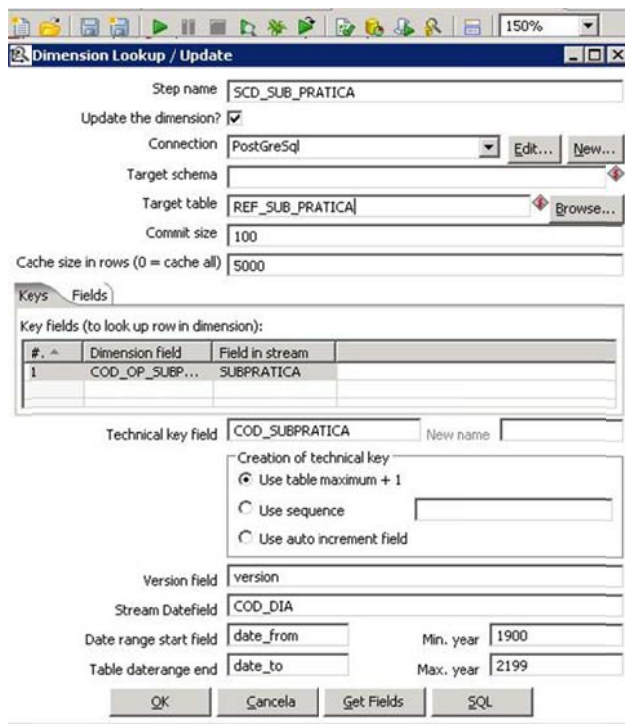


Figura 18 – Slowly Changing Dimensions

O carregamento das tabelas de referência foi dificultado pelas fontes de informação. Estas só continham as tabelas factuais e todas as tabelas de referência tiveram de ser criadas a partir delas. No entanto, e através de todos os operadores fornecidos pelo Kettle, as dificuldades foram ultrapassadas e foi conseguido, de uma forma eficiente, o carregamento de todas as tabelas.

### 4.2.3 – Transformação e carregamento das tabelas Factuais

O carregamento das tabelas factuais teve por base transformações feitas a partir das DSA's e lookups às tabelas de referência. Foi criado um Job Global, apresentado na Figura 19

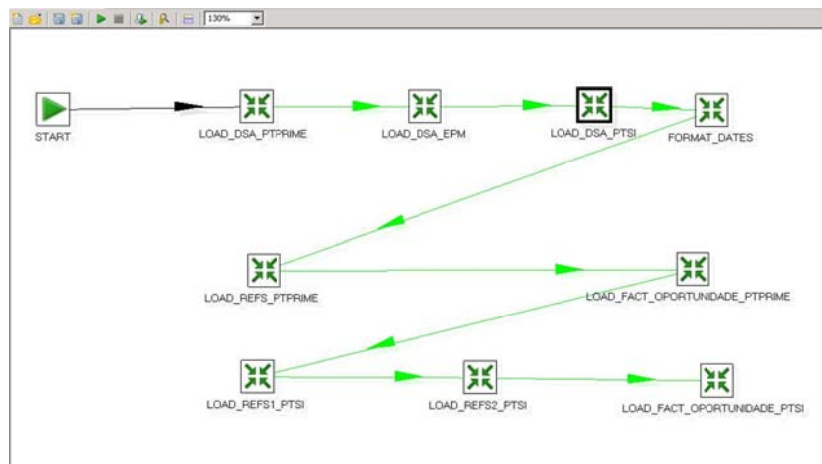


Figura 19 – Job Global de Carregamento

O operador Start inicia o processo de ETL. Seguidamente é extraída a informação dos 3 sistemas distintos e carregada nas DSA's. Posteriormente existe uma transformação para uniformização de todas as datas do sistema para o formato YYYYMMDD hh24:mi:ss. Segue-se o carregamento das referências de suporte à tabela factual PT-PRIME e o seu carregamento da própria factual. O job continua com o carregamento das tabelas de referência da PTSI do 1º nível ou nível pai, seguindo-se o carregamento das do nível inferior, finalizando com o carregamento das tabelas factuais da PT-SI.

Embora a volumetria de dados seja um pouco pequena, 20K registos, a performance do sistema mostrou-se bastante boa sendo que o tempo total de processamento não ultrapassou os 5 minutos.

## 4.3 Reporting

Depois do Datawarehouse alimentado, foram criados os indicadores pretendidos em forma de relatórios estáticos, cubos olap e dashboard.

### 4.3.1 Relatórios estáticos

A ferramenta utilizada para a criação dos relatórios estáticos, já mencionada no capítulo 3.2.2.2, foi o Report Designer com o auxílio do Design Studio. O relatório apresentado na Figura 20 mostra a evolução das propostas ao longo dos anos, tanto a nível de número de propostas, como a nível de custos da elaboração das mesmas e finalmente o preço de venda ao público

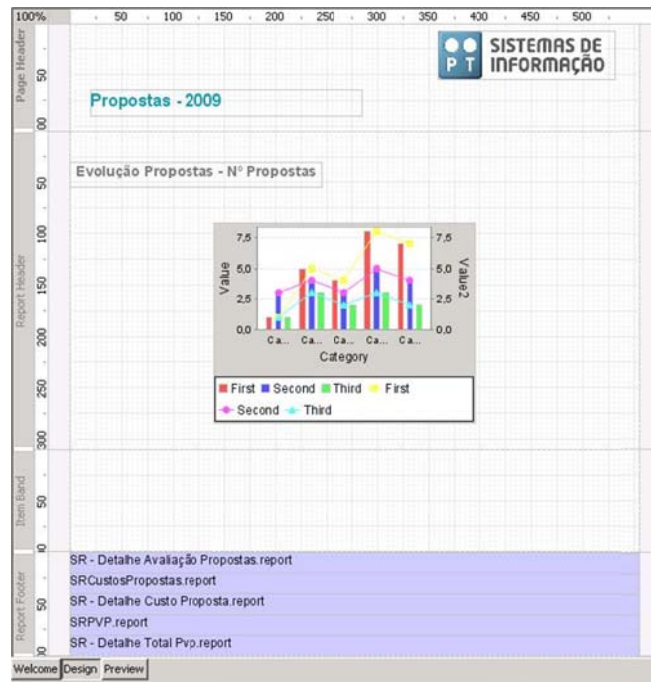
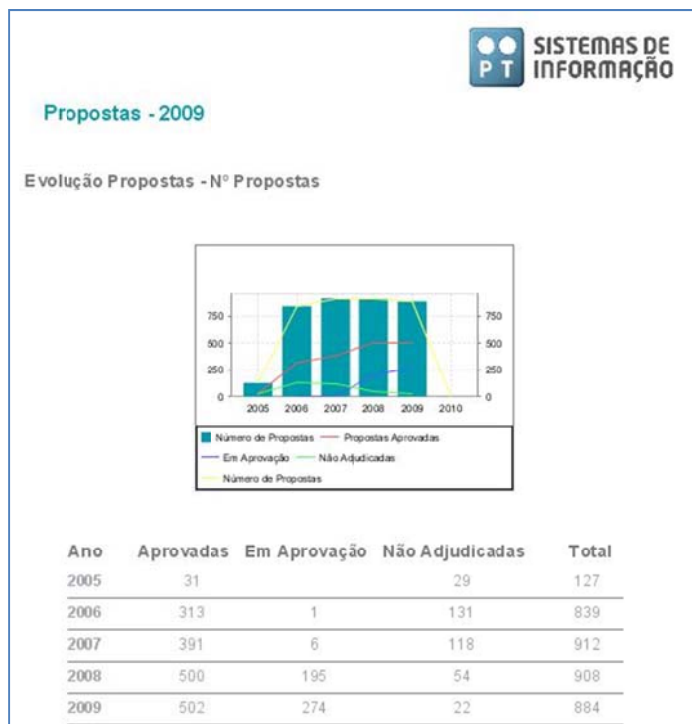


Figura 20 - Evolução de Propostas

Este relatório gerou 3 ficheiros, um Report , um XML e um Xaction. O ficheiro XML é enviado para o servidor e serve como modelo do relatório. Já o ficheiro Xaction

contém as “acções” que serão executadas na abertura do relatório, é nele que estão configuradas as variáveis de entrada (prompts). Neste caso, a única variável que o utilizador pode escolher é o tipo de formato que quer o relatório.

Na imagem 21 podemos ver o relatório depois de processado



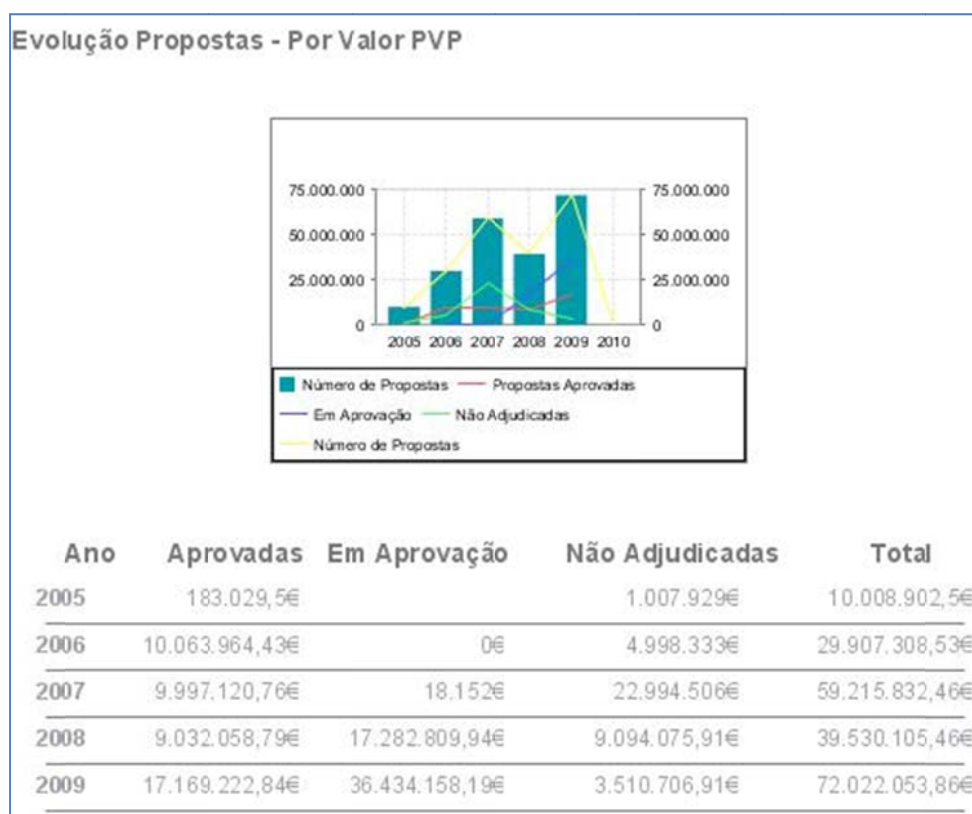
**Figura 21 - N° de Propostas**

Depois de analisado o relatório, observa-se que apesar do nº total de propostas em 2009 ser menor que 2008, o nº de propostas aprovadas foi ligeiramente superior, o que indica que o ratio propostas ganhas / propostas não ganhas é maior, logo é um indicador positivo para a empresa. Para analisar melhor esse indicador, o mesmo relatório também disponibiliza o valor gasto na elaboração das propostas. A Figura 22 mostra que o valor gasto na elaboração de propostas em 2009 foi quase o dobro do valor gasto que em 2008 e o nº de propostas é menor. O custo de elaboração de propostas em 2007 é quase nulo uma vez que o EPM da PT-SI onde estão guardados os custos das propostas só tem um histórico de 2 anos.



**Figura 22 - Custo de Elaboração de Propostas**

No entanto falta verificar o preço de venda ao público dessas propostas. Esses dados são mostrados na figura 23, onde se conclui que o valor de preço de venda ao público das propostas aprovadas e em aprovação em 2009 são bastante superiores aos registados em 2008.



**Figura 23 - Propostas por PVP**

Uma análise mais detalhada poderia ser feita caso o sistema tivesse acesso às ordens de SAP onde estão detalhados os custos com os recursos. Assim sendo poderia ser feita a comparação entre o lucro gerado em 2008 e o lucro gerado em 2009.

### 4.3.2 Cubos Olap

Com a criação dos 2 cubos, os utilizadores podem criar os seus relatórios dinâmicos via web, relatórios esses que possuem imensas funcionalidades: podem manusear grelhas adicionando dimensões e métricas; criar gráficos; fazer drill up e drill down nos diversos níveis de informação; extrair os relatórios no formato pretendido. O cubo pode ser gerado de duas maneiras: utilizando o Schema Workbench (representado na Figura 24); editando um ficheiro XML. O ficheiro XML contém a modelação do Cubo (definição da metadata). É nele que estão contidas as associações entre a tabela factual e as dimensões, e o cálculo das métricas

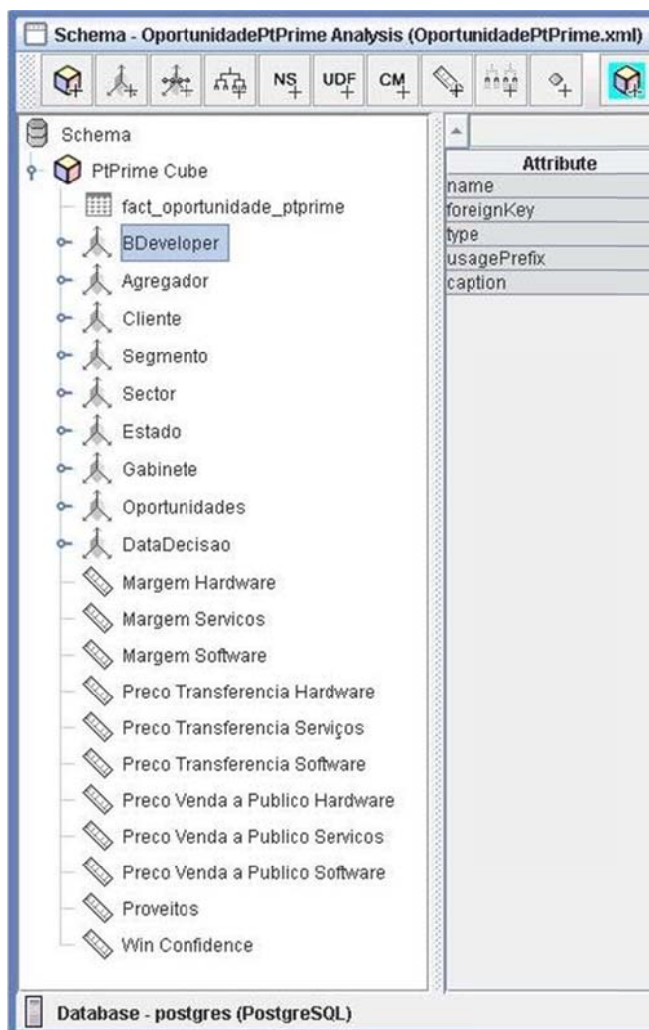


Figura 24 - Cubo PT-PRIME

O cubo PT-PRIME é constituído por uma tabela factual, suportada por nove dimensões, e onze métricas. Ele permite a associação entre todas as dimensões, mostrando os valores das métricas que o utilizador desejar.

Sector	Cliente		Measures				
Category	(All)	Cliente	● Hardware	● Servicos	● Software	▼ Proveitos	● Win Confidence
Comércio & Distribuição	All Clientes		€4.475.131,36	€3.908.266,56	€1.054.051,94	€9.437.449,86	€44,808
	All Clientes	Auchan	€1.039.212	€171.083	€	€1.210.295	€20
		Moviflor	€555.268,05	€207.669,17	€173.329,65	€936.266,87	€59,412
		Pap.Fernandes	€53.410,52	€807.516,04	€56.119,59	€917.046,15	€20
		Transdev	€717.876	€	€	€717.876	€0
		Urbanos	€125.974,38	€462.670,62	€68.893,06	€657.538,06	€70

Slicer: [(All)=All Agregador] [Ano=2009]

Figura 25 - Top 5 Clientes

Drill Through Table for Proveitos												
BDeveloper	Tipo	Cliente	Segmento	Category	Estado	Gabinete	Descricao	Ano	Mes	Dia	Proveitos	
Feliciano, Nuno	Ganha	Urbanos	Logística e Distribuição	Comércio & Distribuição	Adjudicada	SNE	Abertura automática de contas analíticas	2009	2	16	3.072,00	
Feliciano, Nuno	Ganha	Urbanos	Logística e Distribuição	Comércio & Distribuição	Adjudicada	SNE	Adicionais à Solução de Printing	2009	12	31	25.778,88	
Feliciano, Nuno	Ganha	Urbanos	Logística e Distribuição	Comércio & Distribuição	Adjudicada	SNE	Adicional Solução Hosting	2009	5	21	109.609,02	
Feliciano, Nuno	Ganha	Urbanos	Logística e Distribuição	Comércio & Distribuição	Adjudicada	SNE	CCTV adicionais Albufeira	2009	4	27	1.427,67	
Feliciano, Nuno	Ganha	Urbanos	Logística e Distribuição	Comércio & Distribuição	Adjudicada	SNE	CCTV adicionais LBP	2009	4	27	6.325,95	
Feliciano, Nuno	Ganha	Urbanos	Logística e Distribuição	Comércio & Distribuição	Adjudicada	SNE	CCTV adicional Urbanos Expresso	2009	4	27	6.938,04	
Feliciano, Nuno	Ganha	Urbanos	Logística e Distribuição	Comércio & Distribuição	Adjudicada	SNE	CCTV para o site da Mealhada	2009	5	5	9.774,75	
Feliciano, Nuno	Ganha	Urbanos	Logística e Distribuição	Comércio & Distribuição	Adjudicada	SNE	CCTV Trofa	2009	4	27	7.512,82	
Feliciano, Nuno	Ganha	Urbanos	Logística e Distribuição	Comércio & Distribuição	Adjudicada	SNE	Expansão da solução call manager	2009	4	27	18.073,45	
Feliciano, Nuno	Ganha	Urbanos	Logística e Distribuição	Comércio & Distribuição	Adjudicada	SNE	Fornecimento de Protáteis	2009	12	31	15.551,64	

Figura 26 - Detalhe Top 5

Na Figura 25 é apresentado o Top 5 de clientes (em proveitos) do Sector Comércio & Distribuição, no ano de 2009. Podemos ver também o que é que foi gasto em Hardware, Serviços e Software assim como o Win Confidence desse Cliente. Com a opção de Drill Down activa, o utilizador pode ver o detalhe das oportunidades do cliente. (Figura 26)

### 4.3.2 OnePage Dashboard

Foi criado um dashboard (Figura 27) que tem os vários sectores no gráfico da esquerda e o Top 10 dos business developer por sector no lado direito. Sempre que se clicar em algum sector, o código deste será enviado para o gráfico da direita e os seus valores serão actualizados. Existe também uma grelha de valores (iguais ao da Figura 25) que listam as oportunidades por sector e por business developer. Este dashboard é criado através de um ficheiro HTML, onde são definidos os objectos (gráficos e grelha), um ficheiro XML onde é definido o modelo do dashboard e um ficheiro xaction que controla a transferência de informação entre objectos.

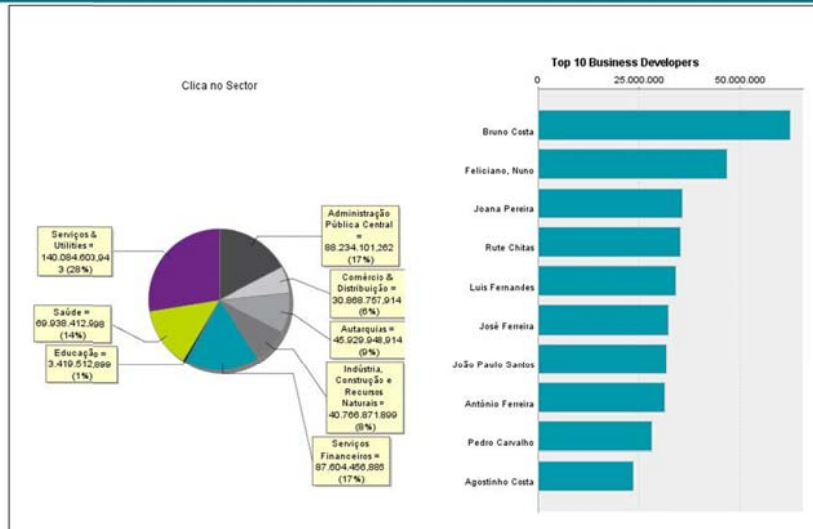


Figura 27 – Dashboard

## 4.4 Conclusões

Todas as opções tomadas na implementação mostraram-se capazes de solucionar o problema descrito.

Apesar de tudo ter sido solucionado, a implementação acabou por se tornar um pouco mais complicada do que se previa, uma vez que a tecnologia adoptada era algo novo, ainda a sofrer alterações por parte da equipa de desenvolvimento para correcção de bugs, e sem suporte, a não ser o já citados fóruns de discussão.

## **5. Conclusões e Trabalho Futuro**

### **5.1 Satisfação dos Objectivos**

Uma das características deste projecto foi a possibilidade dada ao autor de fazer um estudo sobre a problemática envolvente, de discutir, propor soluções, permitindo encontrar alternativas e elevando o conhecimento pessoal. A descoberta da ferramenta Pentaho Bi Suite com todas as suas potencialidades é um exemplo que foi fruto dessa característica

Considerando as dificuldades inerentes ao projecto, os conhecimentos do autor relativamente à área envolvida, todo o estudo que foi necessário para o desenvolvimento do projecto e tendo em atenção os objectivos propostos no início do mesmo, pode-se concluir que estes foram todos concretizados e que o trabalho realizado é satisfatório.

### **5.2 Trabalho Futuro**

Apesar da maioria das funcionalidades pretendidas pela PT-SI estarem implementadas, já estão detectadas lacunas e melhorias a ser efectuadas. A integração com SAP e Remedy é essencial para uma monitorização mais pormenorizada de todos os projectos que circulam na PT-SI. A partir das ordens de SAP, consegue-se saber o custo real de um projecto , possibilitando uma nova visão de todas as áreas envolvidas em cada proposta. O acesso ao remedy irá permitir monitorizar a manutenção correctiva, área essa que ainda não é contemplada no CRM e no EPM da PT-SI.

## Referências

- [1] - PT - Sistemas de Informação, [Online] Disponível em [http://www.ptsi.pt/PTSI/quem\\_somos.html](http://www.ptsi.pt/PTSI/quem_somos.html) , [consultado em 2010-02]
- [2] - Stallman, Richard (N.D.). "Richard Stallman's 1983 biography"
- [3] - Stallman, Richard (1983-09-27). "Initial GNU announcement"
- [4] - The FreeBSD Project , [Online] Disponível em <http://www.freebsd.org> , [consultado em 2010-02]
- [5] - Open Sources: Voices from the Open Source Revolution, O'Reilly 1st edition, January 1999
- [6] - Gartner Group, [Online] Disponível em <http://www.gartner.com/>, [consultado em 2010-02]
- [7] - Sergio Delfino, [Online] Disponível em [http://galileu.fundanet.br/jornada/artigos/computacao/Sergio\\_Delfino.pdf](http://galileu.fundanet.br/jornada/artigos/computacao/Sergio_Delfino.pdf), [consultado em 2010-02]
- [8] - KIMBALL, Ralph. Data Warehouse Toolkit. Tradução Mônica Rosemberg; Revisão Técnica Ronal Stevis Cassiolato. São Paulo: Makron Books, 1998.
- [9] - Fábio Silva Gomes da Gama e Abreu- FSMA, Revista de Sistemas de Informação, nº 2 Jul/Dez 2008, [Online] Disponível em [http://www.fsma.edu.br/si/Artigos/V2\\_Artigo1.pdf](http://www.fsma.edu.br/si/Artigos/V2_Artigo1.pdf), [consultado em 2010/02]
- [10] - Codd, Edgar Frank: "Is Your DBMS Really Relational?", ComputerWorld, 14. October 1985
- [11] Cynthia Aurora Anzanello, Olap Conceitos e Utilização, [Online] Disponível em [http://svn.assembla.com/svn/odinIDS/Egio/artigos/SolucoesIA/Datamining/artigo\\_cynthia.pdf](http://svn.assembla.com/svn/odinIDS/Egio/artigos/SolucoesIA/Datamining/artigo_cynthia.pdf), [consultado em 2010-02]
- [12] - Pentaho, [Online], Disponível em <http://www.pentaho.com/>, [consultao em 2010-02]
- [13] - Eduardo Moreira, Modelo Dimensional para Datawarehouse, [Online] disponível em [http://imasters.uol.com.br/artigo/3836/modelo\\_dimensional\\_para\\_data\\_warehouse](http://imasters.uol.com.br/artigo/3836/modelo_dimensional_para_data_warehouse) , [consultado 2010-02]