

Autor: Carlos Miguel Inácio Camejo

Orientador: Professor Doutor Carlos Pampulim Caldeira

**Desenvolvimento de um Sistema de Data Warehousing para a
Consolidação da Informação num Grupo Empresarial**

Universidade de Évora

2010

Universidade de Évora
Escola de Ciências e Tecnologia
Departamento de Informática

**Desenvolvimento de um Sistema de Data Warehousing para a
Consolidação da Informação num Grupo Empresarial**

Esta dissertação foi submetida ao Departamento de Informática da Universidade de Évora com vista à obtenção do grau de Mestre em Engenharia Informática.



177 948

Autor: Carlos Miguel Inácio Camejo

Orientador: Professor Doutor Carlos Pampulim Caldeira

Novembro 2010

Tabela de Conteúdo

Lista de Figuras.....	iii
Lista de Tabelas	iv
Resumo.....	v
Abstract	vi
1. Introdução	1
1.1. Motivação e Objectivos.....	1
1.2. Estrutura do Documento.....	3
2. Data Warehouse Empresarial vs Data Warehouse baseado em Data Marts.....	5
3. Online Transactional Processing (OLTP) vs Online Analytical Processing (OLAP).....	7
4. Data Warehouse.....	9
4.1. O que é?	9
4.2. Objectivos.....	11
4.3. Componentes	12
4.3.1. Sistemas Operacionais	14
4.3.2. Área de Estágio.....	14
4.3.3. Apresentação dos Dados.....	15
4.3.4. Acesso aos Dados	15
4.4. Metadados	16
4.5. Concepção do Data Warehouse.....	17
4.5.1. Modelos Dimensionais	18
4.5.1.1. Esquema em Estrela	18
4.5.1.2. Esquema em Floco de Neve	20
4.5.2. Processo de Negócio	21
4.5.3. Matriz em Bus	21
5. Extracção, Transformação e Carregamento.....	23
6. Tabelas de Dimensão	25
6.1. Surrogate Key	26
6.2. Dimensões Planas e Dimensões em Floco de Neve	27
6.3. Alterações nos valores dos atributos	28
6.3.1. Alterações Tipo I.....	28
6.3.2. Alterações Tipo II.....	29
6.3.3. Alterações Tipo III.....	29

6.4.	Dimensões Degeneradas.....	30
6.5.	Mini Dimensões.....	31
6.6.	Dimensão Lixo	31
6.7.	Views	33
7.	Tabelas de Factos	33
7.1.	Tipos de Factos.....	34
7.2.	Tipos de Tabelas de Factos.....	35
7.2.1.	Tabelas de Factos Transaccionais.....	36
7.2.2.	Tabelas de Factos de Sumarização Periódica.....	36
7.2.3.	Tabelas de Factos de Sumarização Acumulada.....	37
7.2.4.	Tabelas de Factos Sem Factos	37
8.	Ferramentas Utilizadas Para o Caso Prático.....	39
8.1.	BIDS	39
8.1.1.	SQL Server Integration Services	39
8.1.2.	SQL Server Analysis Services	39
8.2.	Microsoft Office Excel	40
9.	Caso de Estudo	41
9.1.	Análise do Processo de Negócio.....	41
9.2.	Modelo de Dados	44
9.3.	Fonte de Dados	46
9.4.	Processo ETL.....	48
9.5.	Cálculo de Medidas	53
9.6.	Requisitos de Segurança	55
9.7.	Análise de Resultados	56
10.	Conclusões.....	58
10.1.	Trabalho Futuro.....	60
	Anexos.....	61
	Referências.....	75

Lista de Figuras

Figura 1 – Componentes do Data Warehouse (adaptado de Sá, 2009)	13
Figura 2 – Exemplo de Estrela em Estrela (Sá, 2009)	19
Figura 3 – Exemplo de Esquema em Floco de Neve (Sá, 2009).....	20
Figura 4 – Processo ETL	23
Figura 5 – Arquitectura do processo ETL do caso prático	52
Figura 6 – Cálculo de Medidas em MDX (Cálculo da Margem Bruta)	54
Figura 7 – Organograma do Grupo Empresarial	55

Lista de Tabelas

Tabela 1 – OLTP vs OLAP (adaptado de Sá, 2009; Cledes, 2001; Come, 2001).....	8
Tabela 2 – Exemplo de Matriz em Bus	22
Tabela 3 – Utilização das Surrogate Key	27
Tabela 4 – Exemplo de Alteração Tipo 1 (Antes)	28
Tabela 5 – Exemplo de Alteração Tipo 1 (Depois).....	28
Tabela 6 – Exemplo de Alteração Tipo 2	29
Tabela 7 – Exemplo de Alteração Tipo 3	30
Tabela 8 – Comparação de Tipos de Tabelas de Facto (Kimball e Ross, 2002)	35
Tabela 9 – Exemplo de Tabela de Factos Sem Factos	38
Tabela 10 – Matriz bus do caso prático	43
Tabela 11 – Granularidade dos processos de negócio	43
Tabela 12 – Exemplo do mesmo cliente em sistemas transaccionais de empresas distintas	47

Resumo

Pretende-se desenvolver um Data Warehouse para um grupo empresarial constituído por quatro empresas, tendo como objectivo primordial a consolidação de informação. A consolidação da informação é de extrema utilidade, uma vez que as empresas podem ter dados comuns, tais como, produtos ou clientes.

O principal objectivo dos sistemas analíticos é permitir analisar os dados dos sistemas transaccionais da organização, fazendo com que os utilizadores que nada percebem destes sistemas consigam ter apoio nas tomadas de decisão de uma forma simples e eficaz.

A utilização do Data Warehouse é útil no apoio a decisões, uma vez que torna os utilizadores autónomos na realização de análises. Os utilizadores deixam de estar dependentes de especialistas em informática para efectuar as suas consultas e passam a ser eles próprios a realizá-las. Por conseguinte, o tempo de execução de uma consulta através do Data Warehouse é de poucos segundos, ao contrário das consultas criadas anteriormente pelos especialistas que por vezes demoravam horas a ser executadas.

Development of a Data Warehousing System for the Information Consolidation in a Business Group

Abstract

It is intended to develop a Data Warehouse for a business related group of four companies, having by main goal the information consolidation. This information consolidation is of extreme usefulness since the companies can have common data, such as products or customers.

The main goal of the analytical systems is to allow analyze data from the organization transactional systems, making that the users that do not understand anything of these systems may have support in a simple and effective way in every process of taking decisions.

Using the Data Warehouse is useful to support decisions, once it will allow users to become autonomous in carrying out analysis. Users will no longer depend on computer experts to make their own queries and they can do it themselves. Therefore, the time of a query through the Data Warehouse takes only a few seconds, unlike the earlier queries created previously by experts that sometimes took hours to run.

1. Introdução

1.1.Motivação e Objectivos

Hoje em dia o mundo empresarial vive uma constante evolução nas tecnologias de informação e a competitividade aumenta a cada dia. É então imprescindível uma empresa manter-se competitiva, para tal, tem que ter uma boa capacidade de análise do seu negócio e das suas capacidades, bem como de oportunidades ou problemas que a afectam. Deste modo, os gestores necessitam de recolher diversos dados derivados de múltiplos sistemas que dão suporte à organização. A qualidade destes dados e a forma como são apresentados perante quem toma decisões pode ser um factor fulcral para se obter vantagem na competitividade empresarial. Os Data Warehouse nasceram fruto desta necessidade, proporcionando aos decisores uma forma de disponibilização da informação simples e eficiente.

O objectivo deste trabalho é desenvolver um Data Warehouse para um grupo empresarial que irá servir de apoio nas tomadas de decisões. Este grupo é constituído por quatro empresas que por uma questão de confidencialidade e segurança os nomes não serão divulgados, bem como qualquer outro dado respeitante a nomes de utilizadores, clientes, fornecedores, vendedores ou produtos. Assim, todo e qualquer dado presente neste documento serão fictícios. O sistema analítico que será desenvolvido pretende disponibilizar informação comercial da organização. Assim, o Data Warehouse terá que conseguir responder a questões tais como: quais os produtos mais vendidos, identificando por exemplo, o País para onde foi efectuada essa venda; qual o vendedor que mais contribui para a organização no primeiro trimestre do ano; qual o cliente com mais dividas e/ou pagamentos atrasados; quais os produtos

mais comprados pela empresa e quem é o utilizador que efectua essas compras. Estes são alguns exemplos de questões que surgem a quem analisa estes factos e, que procurará certamente respostas neste sistema que o apoiem ou possam apoiar a tomar uma decisão que torne a cada dia que passa a organização mais competitiva.

Cada empresa tem os seus próprios dados afectos à organização, que por sua vez podem ser comuns às restantes empresas. É então crucial que dados comuns entre empresas sejam consolidados, isto é, dados diferentes ou com determinado grau de semelhança que definam o mesmo objecto sejam representados apenas uma única vez. Por exemplo, um artigo com nome X numa empresa e com nome Y em outra mas que sejam o mesmo artigo devem aparecer, de alguma forma, uma única vez no Data Warehouse. Esta consolidação dos dados entre empresas terá como principal vantagem a análise global do grupo empresarial, onde será necessário analisar apenas um nome ao contrário de dois, tomando o exemplo anterior do artigo X e Y. Caso contrário, sem a existência da consolidação de dados, em qualquer análise que fosse feita existiriam pelo menos dois artigos que aparentemente seriam diferentes mas que na realidade é apenas um. Isto levaria a análises erróneas por parte dos utilizadores. A existência de mais do que um nome para definir o mesmo objecto levaria à dificuldade de chegar a conclusões e tomadas de decisões, assim como para a obtenção de determinados valores. Ter numa análise o artigo X e Y não é a mesma coisa que analisar o artigo X e Y consolidado num único artigo.

A informação consolidada no Data Warehouse é então, uma mais-valia para o grupo onde cada empresa descreve as entidades intrínsecas ao seu negócio de uma forma particular. A consolidação da informação permitirá

análises mais eficazes e eficientes para um aumento contínuo da competitividade do grupo.

1.2.Estrutura do Documento

Além do presente capítulo, o documento contém mais 9 capítulos, estando organizados da seguinte forma:

Capítulo 2 – Data Warehouse Empresarial vs Data Warehouse baseado em Data Marts

É possível ter duas estruturas de desenvolvimento de um Data Warehouse, este capítulo confronta as duas hipóteses.

Capítulo 3 – Online Transactional Processing (OLTP) vs Online Analytical Processing (OLAP)

Comparação entre os sistemas transaccionais e sistemas analíticos, dando a conhecer cada um deles é objectivo deste capítulo.

Capítulo 4 – Data Warehouse

Este capítulo tem como objectivo explicar o que é um Data Warehouse e quais os seus objectivos, explicando o seu processo de desenvolvimento.

Capítulo 5 – Extração, Transformação e Carregamento

Neste capítulo é descrito em que consiste o processo ETL e como ele é importante no desenvolvimento do Data Warehouse.

Capítulo 6 – Tabelas de Dimensão

Neste capítulo é feita uma análise pormenorizada a estas tabelas, analisando vários tipos de tabelas de dimensões que podem ser encontradas

num Data Warehouse e explicando os tipos de alteração nos valores dos seus atributos.

Capítulo 7 – Tabelas de Factos

Os diferentes tipos de factos existentes, bem como as distintas tabelas de factos são alvo de análise neste capítulo.

Capítulo 8 – Ferramentas Utilizadas no Caso Prático

Será feita uma breve análise das ferramentas utilizadas para o desenvolvimento do caso prático.

Capítulo 9 – Caso Prático

Após a análise e explicação teórica passa-se para a prática. Este capítulo dedica-se ao desenvolvido um Data Warehouse onde se explica cada passo e decisão tomada.

Capítulo 10 - Conclusões

Neste último capítulo é feita a análise crítica ao desenvolvimento do caso prático do capítulo anterior.

2. Data Warehouse Empresarial vs Data Warehouse baseado em Data Marts

Para desenvolver um sistema analítico existem duas arquitecturas possíveis, o Data Warehouse empresarial ou normalizado e o Data Warehouse baseado em Data Marts.

O primeiro tem como objectivo o desenvolvimento do sistema como sendo uma aplicação global, ou seja, olhando para a organização como um todo. Este é o paradigma da normalização (Inmom, 2002), cujo modelo de dados utilizado é o mesmo que se utiliza para o desenvolvimento de bases de dados relacionais, podendo até ser confundido como uma destas. Este modelo obriga a desenvolver uma arquitectura complexa e de difícil compreensão para os utilizadores do Data Warehouse. Segundo Caldeira (2008), esta metodologia tem como principal vantagem a agregação das medidas ou factos, mas em contrapartida, a sua arquitectura é bastante complexa o que torna bastante complexa a sua compreensão por parte de utilizadores não especialistas nesta matéria.

O Data Warehouse com base em Data Marts é desenvolvido gradualmente, faseando os custos de investimento. Esta metodologia de desenvolvimento permite entregar e/ou disponibilizar o Data Warehouse mais depressa aos utilizador, embora que não seja no seu todo. Esta arquitectura assenta no modelo dimensional, em que o esquema em estrela tem um papel preponderante no Data Warehouse. A arquitectura em estrela consiste numa tabela central que contém as medidas ou factos, rodeada por outras tabelas denominadas dimensões que descrevem os factos. Deste modo, é necessário ter mais do que um Data Mart para se obter um Data Warehouse (Kimball e Ross, 2002).

As duas arquitecturas descritas anteriormente não podem ser combinadas entre si uma vez que são antagónicas. O Data Warehouse empresarial ou normalizado é muito complexo o que dificulta de certa forma a interacção com os utilizadores e a sua percepção do Data Warehouse, prejudicando ainda o desempenho do Data Warehouse devido ao elevado número de ligações (do inglês *joins*) entre tabelas. Estas ligações são essenciais para ligar diferentes tabelas nas pesquisas de informação. Por outro lado, a segunda arquitectura de Data Warehouse tem um desenho que é facilmente legível pelos utilizadores e tem melhores níveis de desempenho uma vez que o número de ligações entre tabelas é muito mais reduzido (Caldeira, 2008).

Não é apenas no desempenho que a arquitectura baseada em Data Marts é mais vantajosa, analisando os custos de desenvolvimento das duas arquitecturas, esta tem um menor custo segundo Caldeira (2008), onde afirma que as tarefas de desenvolvimento podem ser distribuídas por diversos especialistas e que a sua posterior integração é facilmente realizada.

Os Data Marts podem ser classificados em dois tipos, dependentes e independentes. Esta distinção faz-se consoante origem dos dados. Deste modo, o Data Mart é considerado independente quando os dados são extraídos dos sistemas transaccionais e por sua vez, são considerados dependentes quando os dados são extraídos a partir do Data Warehouse. Caldeira (2008) defende que esta classificação “não faz muito sentido” uma vez que construir um Data Mart através do Data Warehouse é trabalho duplicado aumentando os custos de desenvolvimento.

3. Online Transactional Processing (OLTP) vs Online Analytical Processing (OLAP)

É importante ter conhecimento destes dois conceitos distintos que são o OLTP e o OLAP, onde o primeiro serve de suporte ao dia-a-dia da organização e tem como finalidade registar as transacções que ocorrem no seu funcionamento, tal como, as encomendas, facturas ou o processamento de vencimentos. Estas operações são acontecimentos estruturados, repetitivos e isolados (Sá, 2009). Já o OLAP fornece apoio na análise dos dados da organização, que tem como utilizadores finais os executivos, gestores e analistas que têm uma forte necessidade de consumir análises e ter suporte nas suas decisões. Segundo Sá (2009), os sistemas analíticos complementam os sistemas transaccionais.

Segundo Caldeira (2008), os sistemas OLTP são os sistemas transaccionais de uma organização. Estes sistemas estão desenvolvidos para processar determinados eventos isoladamente, designados por transacções. Deste modo, um sistema OLTP é qualquer aplicação que regista transacções, normalmente, em tempo real e onde estas transacções reflectem a forma como organização opera as suas actividades. A execução destas transacções movimenta uma pequena quantidade de dados e o tempo de execução é bastante reduzido. Estes sistemas têm como principal objectivo armazenar as ocorrências na organização em tempo real, no menor período de tempo possível e suportar centenas de utilizadores e operações em concorrência (Sá, 2009).

Os sistemas OLAP ao contrário dos sistemas OLTP têm como principal objectivo a análise dos dados dos sistemas OLTP. Estes sistemas preocupam-se com grandes volumes de dados e não com registos isolados (Caldeira, 2009).

Sá (2009) refere que a finalidade destes sistemas é o armazenamento histórico das actividades da organização e a integração de informação oriunda de distintas fontes de dados internas ou externas à organização. Deste modo, o Data Warehouse permite fazer análises do negócio da organização, executar consultas complexas que podem englobar um volume significativo de informação, carregar informação sem necessidade de recorrer a mecanismos de validação (os dados são validados antes do seu carregamento no sistema analítico) e é desenvolvido para suportar poucos utilizadores em concorrência.

As diferenças entre os sistemas são assim resumidas na Tabela 1.

Tabela 1 – OLTP vs OLAP (adaptado de Sá, 2009; Cledes, 2001; Come, 2001)

		Base de Dados Operacional	Data Warehouse
Características	Destinatários	Aplicações informáticas operacionais Colaboradores	Analistas de negócio Gestores Executivos
	Função	Registo de operações diárias	Suporte à decisão
	Informação	Correntes Actualizados Atómicos Relacionais (normalizada) Isolados	Históricos Sumariados Multidimensionais Integrados
	Utilização	Repetitivos, rotineiros	<i>Ad-hoc</i>
	Acessos	Leitura e Escrita Transacções simples	Maioritariamente leitura Questões complexas
	Requisitos elementares	Débito de transacções Consistência dos registos informáticos	Débito de questões Precisão da informação

Em suma, os dados dos sistemas OLAP e OLTP serão os mesmos, mas armazenados e organizados de forma a maximizar as suas características. Os sistemas OLTP estão vocacionados para a criação e/ou edição de dados, enquanto os sistemas OLAP servem para confrontar grupos de dados relacionando-os por diversos critérios.

4. Data Warehouse

4.1.O que é?

Para que uma organização tenha sucesso ela tem que ser competitiva de modo a ultrapassar a concorrência. Potencializar as capacidades da organização, identificar pontos menos fortes e identificar novos nichos de mercado são factores fulcrais para o desenvolvimento e estabilidade da organização ou da empresa. Assim, os Data Warehouse são uma mais-valia oferecendo apoio nas decisões ajudando a escolher a melhor opção.

Os sistemas de Data Warehouse são aplicações desenvolvidas de origem e à medida da organização, com o único propósito de dar suporte na análise de dados. Qualquer utilizador destas aplicações, normalmente são gestores ou decisores, consegue por si só fazer consultas e análises sem necessitar de ter conhecimento SQL ou ser especialista em base de dados. A usabilidade destes sistemas é o factor primordial no seu desenvolvimento, oferecendo a possibilidade de o utilizador, de forma autónoma, seguir o seu próprio caminho na análise e exploração dos dados.

Kimball e Caserta (2004) definem o Data Warehouse como um processo de obtenção de dados de um sistema transaccional, transformá-la em informação organizada para apoiar na tomada de decisões. Assim, um Data Warehouse é um sistema que extrai, limpa e carrega os dados num modelo de armazenamento dimensional, e por fim, permite análises que apoiam nas decisões. O Data Warehouse é constituído por diversos componentes onde cada um é projecto com as suas técnicas e ferramentas e nenhum destes componentes constitui por si só o Data Warehouse. Por exemplo, o sistema ETL é um componente muito importante, mas existem outros igualmente

importantes que não devem ser descartados. O conjunto dos diversos componentes constitui o Data Warehouse. Os autores realçam ainda a importância para o facto de o Data Warehouse ser muitas vezes tratado como um produto. Contrariamente a muitas explicações de vendedores ou fornecedores o Data Warehouse não é um produto que possa ser adquirido, este inclui tarefas fundamentais que não podem ser descartadas. É imprescindível fazer uma análise dos sistemas e negócio da organização, transformar e limpar dados e por fim, desenhar o modelo de dados de acordo com as necessidades da organização. Estas tarefas fazem com que o Data Warehouse seja à medida e específico, cada caso é um caso. É um erro comum pensar-se no Data Warehouse como um projecto. O seu desenvolvimento é constituído por diversos projectos, geralmente, cada Data Mart é um projecto com a sua própria linha de tempo de desenvolvimento e orçamento. Cada Data Mart é composto por dimensões conforme e factos padronizados para que cada um se integre e forme uma única peça, o Data Warehouse. Assim, o Data Warehouse deve ser visto como um processo e não como um projecto.

Segundo Inmon (2002) o Data Warehouse é um conjunto de dados orientado a um assunto, integrado, não volátil e variável no tempo que é desenvolvido com o objectivo de apoiar as decisões. É orientado a um assunto uma vez que são desenhados para a organização analisar os seus dados onde estes são categorizados por temas. Os Data Warehouse armazenam dados provenientes de diversas fontes de informação onde estes são transformados antes de serem guardados, assim diz-se integrado. Ser não volátil significa que os dados armazenados no Data Warehouse não são alterados pelo utilizador, apenas são consultados. As modificações que ocorrem nos sistemas transaccionais não implicam alterações nos dados carregados, podem ser

novos carregamentos de dados. Por fim, o autor refere que são variáveis no tempo, isto deve-se ao facto de nos sistemas transaccionais os utilizadores, normalmente, apenas têm a informação actual dos acontecimentos, enquanto nos Data Warehouse é possível analisar a informação num determinado momento. Isto porque é armazenando todo o histórico dos acontecimentos. Estas são as características que, segundo Inmon (2002), diferem os sistemas Data Warehouse dos sistemas operacionais.

Em suma, o Data Warehouse é um sistema que permite aos utilizadores construir consultas, relatórios e diversos tipos e formas de análises de uma forma transparente e bastante simplificada.

4.2.Objectivos

O Data Warehouse sendo um sistema analítico, o seu objectivo primordial é o processamento de grandes volumes de dados e posterior publicação para análises de dados que sirvam de suporte na tomada de decisões.

Hoje em dia o volume de dados gerado por uma organização é imenso e este acumular de dados nos sistemas transaccionais torna-os de difícil acesso e muitas vezes inúteis para dar algum apoio nas decisões. Para colmatar esta falha nos sistemas OLTP, o Data Warehouse é desenvolvido com o objectivo de os dados se tornarem em informação e facilmente serem analisados dando suporte na tomada de decisões. Assim, o Data Warehouse pode conter dados provenientes de diversas aplicações transaccionais e incompatíveis entre si, que de outra maneira seria muito difícil de serem analisados conjuntamente.

Kimball e Ross (2002) definem seis grandes objectivos para que um Data Warehouse tenha sucesso:

1. Deve tornar a informação da organização de fácil acesso, o conteúdo deve ser compreensível e intuitivo para o utilizador e o tempo de espera deve ser mínimo.
2. Deve apresentar a informação da organização de forma consistente, ou seja, os dados têm que ser credíveis.
3. Deve ser adaptável e resistente à mudança. Quando são feitos novos desenvolvimentos em cima do Data Warehouse, este deve suportar essas alterações garantindo o seu correcto funcionamento.
4. Deve ser seguro de modo a proteger a informação, isto é, o acesso a informação confidencial deve ser controlado.
5. Deve servir para ajudar a escolher a melhor opção na tomada de decisões, uma vez que é um sistema de apoio à decisão.
6. Por fim, a comunidade empresarial deve aceitar o Data Warehouse. O seu uso deve ser simples para que os decisores o utilizem como uma mais-valia para a organização.

4.3.Componentes

O Data Warehouse, tal como acontece em outros sistemas, é composto por subsistemas. Assim, consideram-se quatro componentes: os sistemas operacionais ou transaccionais que funcionam como fonte de dados; a área de estágio que se encarrega de extrair, transformar e limpar os dados dos sistemas operacionais; a apresentação dos dados onde os dados são carregados através da área de estágio; e, o acesso aos dados onde os

utilizadores através de uma interface utilizam as funcionalidades do Data Warehouse.

Estes componentes estão ligados e são indispensáveis para o sucesso do Data Warehouse. A Figura 1 mostra a forma como cada componente se relaciona com os demais.

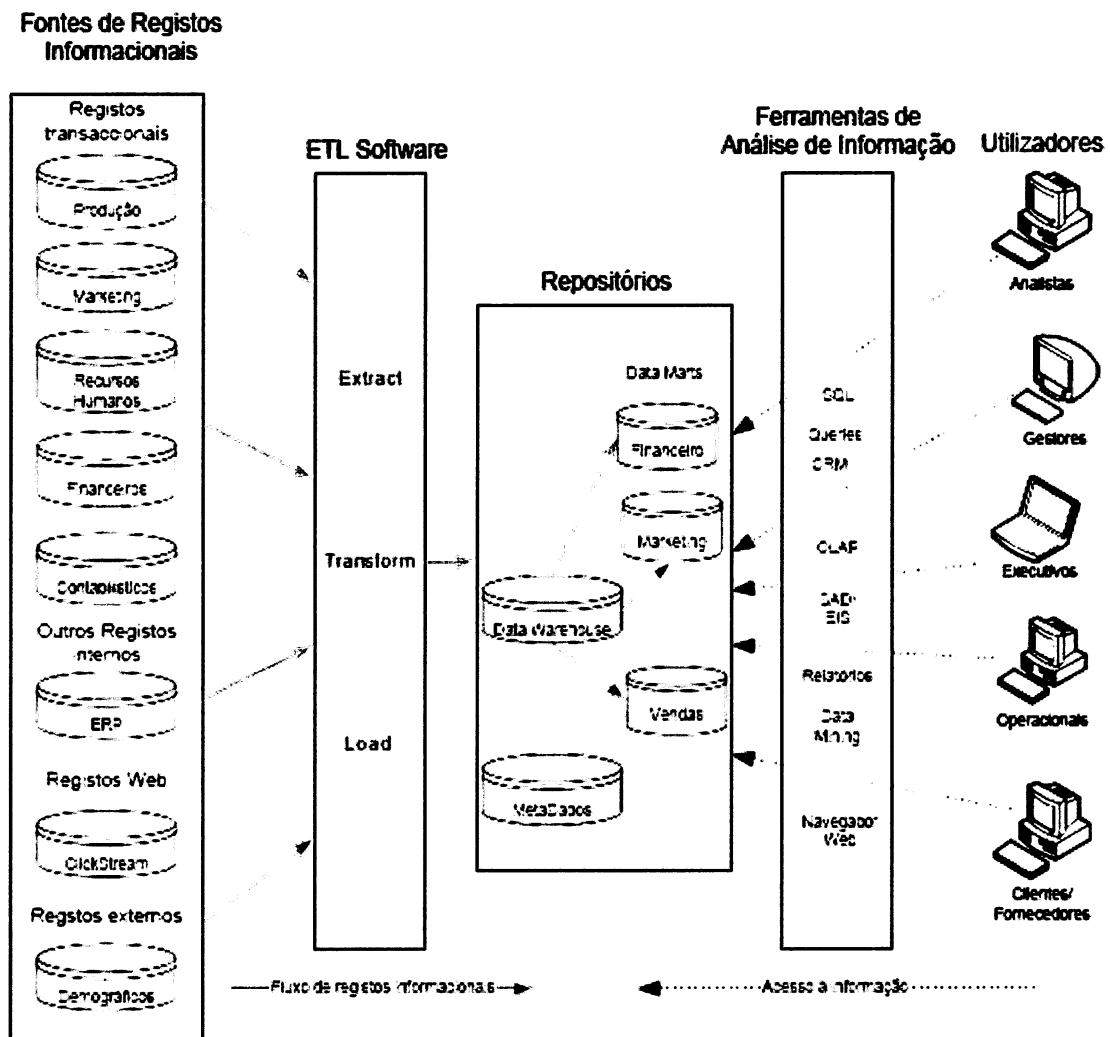


Figura 1 – Componentes do Data Warehouse (adaptado de Sá, 2009)

Por um lado temos a fonte dados onde estão armazenados os registos transaccionais, através do processo de ETL estes registos são transformados e colocados no Data Warehouse e respectivos Data Marts. Por fim, através de ferramentas de acesso e análise de informação os utilizadores têm acesso à informação armazenada no Data Warehouse.

4.3.1. Sistemas Operacionais

Os sistemas operacionais são as fontes de dados que irão alimentar o Data Warehouse. Estas fontes de dados podem ser classificadas em internas ou externas. As fontes de dados internas são as mais frequentes e são aquelas que a organização utiliza no seu dia-a-dia para registar e armazenar as suas transacções. Enquanto as externas, são fontes de informação que não existem na organização, portanto são externas à empresa, mas que fazem todo sentido no Data Warehouse porque são uma mais-valia fornecendo um maior detalhe sobre os dados organizacionais (Caldeira, 2008; Kimball e Ross, 2002). São exemplos de dados externos os códigos postais, dados demográficos, económicos, estudos de mercado ou câmbios que podem ser adquiridos através de empresas que prestam este tipo de serviços ou de organização que disponibilizam este conjunto de informação de forma gratuita. Deste modo, o Data Warehouse engloba múltiplas fontes de dados passando desde simples ficheiros de texto, a folhas de cálculo e/ou diversas bases de dados.

4.3.2. Área de Estágio

Este componente é o elo de ligação entre múltiplas fontes de dados e o Data Warehouse. Este componente é constituído por um armazenamento temporário de dados e o sistema ETL (Extract, Transform and Load). Neste componente faz-se a extracção da informação das diversas fontes de dados, depois fazem-se as transformações necessárias sobre os dados extraídos e por fim carregam-se esses dados no Data Warehouse. Por algum motivo pode-se armazenar temporariamente estes dados antes do carregamento no Data Warehouse (Caldeira, 2008; Kimball e Ross, 2002). É através do processo ETL que se faz a união das diversas fontes de dados, este processo é responsável pela obtenção dos dados nos diferentes formatos de origem, transformando e

limpando esses dados, colocando-os em conformidade para que possam ser carregados no Data Warehouse. Todo este complexo processo será explicado mais detalhadamente no capítulo 5.

4.3.3. Apresentação dos Dados

Este subsistema é onde os dados estão organizados, armazenados e disponibilizados aos utilizadores, para que estes possam efectuar consultas através de ferramentas de acesso aos dados.

Cada processo de negócio da organização dá origem a um Data Mart e o conjunto dos Data Mart constituem o Data Warehouse. Cada Data Mart representa apenas um processo de negócio, cujos seus dados devem ser guardados com o menor grão possível, isto é, com o máximo de detalhe. Estes diferentes níveis de detalhe presente nos Data Mart denomina-se de atomicidade dos dados e o dado mais atómico representa maior detalhe. No caso de o Data Mart armazenar as vendas da organização e o dado mais atómico for a venda diária, através de agregações poder-se-ão realizar análises mensais, o que já não acontece no inverso, se o dado mais atómico for a venda mensal, não é possível analisar as vendas num determinado dia (Caldeira, 2008; Kimball e Ross, 2002). Deste modo, o detalhe dos dados é fulcral para que o Data Warehouse consiga dar resposta a todas as questões colocadas pelos utilizadores.

4.3.4. Acesso aos Dados

O último componente do Data Warehouse é o acesso aos dados. As ferramentas de acessos aos dados dão aos utilizadores a possibilidade e capacidade de consultar e analisar a informação armazenada e disponibilizada no Data Warehouse. Estas ferramentas podem ser aplicações tão simples

como uma folha de cálculo, que com um simples arrastar de campos se elaboram as consultas pretendidas (Caldeira, 2008; Kimball e Ross, 2002).

4.4.Metadados

A definição mais comum diz que são dados sobre dados, embora seja uma definição muito vaga, realmente, os metadados são um conjunto de dados sobre os dados (Clemes, 2001). Os metadados são informação sobre os dados e assumem um papel fundamental no Data Warehouse, estes descrevem os dados, o ambiente que os rodeia, a forma como são tratados e sem eles os dados não têm significado (Come, 2001). Como já foi dito anteriormente, o Data Warehouse pode conter dados derivados de diversas fontes de dados, representados e tratados de diferentes formas, onde a obtenção de um sistema homogéneo pode ser uma tarefa bastante complexa colocando bastantes entraves (Caldeira, 2008).

Os metadados são muito importantes na medida em que estes dão origem aos primeiros diálogos com os utilizadores do sistema. Segundo Come (2001) um Data Warehouse sem metadados é como arquivo de documentos sem estes estarem classificados em pastas. Perante este cenário os metadados devem existir desde o início do Data Warehouse mas, segundo Kimball e Ross (2002), isto muitas vezes não acontece porque quem desenvolve tem uma certa aversão à documentação e acaba por ignorar os metadados apesar de ser do conhecimento de todos que estes são essenciais.

Clemes (2001) afirma existirem duas classificações possíveis para os metadados do Data Warehouse. Os metadados técnicos descrevem as regras de transformação dos dados, nomes de atributos ou tipos de dados, onde esta informação é fulcral para a extracção e transformação dos dados bem como

para a manutenção do Data Warehouse. Os metadados de negócio são mais voltados para os utilizadores finais, onde estes tendem a encapsular toda a complexidade do Data Warehouse permitindo aos utilizadores finais perceber de forma clara e evidente as regras de negócio. Informação sobre campos e mapeamento entre tabelas são exemplos desta classificação de metadados.

Existem duas fontes para os metadados, as formais e as informais, e ambas abrangem os metadados técnicos e de negócio. As fontes de metadados formais são aquelas que já foram discutidas e documentadas (Come, 2001). Documentam a origem dos dados e fazem parte da documentação formal da empresa (Clemes, 2001). Deste modo, os metadados formais são aqueles que foram discutidos e documentados na empresa. Normalmente são armazenados em documentos onde são mantidos e distribuídos por toda a organização. Os metadados informais são os que existem na organização mas não estão registados na documentação formal, é a informação que faz parte da organização e que os colaboradores têm conhecimento mas não está documentada (Clemes, 2001; Come, 2001).

De uma forma breve, os metadados são importantes para o desenvolvimento do Data Warehouse e obrigatórios desde o seu começo.

4.5. Concepção do Data Warehouse

O desenvolvimento de um Data Warehouse é muito mais que um simples projecto. É algo que necessita de ser bem analisado e pensado, seguindo determinadas tarefas que não podem de maneira nenhuma ser deixadas para trás ou simplesmente ignoradas. Todo e qualquer passo no desenvolvimento terá que ser previamente analisado e fundamentado, sob pena de o resultado final não ser o esperado e os utilizadores simplesmente

não usarem um sistema que quando bem construído, é uma mais-valia para a organização.

Para se construir um Data Warehouse antes de mais, é imprescindível conhecer a fonte ou fontes de dados que são utilizados no sistema transaccional da organização e que irão alimentar o sistema analítico, sendo igualmente necessário saber com clareza de onde vêm os dados e de que tipos são. A identificação dos processos de negócio que irão fazer parte do Data Warehouse é uma tarefa bastante importante e que deve ter especial atenção, é com base nesta análise que todo o desenvolvimento vai incidir. Depois de definidos quais os processos de negócio, chega então a fase de desenho dos modelos dimensionais, definindo os factos e as dimensões que irão descrever estes factos. Por último, deve-se manter sempre a matriz em bus actualizada, para que de uma forma simples e rápida os utilizadores percebam quais as dimensões que descrevem cada processo de negócio e se consiga identificar que factos partilham os mesmos descritores.

4.5.1. Modelos Dimensionais

O modelo dimensional é uma forma de relacionar os dados. Este relacionamento forma um cubo que pode ser subdividido pelas dimensões, resultando dados com mais ou menos detalhe (Santos, 2010).

4.5.1.1. Esquema em Estrela

A característica principal do esquema em estrela é a sua elevada redundância (Santos, 2010). Esta redundância melhora o desempenho uma vez que diminui bastante o número de ligações entre tabelas.

Um esquema em estrela é constituído por uma tabela de factos rodeada por várias tabelas de dimensões. Os atributos destas dimensões descrevem as medidas presentes nas tabelas de factos.

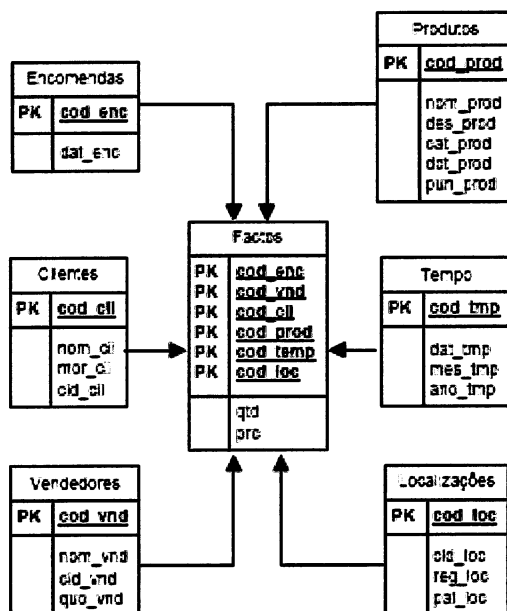


Figura 2 – Exemplo de Estrela em Estrela (Sá, 2009)

A Figura 2 é um exemplo de um esquema em estrela, onde a Tabela de Factos contém as medidas que serão alvo de análises, rodeada pelas diversas dimensões que caracterizam estes factos.

O desenhar desta estrutura implica seguir quatro passos fundamentais, segundo Caldeira (2008). Em primeiro lugar é necessário escolher os processos de negócio afectos à organização, estes são actividades ou fluxos de informação que atravessam uma organização horizontalmente bem como vários departamentos, produzindo um conjunto de medidas. Em seguida é necessário definir o grão isto é, definir o nível de detalhe; quanto menor for a granularidade mais será o nível de detalhe da informação contida na tabela. Após esta definição, a próxima tarefa é determinar que dimensões são importantes para o Data Warehouse, sendo que para este passo, a maneira mais simples de decidir que dimensões fazem parte do esquema em estrela é

recorrer à matriz em bus. Por fim devem-se identificar as medidas da tabela de factos, ou seja, identificar os valores numéricos ou não numéricos que serão preenchidos na tabela de factos, devendo os diversos valores ficar guardados com o mesmo nível de detalhe.

4.5.1.2. Esquema em Floco de Neve

O esquema em floco de neve é uma variante do esquema em estrela e, segundo (Santos, 2010) consiste na normalização das dimensões. Tal como o esquema anterior, as tabelas de dimensão ligam-se à tabela de factos que se encontra no centro do esquema. A diferença entre os dois esquemas encontra-se nas dimensões. No esquema em floco de neve as dimensões são divididas em várias tabelas normalizadas, ficando o esquema o Data Warehouse semelhante a um esquema relacional de uma base de dados transaccional, tal como mostra a Figura 3.

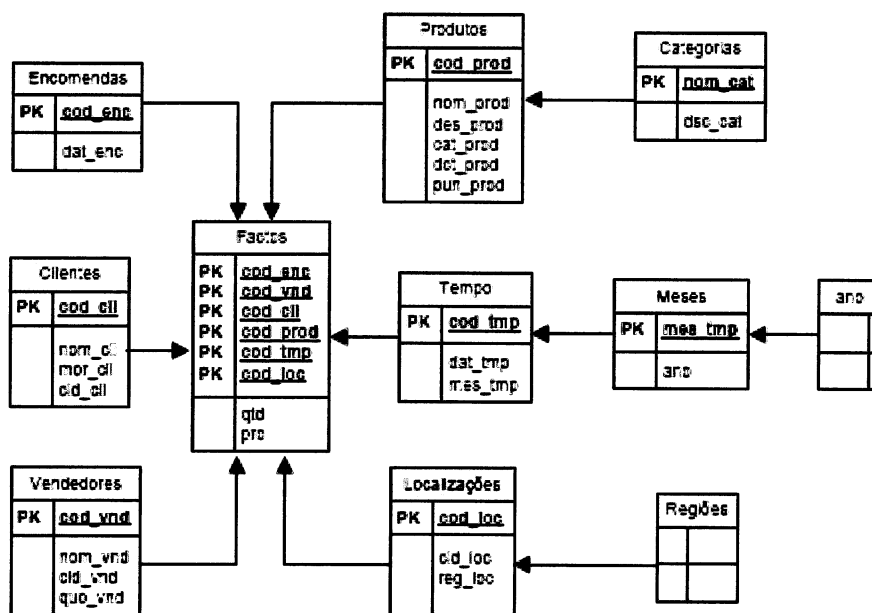


Figura 3 – Exemplo de Esquema em Floco de Neve (Sá, 2009)

A normalização das tabelas de dimensão tem um efeito negativo no desempenho do Data Warehouse. O número de ligações aumenta substancialmente, aumentando também o tempo de obtenção dos resultados.

Por este motivo, não é aconselhável a sua utilização em sistemas de Data Warehouse (Santos, 2010).

4.5.2. Processo de Negócio

A abordagem para analisar os requisitos para um sistema analítico não é a mesma que se utiliza para os sistemas transaccionais. O desenvolvimento dos sistemas operacionais tem como objectivo primordial a optimização do armazenamento e a redução da redundância dos dados. Por sua vez, os Data Warehouse são optimizados para o utilizador e para este trabalhar os dados de uma forma autónoma. Os sistemas analíticos não devem ter como base qualquer especificação dos sistemas operacionais, mesmo sendo estes a fonte de dados. O estudo das aplicações operacionais e a consolidação em temas de análise é a primeira tarefa da definição dos processos de negócio. Cada um destes temas é analisado, sendo convertidos posteriormente em processos de negócio. Um processo de negócio é qualquer actividade previamente definida que ocorre na organização com o intuito de atingir os objectivos estabelecidos. Cada processo tem o seu método de exploração e a informação gerada pode interessar a vários departamentos da organização, pode-se então dizer que um processo atravessa horizontalmente a organização (Caldeira, 2008).

4.5.3. Matriz em Bus

A matriz em bus é uma ferramenta que tem como objectivo principal servir de suporte para o desenvolvimento e que permite criar e documentar a arquitectura do Data Warehouse, servindo posteriormente, para comunicar aos utilizadores a estrutura do Data Warehouse (Caldeira, 2008; Kimball e Ross 2002). Permite ainda identificar de uma forma simples e rápida os processos de negócio existentes na organização.

É a partir da construção desta matriz que se dá início ao desenvolvimento do Data Warehouse porque é de fácil compreensão, dando origem às primeiras reuniões com os futuros utilizadores (Sousa, 2010). A matriz é uma boa ferramenta de comunicação com os utilizadores que não entendem da parte técnica, assim conseguem ter a percepção do funcionamento do Data Warehouse e ajudar na sua definição, uma vez que é simples e directa (Kimball, 1999).

Esta matriz é um quadro cujas linhas representam os processos de negócio afectos à organização e as colunas representam as dimensões que descrevem cada processo. As células da matriz são marcadas para identificar quais dimensões estão associadas aos processos (Ross, 2002). Cada linha da matriz representa um Data Mart e o desenvolvimento do Data Warehouse deve-se iniciar pelos Data Marts de maior dimensionalidade, isto é, aqueles que contém mais dimensões que o caracterizam. Segundo Ross (2005), estas colunas servem para identificar “quem, o quê, onde, quando, porquê e como” para cada actividade transaccional.

A Tabela 2 é um exemplo do que pode ser uma matriz em bus.

Tabela 2 – Exemplo de Matriz em Bus

	Dim1	Dim2	Dim3	Dim4	Dim5	Dim6	Dimensionalidade
Processo de Negócio 1	x	x	x	x		x	5
Processo de Negócio 2	x	x			x	x	4

Neste caso, o desenvolvimento do Data Warehouse deveria iniciar-se pelo Processo de Negócio 1, uma vez que tem dimensionalidade 5 sendo esta superior à do Processo de Negócio 2 cuja dimensionalidade é 4.

A grande vantagem da utilização desta matriz é a sua elasticidade e não deve ser vista como uma representação estática da arquitectura do Data

Warehouse. A matriz deve reproduzir todas as alterações realizadas no Data Warehouse e deve estar sempre actualizada, mantendo um registo histórico detalhado para posteriores consultas (Sousa, 2010).

5. Extracção, Transformação e Carregamento

Após a identificação dos processos de negócio da organização, construção da matriz em bus e dos modelos dimensionais, chega então a fase de passar os dados do sistema transaccional para o sistema analítico. Esta passagem denomina-se de processo ETL (Extract, Transform and Load).

Este processo implica três fases distintas: extracção de dados de uma ou diversas fontes provenientes dos sistemas transaccionais, transformação e limpeza destes mesmos dados e posterior carregamento no Data Warehouse. Estas três fases devem ser vistas como um único processo e não como processos independentes, tal como mostra a Figura 4.

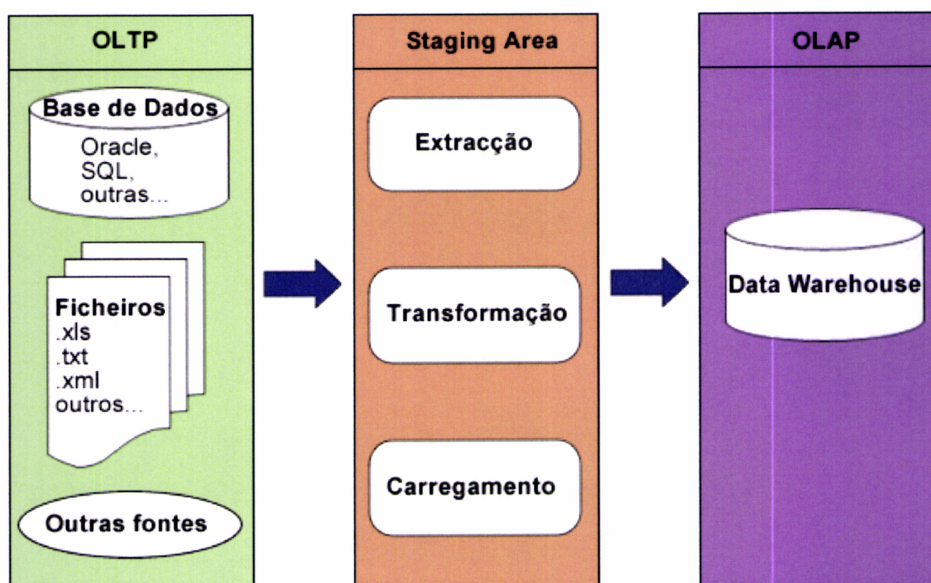


Figura 4 – Processo ETL

A fase de extracção é a etapa de aquisição dos dados provenientes dos múltiplos sistemas transaccionais. Durante esta etapa podem ocorrer diversos erros sobre os dados, assim, a fase seguinte consiste na transformação e

limpeza dos dados extraídos, sendo esta etapa também bastante complexa. Deste modo, é importante limpar os dados tornando-os consistentes, sem erros e não duplicados. Por último, a fase de carregamento dos dados no Data Warehouse.

A fase intermédia deste complexo processo é uma das mais importantes, transformar e limpar os dados tornando-os consistentes e sem erros é imprescindível. A qualidade dos dados é um factor preponderante para o Data Warehouse. Para Kimball e Caserta (2004) a qualidade dos dados garante-se através de quatro características fundamentais, os dados têm que ser correctos, ou seja, verdadeiros; não devem de modo algum ser ambíguos, os valores e as descrições devem apontar numa só direcção, isto é, precisos; devem ser consistentes, para tal, determinada descrição com um certo significado deve ser sempre igual; e, por último, devem ser completos, o sistema ETL deve garantir que não existem falhas ao longo do processo e que tudo é carregado no Data Warehouse.

Devido à existência das surrogate keys, as tabelas de dimensão são carregadas primeiro e, só depois deste carregamento, são preenchidas as tabelas de factos. Todo este processo é complexo e moroso fazendo com que chegue ocupar cerca de 80% do trabalho de desenvolvimento de um sistema analítico (Caldeira, 2008).

Após se ter os dados carregados no Data Warehouse é necessário processar os cubos, sendo que o tempo de execução deste processo depende directamente da quantidade de registos presentes. É nesta fase que são feitas todas as agregações e respectivos cálculos das medidas presentes nos cubos permitindo a posterior consulta em diferentes perspectivas.

Os sistemas de ETL devem estar preparados para superar situações inesperadas. Estes estão sujeitos, ao longo do tempo, a diversas situações. Ao se desenvolver um mecanismo de ETL deve-se ter em conta todos os factores possíveis e imaginários para garantir que este processo não falhe e se comporte como é esperado e de uma forma estável (Silvers, 2008).

6. Tabelas de Dimensão

As tabelas de dimensão descrevem as medidas registadas nas tabelas de facto, estas são conhecidas por terem muitos atributos e são consideradas conforme quando se incluem em mais do que um processo de negócio, possuem muitos atributos e cujos nomes estão escritos correctamente, que no nosso caso é em bom português. Para salvaguardar medidas que podem não conter determinada dimensão, quer por erro informático ou porque ainda não ocorreu, todas as dimensões devem conter pelo menos uma linha cujos seus atributos devem ser do tipo “Não disponível” ou “Ainda não aconteceu” para que, de forma amigável, o utilizador perceba o que está a acontecer (Caldeira, 2008; Kimball e Caserta, 2004). Da mesma forma, deve-se prevenir que os registos dimensionais tenham campos vazios, pode acontecer que determinado registo tenha um ou mais atributos a NULL (vazios). Assim, deve colocar também esse atributo com um valor padrão como por exemplo “Não Aplicável” (Ross, 2009).

Dentro de cada dimensão existe uma estrutura hierárquica constituída pelos atributos da própria dimensão. Esta estrutura é a forma como os atributos se organizam por nível de detalhe, estando no topo o atributo com menor detalhe e na base da hierarquia o atributo mais atómico. Esta estrutura permite efectuar análises com mais ou menos detalhe, conforme o nível da

hierarquia. Os dados são agrupados e as medidas sumarizadas. Um exemplo claro desta situação é a data em que podemos ter a hierarquia Ano, Mês e Dia; em que o dia teria os dados na sua forma atómica e, para mês e ano as medidas seriam sumarizadas quando os factos assim o permitam.

6.1.Surrogate Key

As chaves primárias destas tabelas denominam-se de surrogate key e são especiais na medida em que são um valor numérico gerado sequencialmente permitindo assim, a integração de vários sistemas transaccionais e são importantes para o Data Warehouse suportar as alterações tipo 2 das dimensões. A gestão destas chaves é bastante complexa e tem que ser muito minuciosa. As primeiras tabelas a serem carregadas são as dimensões, onde as chaves primárias vão sendo criadas e, só depois são carregadas as tabelas de factos que fazem referência às surrogate keys através das suas chaves estrangeiras. (Caldeira, 2008; Kimball e Ross, 2002).

A surrogate key é a chave substituta da chave natural. A chave natural é a chave primária de determinado registo no sistema transaccional. Deste modo, é possível integrar registos de diversas fontes de dados, por exemplo o seguinte caso, uma empresa que tem duas lojas, uma em Évora e a outra em Portalegre. Cada uma das lojas tem os seus clientes e ambas têm o cliente cujo código é C01. São clientes distintos, numa é o Joaquim e na outra é o Manuel, que não partilham nada nem se conhecem. Através das surrogates key é possível carregar estes dois clientes numa dimensão que seria denominada por Cliente e que seria semelhantes à Tabela 3.

Tabela 3 – Utilização das Surrogate Key

Chave do Cliente	Código do Cliente	Nome do Cliente	...
23	C1	Joaquim	...
...
64	C1	Manuel	...

Ainda poderíamos ter outra situação, onde numa das lojas o código do cliente seria um campo numérico e, na outra, alfanumérico. Sem a utilização destas chaves não seria possível juntar todos os clientes numa única dimensão.

6.2.Dimensões Planas e Dimensões em Floco de Neve

Existem duas classificações possíveis para as dimensões segundo os autores Kimball e Ross (2002), as dimensões planas (do inglês, *Flat Dimension*) e as dimensões em floco de neve (do inglês, *Snowflaked Dimension*). As dimensões planas são as mais utilizadas e aconselháveis para os Data Warehouse, sendo completamente desnormalizadas. Deste modo, todas as hierarquias presentes na origem de dados encontram-se todas juntas numa mesma tabela de dimensão do Data Warehouse, ou seja, a hierarquia mantém-se mas de uma forma desnormalizada. As dimensões em floco de neve, ou seja, normalizadas, resultam da existência de dimensões ligadas a outras dimensões através dos seus atributos. A sua utilização é pouco recomendável uma vez que são mais complexas de gerir e diminuem o grau de usabilidade que é um dos principais objectivos dos sistemas analíticos. Só se deve fazer uso deste tipo de dimensões em casos muito específicos, por exemplo, quando se pretende agrupar os dados de uma dimensão por classes e denomina-se por *outrigger*.

As dimensões planas são as que advêm da utilização do esquema em estrela, enquanto as dimensões em floco de neve são as que resultam da utilização de um modelo normalizado, o esquema em floco de neve.

6.3. Alterações nos valores dos atributos

Um ponto muito importante e imprescindível de ser analisado é a alteração aos dados contidos nas tabelas de dimensões. Assim, temos três tipos de alterações sendo que um deles não é muito utilizado.

6.3.1. Alterações Tipo I

As alterações do Tipo 1 são aquelas em que a linha que foi modificada é substituída pela nova linha e apenas deve ser utilizado excepcionalmente, uma vez que se perde o histórico referente a esse mesmo registo. O controlo deste tipo de alteração pode ser feito adicionando uma coluna à dimensão, em que essa contém um checksum de todos os campos. Sendo esse checksum diferente então a linha foi modificada, e terá que se proceder à sua alteração (Caldeira, 2008; Kimball e Caserta, 2004; Kimball e Ross, 2002).

O exemplo seguinte mostra a imagem de uma dimensão antes e depois de uma alteração deste tipo, Tabela 4 e Tabela 5 respectivamente.

Tabela 4 – Exemplo de Alteração Tipo 1 (Antes)

Chave do Colaborador	Número de Colaborador	Nome do Colaborador	Departamento
21	10	Carlos	Comercial

Tabela 5 – Exemplo de Alteração Tipo 1 (Depois)

Chave do Colaborador	Número de Colaborador	Nome do Colaborador	Departamento
21	10	Carlos	Financeiro

Como podemos verificar, os campos alterados são substituídos no próprio registo perdendo-se todo o histórico do colaborador ficando apenas a informação actual.

6.3.2. Alterações Tipo II

Estas alterações são as mais utilizadas, porém um pouco mais complexas uma vez que guardam o histórico de um registo. Ao existir uma alteração num determinado campo de uma tabela de dimensão, esta passará a ter uma nova linha, ficando com a precedente armazenada tal como estava anteriormente. A forma de gerir estas alterações, é inserir na dimensão dois campos do tipo data em que um será a data de início e o outro a data de fim, assim saber-se-á o intervalo de datas de cada registo (Caldeira, 2008; Kimball e Caserta, 2004; Kimball e Ross, 2002).

O exemplo da Tabela 6 reflecte uma alteração tipo 2 num registo.

Tabela 6 – Exemplo de Alteração Tipo 2

Chave do Colaborador	Número de Colaborador	Nome do Colaborador	Departamento
21	10	Carlos	Comercial
27	10	Carlos	Financeiro

O registo que tem a chave 21 foi alterado no campo departamento, assim foi inserido um novo registo, o 27, mantendo o anterior.

6.3.3. Alterações Tipo III

As alterações deste tipo só são utilizadas em casos específicos, uma vez que estas alterações implicam uma alteração estrutural. Quando ocorre uma alteração Tipo 2, é inserido um novo registo na dimensão, ou seja, uma nova linha. Numa alteração Tipo 3, o registo alterado é colocado em colunas, isto é, o valor da célula é actualizado com o novo valor, como se procede para as alterações Tipo 1, e o valor antigo é colocado numa nova coluna da dimensão (Kimball e Caserta, 2004; Kimball e Ross, 2002).

Tabela 7 – Exemplo de Alteração Tipo 3

Chave do Colaborador	Número de Colaborador	Nome do Colaborador	Departamento Actual	Departamento Antigo
21	10	Carlos	Financeiro	Comercial
27	10	Carlos	Financeiro	Financeiro

Como é mostrado na Tabela 7, inicialmente o valor do campo Departamento Actual era Comercial, foi alterado para Financeiro e todos os registos foram actualizados, ficando o histórico no atributo Departamento Antigo.

6.4.Dimensões Degeneradas

Uma noção interessante é a definição de dimensão degenerada, visto que esta tem apenas um único atributo, mas é um atributo muito relevante. Esta dimensão ocorre quando ao se retirar tudo de uma transacção apenas fica a identificação dessa transacção. Muitas tabelas de factos transaccionais contêm um número de controlo, tais como o número de uma factura, o número de uma ordem de compra ou o número de uma apólice (Kelley et al., 2003), é esse valor que dá origem à dimensão degenerada. Um caso típico é o talão de supermercado (Caldeira, 2008), retira-se do talão o cliente, a caixa, cada produto e respectivo valor, etc. Isto por sua vez dará origem a outras dimensões e, após se “retirar tudo” apenas sobra o número que identifica esse talão específico. Esse número terá que ser armazenado para que seja possível identificar cada talão e quais produtos foram comprados nesse talão. É aqui que entra a dimensão degenerada que apenas terá esse valor, que de outra maneira era impossível identificar o talão.

Esta dimensão, ao contrário das demais, não faz uso das surrogate key. Normalmente, os valores para a dimensão degenerada são únicos e com um

tamanho razoável. Contudo, se o valor for um atributo alfanumérico muito grande deve-se fazer uso das surrogate key (Becker, 2003).

Segundo Becker (2003), se no processo de carregamento das dimensões, existir uma cujas suas linhas aumentem ao mesmo ritmo que uma tabela de factos, então nessa dimensão existe um atributo que deveria ser uma dimensão degenerada.

6.5. Mini Dimensões

Segundo Caldeira (2008), existem algumas dimensões que ao longo do tempo acumulam muitos registos tornando-se pouco prático a utilização das alterações tipo 2, uma vez que contribuem ainda mais para o seu crescimento. Suponhamos que temos uma dimensão com 50 atributos, dos quais 5 são alterados frequentemente e muito analisados, ao contrário dos outros que raramente são modificados. Assim, esses 5 atributos são transferidos para outra dimensão denominada de Mini Dimensão.

As mini dimensões, normalmente, contêm atributos que são imprescindíveis para a análise do negócio. Como estes atributos são frequentemente utilizados e estão misturados com outros, têm um impacto negativo da performance das consultas no Data Warehouse. Deste modo, a colocação de determinados atributos, mais utilizados nas análises ou mais susceptíveis de alterações, numa mini dimensão é uma mais-valia para o sistema analítico (Caldeira, 2008).

6.6. Dimensão Lixo

Existe ainda outro tipo de dimensão que se intitula de Dimensão Lixo. Esta contém atributos que são frequentemente descartados por não se

enquadrarem nas outras dimensões, mas que podem ser úteis na análise dos processos de negócio.

Quando nos confrontamos com as fontes de dados faz-se uma identificação dos campos que podemos obter e começamos a agrupá-los em dimensões, posteriormente identificam-se as medidas para a tabela de factos. Nesta fase, são deixados de lado diversos indicadores por não se enquadrarem em nenhuma das dimensões, mas que até poderão ser úteis nas análises. A sua inclusão no Data Warehouse passa por agrupar estes campos em uma ou mais dimensões lixo (Kimball e Ross, 2002).

Se estes indicadores não fossem agrupados neste tipo de dimensão e cada um que não seja relativo a um mesmo assunto é separado, estaríamos a criar um problema para o Data Warehouse. Suponhamos a existência de 5 atributos, se cada um destes atributos der origem a uma dimensões, a tabela de factos teria que ter mais 5 campos para fazer referência a cada uma das tabelas e consultas mais complexas poderiam ter até cinco mais ligações entre tabelas. Por sua vez, se esses mesmos atributos forem agrupados numa dimensão lixo, a tabela de factos necessita de apenas mais um campo. Assim, esta dimensão contém atributos que em nada se relacionam, mas que podem ser bastante úteis fornecendo informações adicionais.

Segundo Caldeira (2008), os atributos desta dimensão têm a característica de variarem em poucas possibilidades e no caso de esta dimensão crescer em demasia os atributos devem ser distribuídos por outra dimensão lixo.

6.7.Views

É frequente que uma determinada dimensão tenha ligações a várias dimensões; é o caso da Dimensão Data que pelas suas características temporais tem uma enorme importância nos Data Warehouse (Caldeira, 2008). Quando é necessário ligar uma dimensão a outras dimensões não são estas que fazem essas ligações, uma vez que se assim fosse estaríamos a construir um “comboio” de ligações entre tabelas frequentemente encontrado nos sistemas transaccionais. Quando é necessário efectuar estes tipos de ligações são usadas as vistas ou views. As views não são mais que vistas sobre uma dimensão, em que ao se alterar o conteúdo da dimensão a vista também altera o seu conteúdo de forma automática. Voltando à Dimensão Data, vamos admitir que esta teria que se ligar a uma tabela de factos e a uma outra dimensão. Então ligar-se-ia a própria dimensão à tabela de factos e criar-se-ia uma vista da dimensão data que se ia ligar à outra dimensão. É muito importante rever os nomes da dimensão na vista ou seja, é necessário alterar os nomes dos atributos na vista para que o utilizador não se confunda entre estes e os da dimensão (Caldeira, 2008). Assim, olhando para o nome do atributo sabe-se rapidamente onde esse pertence.

7. Tabelas de Factos

As tabelas de factos são o elemento central do esquema em estrela, representando as transacções ou acontecimentos de determinado processo de negócio, armazenando todos os factos que a este dizem respeito. Os factos ou medidas são métricas resultantes de um processo.

Estas tabelas são constituídas por dois tipos de atributos, as chaves estrangeiras e os factos. Estes últimos são utilizados para calcular os resultados

do processo de negócio, enquanto as chaves estrangeiras ligam os factos às tabelas de dimensão. Esta ligação garante a integridade referencial e os atributos da dimensão descrevem os factos, caracterizando as métricas armazenadas. O conjunto ou subconjunto destas chaves estrangeiras constituem a chave primária da tabela de factos.

Todos os factos têm um nível de granularidade, isto é, o nível de detalhe que se pode caracterizar os factos. Segundo Kimball (2008) a primeira etapa de desenho das tabelas de factos é a definição deste grão. A granularidade é um aspecto muito importante no Data Warehouse, uma vez que é benéfico para os utilizadores permitindo olhar para os dados com diferentes níveis de detalhe. Quanto menor for a granularidade de um facto maior é o seu detalhe (Inmon, 2002). Kimball e Caserta (2004) defendem que deve ser armazenado o grão mais atómico para assim se conseguir produzir factos com outros níveis de granularidade, isto é, ao serem armazenados os factos com um elevado grau de atomicidade é possível agregar os dados de diversas formas através das dimensões. Estas agregações permitem calcular medidas sumarizadas a partir de tabelas de factos atómicas, ou seja, com muito detalhe. Cada tabela de factos apenas pode armazenar factos como mesmo nível de detalhe, ou seja, não é possível, por exemplo, numa mesma tabela guardar cada venda efectuada juntamente com a venda mensal.

7.1. Tipos de Factos

As medidas presentes nas tabelas de factos podem ser de três tipos, segundo o autor (Caldeira, 2008): aditivas, semi-aditivas ou não-aditivas. Os factos aditivos são aqueles que pelas suas características podem ser somados relativamente a qualquer dimensão. As agregações anteriormente referidas

são operações aditivas que apenas podem ser executadas sobre estes tipos de factos. Assim, uma medida aditiva quando é agregada, é somada, por exemplo, as vendas diárias, quando são agrupadas pela dimensão data para análises mensais, vendas diárias são somadas em cada mês. Os factos semi-aditivos são aqueles que apenas podem ser somados relativamente a algumas dimensões. Um exemplo de factos semi-aditivos é as contagens. Por último, os não-aditivos, que tal como o nome indica não podem ser somados ou agregados. São exemplos deste tipo as percentagens ou valores unitários.

7.2. Tipos de Tabelas de Factos

Além dos tipos de medidas existem dois grandes grupos de tabelas de factos, as tabelas de factos propriamente ditas e as tabelas de factos sem factos. As primeiras podem dividir-se em três tipos de tabelas, as tabelas de factos transaccionais, de sumarização periódica e de sumarização acumulada.

A Tabela 8 compara estes três tipos de tabelas de factos.

Tabela 8 – Comparação de Tipos de Tabelas de Facto (Kimball e Ross, 2002)

Característica	Transaccional	Sumarização Periódica	Sumarização Acumulada
Período	Ponto no tempo	Intervalos regulares	Indeterminado, duração do ciclo
Grão	Uma linha por transacção	Uma linha por período	Uma linha por ciclo de vida
Dimensão Data	Data da transacção	Data de fim do período	Várias datas para metas padrão
Factos	Actividade transaccional	Desempenho para intervalo de tempo pré-definido	Desempenho ao longo do ciclo

Estes tipos de tabelas de factos não são totalmente diferentes, na realidade complementam-se e partilham dimensões conforme. Deste modo, podem ser utilizadas em conjunto, onde as dimensões conforme são utilizadas

como filtros ou rótulos. Contudo, a gestão e ritmos de crescimento destas tabelas são bastante diferentes.

7.2.1. Tabelas de Factos Transaccionais

Neste tipo de tabelas de factos são armazenadas um linha por cada transacção e/ou acontecimento ocorrido na organização (Kimball e Ross, 2002). Deste modo, cada acontecimento que ocorre no sistema transaccional dá origem a um registo no modelo analítico, assim, a tendência para o seu crescimento é forte, sendo este crescimento proporcional ao ritmo das transacções (Caldeira, 2008).

Segundo Kimball e Caserta (2004), o exemplo padrão para este tipo de tabelas é as vendas num supermercado. Quando numa venda, o artigo passa pelo leitor do código de barras e soa um sinal sonoro, é inserido um registo no sistema transaccional. Este acontecimento dá origem a um registo no Data Warehouse. Assim, este tipo de tabelas é, dos três tipos de tabelas de factos, a maior e mais detalhada e normalmente está ligada a muitas dimensões.

7.2.2. Tabelas de Factos de Sumarização Periódica

Os acontecimentos da organização podem ser sumariados ou agregados segundo determinados intervalos de tempo, isto é, as tabelas de sumarização periódica agregam os dados periodicamente, fazendo com que estas tenham um grão mais grosseiro mas que pode variar dependendo da periodicidade (Caldeira, 2008). Kimball e Caserta (2004) referem que esta técnica é muito adequada a processos de longa duração, normalmente associado a dados financeiros. As tabelas de factos de sumarização periódica agrupam os factos periodicamente, ao contrário das tabelas de facto transaccionais onde são armazenados uma linha por acontecimento, neste tipo de tabelas de factos cada linha guardada contém um conjunto de dados

correspondente a um período de tempo especificado, por exemplo, um dia, uma semana ou um mês (Kimball e Ross, 2002).

Deste modo, estas tabelas armazenam em cada linha uma fotografia no final do período pré-determinado. Desta forma, como são armazenadas diversas fotografias dos períodos, esta é uma forma consistente de analisar a evolução do negócio da organização (Caldeira, 2008).

Segundo Caldeira (2008), o ritmo de crescimento destas tabelas é lento e está directamente relacionado com a periodicidade.

7.2.3. Tabelas de Factos de Sumarização Acumulada

As tabelas de factos de sumarização acumulada não são tão comuns como as anteriores. Estas armazenam uma linha que representa um período de tempo indeterminado que abrange todo o ciclo de duração de um evento (Kimball e Ross, 2002). Segundo Kimball e Caserta (2004), são utilizadas para descrever processos que têm um início e um fim, tais como, o atendimento e processamento de pedidos. Estas tabelas descrevem o início, a evolução e a conclusão de determinado acontecimento.

As tabelas de sumarização acumulada distinguem-se facilmente das outras anteriores, uma vez que possuem muitos indicadores temporais que correspondem às etapas do tempo de vida do evento. Como à partida não são conhecidos todos os atributos temporais, a actualização sucessiva destes atributos é uma característica particular destas tabelas (Caldeira, 2008).

7.2.4. Tabelas de Factos Sem Factos

Finalmente as tabelas de factos sem factos, segundo Caldeira (2008) diferem das anteriores e têm características muito especiais, pois não contêm nenhuma medida. Ao contrário das anteriores que têm dois grupos de

atributos, os factos ou métricas que são os indicadores do negócio e as chaves estrangeiras que caracterizam os factos, as tabelas de factos sem factos apenas contêm o grupo de chaves artificiais. Estas tabelas apenas servem para registar a ocorrência de determinado acontecimento, por exemplo, a contagem de cliques numa página Web ou o registo da assiduidade em determinada actividade. Caldeira (2008) mostra como este tipo de tabelas de factos pode ser útil na análise da assiduidade dos alunos, possibilitando responder a questões tais como, qual o docente com mais audiência e qual a disciplina com mais assiduidade. Assim, através da tabela de factos da Tabela 9 é possível obter estas análises.

Tabela 9 – Exemplo de Tabela de Factos Sem Factos

Chave da Data	Chave do Aluno	Chave da Disciplina	Chave do Docente	Chave da Sala de Aulas	Indicador de Presença
2	2735	10	8	3	1
2	2736	10	8	3	0
2	2739	10	8	3	1
3	2846	12	4	6	1
3	2043	13	10	2	1

Como podemos observar pela Tabela 9, o conjunto das cinco primeiras colunas são chaves estrangeiras para tabelas de dimensão, enquanto a coluna Indicador de Presença é um pseudo facto que contém o valor 1 para indicar a presença do aluno numa aula e o valor 0 quando o aluno falta. Este pseudo facto apenas serve para contar registos.

8. Ferramentas Utilizadas Para o Caso Prático

O caso prático foi desenvolvido em tecnologia Microsoft, nomeadamente através do Microsoft Business Intelligence Development Studio, que está incluído no SQL Server, neste caso na versão SQL Server 2005 Developer Edition.

8.1.BIDS

O BIDS (Business Intelligence Development Studio) é o Microsoft Visual Studio mas com tipos de projectos específicos do SQL Server Business Intelligence. É utilizado para desenvolver soluções de Business Intelligence e inclui ferramentas de extracção e transformação da informação, de processamento da informação e de visualização de informação. Cada uma destas ferramentas pode ser utilizada de forma independente, e neste caso, apenas foram utilizadas duas das ferramentas, ETL e OLAP.

8.1.1. SQL Server Integration Services

O SSIS é a ferramenta que permite extrair a informação, ou seja, permite desenvolver o processo de ETL. Contém objectos que permitem a ligação a distintas fontes de dados, tais como, ficheiros de texto, folhas de cálculos ou bases de dados, outros que permitem fazer operações e/ou transformações sobre os dados, e por fim, contém ainda objectos que permitem inserir os dados num destino, podendo este ser uma base de dados ou um simples ficheiro.

8.1.2. SQL Server Analysis Services

O SSAS é uma ferramenta de processamento analítico. Permite criar e gerir estruturas multidimensionais para posteriormente serem alvo de análises. A fonte de dados para o SSAS é os dados carregados pela ferramenta

anterior, que processa esses dados e armazena-os de uma forma especial, numa estrutura própria, em forma de cubo. Esta ferramenta está otimizada para permitir bons desempenhos ao nível de velocidade de acesso aos dados (Jacobson e Misner, 2006).

8.2. Microsoft Office Excel

O Microsoft Office Excel foi a ferramenta escolhida para a visualização da informação. Esta é uma aplicação com a qual os gestores estão familiarizados e a sua simplicidade em se ligar ao Data Warehouse foram as razões que levaram à sua escolha. A versão utilizada foi a 12.0, mais conhecida por Excel 2007. Esta ferramenta é uma folha de cálculo que através da funcionalidade Tabelas e Gráficos Dinâmicos permite obter dados a partir do Data Warehouse e criar tabelas ou gráficos com um simples arrastar de campos.

9. Caso de Estudo

Como já foi referido anteriormente, o projecto desenvolvido tem como objectivo a construção de um Data Warehouse para um grupo empresarial. Este grupo é detentor de 4 empresas e o objectivo final é oferecer aos administradores/gestores uma ferramenta que permita efectuar análises sobre o grupo de uma forma transparente.

É possível obter a partir do ERP de cada empresa relatórios que permitem efectuar alguma análise sobre os dados organizacionais, mas de uma forma isolada e independente. Quando se pretende uma análise do grupo, torna-se bastante mais complicado obter esses mapas e dependendo do relatório pode ser mesmo impossível realizar uma análise fiável.

Assim, através da construção de um Data Warehouse que contém dados relativos às empresas do grupo será possível, por uma lado, aos administradores do grupo efectuarem consultas sobre os dados de ambas as empresas e, por outro, aos gestores de cada empresa apenas terem conhecimento dos dados dessa mesma empresa, uma vez que não lhes interessa e nem convém terem conhecimento do estado das demais empresas.

O Data Warehouse terá um papel fundamental na análise e identificação de factores que podem tornar o grupo mais competitivo no mercado.

9.1. Análise do Processo de Negócio

Cada uma das empresas tem o seu sistema operacional independente onde a informação é armazenada em bases de dados. Deste modo, existem

quatro bases de dados, uma para cada sistema operacional de cada empresa, mais dados externos à empresa.

Após a análise do negócio do grupo, foram identificados processos de negócio. Deste modo, foram identificados 5 processos de negócio que levaram à criação da matriz em bus presente na Tabela 10.

- Processo Venda – cada empresa do grupo efectua vendas a determinados clientes, assim é essencial analisar e identificar quais são os melhores clientes e o local para onde são efectuadas essas vendas ou, por exemplo, qual a empresa com maior volume de vendas.
- Processo Compra – é necessário saber quais são as compras do grupo para que seja possível minimizar as despesas e identificar melhores negócios para compras de determinados produtos para o grupo.
- Processo Recepção de Encomendas – porque não se sabe se as encomendas se tornam em vendas, este processo é uma mais-valia como suporte de análises. Através dele é possível analisar se determinado cliente efectua encomendas e a respectiva compra do que encomendou.
- Processo Dívidas de Clientes – para ajudar na identificação de clientes que têm o hábito de se atrasar no pagamento das suas dívidas, este é um processo delicado mas que para o grupo fornece informação preciosa.
- Processo Dívidas a Fornecedores – os administradores do grupo por vezes não sabem ou não têm certezas sobre potenciais dívidas das empresas, este processo permite realizar análises sobre eventuais

dívidas que as empresas possam ter e/ou atrasos nos pagamentos aos fornecedores.

Tabela 10 – Matriz bus do caso prático

	Empresa	Data	Relógio	Local	Artigo	Cliente	Fornecedor	Documento	Utilizador	Data Vencimento	Vendedor	Dimensionalidade
Processo Venda	x	x	x	x	x	x	x	x	x	x		9
Processo Compra	x	x	x		x	x	x	x				7
Processo Dividas de Clientes	x	x		x		x	x		x			6
Processo Dividas a Fornecedores	x	x				x	x		x			5
Processo Encomenda	x	x	x	x	x	x		x		x		8

Como podemos verificar na matriz, as linhas representam os processos de negócio e as colunas as dimensões que irão caracterizar os factos. Cada processo de negócio dá origem a um Data Mart que irá conter dados das diversas empresas. Cada processo de negócio tem um nível de detalhe. Assim, o grão é definido segundo a Tabela 11.

Tabela 11 – Granularidade dos processos de negócio

Processos de Negócio	Granularidade
Processo Venda	Venda-a-venda de artigos
Processo Compra	Compra-a-compra de artigos
Processo Dividas de Clientes	Dívida-a-dívida de clientes
Processo Dividas a Fornecedores	Dívida-a-dívida a fornecedores
Processo Encomenda	Recepção encomenda-a-encomenda de artigos

Deste modo, o grão do processo de compras é cada compra de determinado artigo efectuada por cada uma das empresas. Por sua vez, o processo de vendas e encomendas representa cada venda e cada encomenda de cada artigo efectuada a cada uma das empresas. Por fim, os processos de dívidas de clientes e a fornecedores terão como grão cada dívida contraída por cada cliente a cada empresa e destas a cada fornecedor, respectivamente.

As dimensões são representadas pelas colunas da matriz em bus e têm uma particularidade bastante interessante. Estas consolidam os dados entre as empresas, isto é, os campos ou atributos das dimensões são preenchidos pelas diversas empresas, mas sem que estes sejam duplicados. Deste modo, determinado fornecedor que está presente em duas ou mais empresas, apenas será inserido na respectiva dimensão uma única vez.

9.2.Modelo de Dados

A partir da construção da matriz em bus desenvolvem-se os esquemas em estrela, onde cada processo de negócio dá origem a um esquema em estrela. Assim temos os seguintes esquemas em estrela:

- Esquema em estrela do processo venda

O esquema em estrela correspondente a este processo de negócio encontra-se no Anexo 1. No centro do esquema temos a tabela de factos que representa cada linha de cada venda que determinado cliente efectua a uma empresa. Deste modo, os campos de medidas que são armazenados são o valor da venda, a quantidade vendida, o valor do desconto efectuado ao cliente, o custo que representa para a empresa o produto vendido e a quantidade de caixas vendidas. Esta unidade é bastante importante uma vez que é a unidade base do grupo, assim uma unidade de determinado artigo será convertido para a unidade caixa. A partir destes factos aditivos pode-se achar ou calcular outras medidas tais como, venda média, valor máximo ou valor mínimo. À volta da tabela de factos e ligada a esta estão as dimensões que caracterizam cada venda, assim temos a empresa que efectuou a venda, o artigo que foi vendido, a data e hora da venda, o cliente, o local, o utilizador

que registou a venda no sistema operacional, o vendedor associado à venda e o respectivo documento de venda.

- Esquema em estrela do processo compra

O processo compra tem o esquema em estrela no Anexo 2, e tal como no processo anterior a tabela de factos está no centro onde as medidas que se encontram armazenadas são o valor da compra, a quantidade de artigos comprados e o desconto realizado. Tal como no processo venda, estas medidas podem levar ao cálculo de outras medidas. Comparativamente com o esquema anterior, este não contém as dimensões Vendedor e Local, mas contém a dimensão Fornecedor ao invés da Cliente. O fornecedor representa a entidade a quem foi efectuada a compra.

- Esquema em estrela do processo recepção de encomendas

Este processo é semelhante ao processo venda, e a sua representação esquemática apresenta-se no Anexo 3. Confrontando os dois esquemas, este não tem a dimensão documento, uma vez que apenas existe um único tipo de documento, Encomenda de cliente. Em relação às medidas, o esquema indica que são armazenados também três campos, o valor e a quantidade encomendada e o respectivo desconto que será efectuado.

- Esquema em estrela do processo dívida de clientes

Este processo representa todas as dívidas de clientes às empresas e o seu esquema em estrela está no Anexo 4. Tal como nos outros esquemas as dimensões à volta da tabela de factos caracterizam as medidas. Assim temos que cada facto é caracterizado pela Empresa a quem determinado Cliente deve, o Local geográfico, a data e o respectivo documento da dívida contraída e, por fim, a Data de Vencimento da dívida. Esta data é o dia limite que o

cliente tem para regularizar a sua situação. Como medidas para análise neste esquema temos o valor total da dívida, valor em dívida (caso o cliente pague parte da dívida), valor não vencido (valor da dívida antes da data de vencimento), valor vencido (valor que à data de vencimento ainda não foi regularizado), valor pago antes do vencimento (valor total pago pelo cliente antes do fim da data de vencimento) e valor pago após vencimento (valor total pago pelo cliente depois da data de vencimento da dívida).

- Esquema em estrela do processo dívida a fornecedores

Este esquema é bastante semelhante ao anterior, mas representa as dívidas contraídas por cada uma das empresas do grupo a uma entidade fornecedora. As medidas presentes no esquema em estrela são as mesmas do esquema anterior, mas sempre das empresas em relação ao fornecedor. Por exemplo, para o valor total da dívida, este representa o valor da dívida contraída por uma empresa a um fornecedor numa determinada data, com uma data de vencimento limite para ser paga, registada por um utilizador e respeitante a um documento específico. Este esquema encontra-se no Anexo 5.

9.3.Fonte de Dados

A análise dos dados que irão preencher os atributos das dimensões é um ponto fulcral no desenvolvimento deste Data Warehouse. Isto porque os valores dos atributos serão preenchidos a partir dos diversos sistemas transaccionais, mas com a nuance que determinado registo pode encontrar-se em mais do que uma base de dados e neste caso deverá ser carregado uma única vez. Deste modo, cada empresa tem os seus clientes, fornecedores, vendedores e artigos. Mas, as empresas também podem partilhar estas

entidades, por exemplo, um mesmo cliente pode ser cliente das diversas empresas e estar presente em cada sistema transaccional. Por sua vez, no Data Warehouse deverá aparecer uma só vez, uma vez que é a mesma entidade.

Deste modo, a consolidação destas dimensões é um factor preponderante para o sucesso deste sistema analítico. É este ponto que faz toda a diferença entre o sucesso e o insucesso do sistema. Se não for possível analisar as vendas a determinado cliente pelas diversas empresas, não estamos a oferecer nada de novo aos administradores do grupo.

Após a análise do processo de negócio da organização e dos diversos sistemas transaccionais conseguiu-se perceber que quando um número de cliente existe numa BD de determinado sistema transaccional, se esse mesmo cliente existe noutro sistema ele terá o mesmo número de cliente. Porém, poderá ter atributos distintos, tal como mostra a Tabela 12:

Tabela 12 – Exemplo do mesmo cliente em sistemas transaccionais de empresas distintas

Sistema Transaccional	Nº de Cliente	Nome do Cliente	Morada	...
ERP Empresa 1	304	João Manuel Baptista	Rua do Bispo	...
ERP Empresa 2	304	João Baptista		...
ERP Empresa 3				...
ERP Empresa 4	304	João M. Baptista	Rua Bispo, 10	...

No exemplo anterior, podemos verificar que o mesmo cliente não tem o nome exactamente igual entre os sistemas, no ERP da Empresa 2 não tem morada e na Empresa 3 não existe.

Assim, após esta conclusão, procedeu-se à actualização dos campos não preenchidos para que de certa forma os registos fossem consistentes. Este foi um processo complexo que necessitou de ser validado por um especialista do ERP da organização. Este processo foi desenvolvido com na linguagem SQL, onde através de consultas foi possível fazer a actualização desses campos.

Após este procedimento, o cliente da tabela anterior ficaria com o campo Morada preenchido com o valor Rua do Bispo.

Depois de executar a actualização dos campos, os dados foram novamente analisados. As diferenças encontradas nesta análise foram nos campos que já haviam sido preenchidos, como é o caso do nome do cliente na tabela anterior. A actualização desses campos não foi possível por uma questão de estrutura organizacional. Chegou-se então à conclusão que dependendo da empresa os dados estavam mais ou menos correctos. Analisando novamente o exemplo anterior e focando o campo Nome, podemos concluir que o registo presente na Empresa 1 é mais completo comparado com o registo da Empresa 2 e o mesmo sucede para os restantes clientes.

9.4. Processo ETL

O processo ETL é uma tarefa bastante importante que ocupa a maior parte do tempo de desenvolvimento de todo o sistema analítico. Esta é a fase do projecto em que são seleccionados os dados dos sistemas transaccionais, transformados e carregados no Data Warehouse.

Como já foi referido anteriormente, o grupo é constituído por quatro empresas, portanto são quatro os sistemas transaccionais de onde serão extraídos os dados. Para um maior detalhe da informação são utilizados dados externos ao grupo, nomeadamente dados geográficos. Estes dados fornecem um maior detalhe geográfico que de outra maneira era impossível usufruir.

Deste modo, as tabelas de factos e as dimensões serão carregados a partir dos sistemas de todas as empresas. Como requisito, e como já foi mencionado, o Data Warehouse tem que consolidar os dados das diversas

empresas. As tabelas de factos armazenam os dados sobre determinado assunto para cada uma das empresas. A questão fulcral sobre a consolidação dos dados encontra-se nas dimensões, uma vez que determinados dados estão inseridos em mais do que um sistema transaccional. Estes dados que de alguma forma estão duplicados entre sistemas do grupo devem aparecer uma única vez no Data Warehouse, ou seja, consolidados no Data Warehouse. Assim, foram estabelecidas regras de prioridade para se ter o conhecimento, aquando a existência dos mesmos dados, de onde devem ser carregados os dados. Como já foi exposto anteriormente, sempre que determinado registo estiver presente na base de dados da Empresa 1 ou numa das outras, prevalece o registo da Empresa 1, caso exista na Empresa 2 e na Empresa 3 ou Empresa 4, prevalece o da Empresa 2, e por fim, se existir na Empresa 3 e na Empresa 4, impera o da Empresa 3. Desta forma, o Data Warehouse irá conter a informação o mais correcta possível e, muito importante, não duplicada.

Assim, o processo de ETL terá que extrair os dados em primeiro lugar da Empresa 1, depois da Empresa 2 à excepção daqueles que existem na 1, depois os da Empresa 3 mas que não existem nem na Empresa 1 nem na Empresa 2, e por fim, os dados da Empresa 4 mas que não existem nas restantes. Esta regra aplica-se à extracção, transformação e carregamentos dos dados das dimensões Artigo, Cliente, Fornecedor e Utilizador. No caso desta última, a dimensão Utilizador é ainda complementada com dados a partir de um outro sistema. Quando um colaborador se liga ao sistema operacional, liga-se através de um utilizador de domínio, a gestão dos utilizadores de domínio é feita a partir de uma Active Directory que contém informação complementar sobre os utilizadores. Como existe um campo que faz a ponte entre o sistema transaccional e a Active Directory é possível obter esses dados. A dimensão

Vendedor é a excepção à regra, cada vendedor apenas existe num sistema transaccional, ou seja, um vendedor pertence apenas a uma só empresa do grupo. Deste modo, todos os vendedores de ambas as empresas têm que ser carregados no Data Warehouse. Os códigos dos vendedores repetem-se entre empresas tornando a sua distinção bastante complexa, isto porque certamente existirão quatro códigos iguais para vendedores distintos, um por cada empresa. Assim, para o processo ETL conseguir distinguir os vendedores de cada empresa para poder aplicar as devidas alterações quando estas ocorram, criou-se mais um campo nesta tabela que indica a que empresa pertence esse registo (vendedor). Este campo serve apenas este propósito pelo que será oculto e não estará disponível no sistema analítico. Como o sistema transaccional tem pouca informação sobre os vendedores e as funcionalidades são vagas, o grupo adquiriu uma aplicação que permite efectuar a gestão de vendedores. Assim, dos dados sobre os vendedores são adquiridos a partir de duas fontes de dados. Relativamente à dimensão Data e Data Vencimento, estas terão valores desde 1 de Janeiro a 31 de Dezembro do menor ao maior ano existentes entre os sistemas operacionais. A dimensão Relógio terá todas as horas e minutos do dia e não depende de nenhum sistema, uma vez que todo o grupo tem as horas no mesmo formato. Quanto à dimensão Documento, esta também é uma dimensão genérica para o grupo, uma vez que todas as empresas utilizam os mesmos tipos de documento com a mesma tipologia, sendo este o formato standard dos sistemas operacionais. Por fim, os dados que alimentam a dimensão Localidade são dados de fontes externas, isto é, não pertencem ao grupo. São dados que são adquiridos através de outras entidades, neste caso, a entidade fornecedora destes dados são CTT – Correios de Portugal. O grupo tem clientes em diversos países e estes países

estarão representados na dimensão, mas apenas o país Portugal terá mais detalhe, ou seja, apenas Portugal terá todos os atributos da dimensão preenchidos enquanto os outros países terão esses campos preenchidos com o valor “Não disponível”. Os CTT fornecem os dados geográficos de Portugal em três ficheiros de texto onde os dados se encontram separados por “;”. É a partir da leitura destes ficheiros em conjunto com a informação geográfica dos clientes e fornecedores que se chega aos dados que alimentam a dimensão Localidade.

Quanto às alterações ocorridas nos registos das dimensões, os tipos de alterações implementados foram os Tipos I e II. Como já foi referido anteriormente, estes dois tipos de alterações são os mais utilizados. O Tipo I substitui no próprio registo o campo que foi alterado, enquanto o Tipo II cria um novo registo guardando o histórico desse registo. A maioria dos campos tem implementado a alteração tipo II, isto porque este tipo de alterações é benéfico para o Data Warehouse uma vez que guarda um histórico dos acontecimentos. O tipo de alteração para cada atributo foi definido através da análise detalhada de cada campo. Existem atributos que pelas suas características e aliadas ao processo de negócio do grupo só são modificados, nos sistemas transaccionais, quando contém valores errados. Por exemplo, o volume de determinado artigo só é alterado se for detectado que esse valor não está correcto. Esse valor é corrigido por um colaborador através do sistema e no Data Warehouse o atributo é também actualizado pelo processo ETL. Deste modo, todos os atributos que detêm estas características foram definidos com alterações do Tipo I.

As medidas que são carregadas nas tabelas de factos são dados numéricos retirados de documentos específicos sobre determinado

acontecimento. Assim, para os processos de vendas e compras as medidas carregadas vêm de cada linha desses documentos específicos, por exemplo Facturas ou Notas de Crédito para os clientes ou vindas de fornecedores. No caso do processo encomendas de clientes as medidas resultam de documentos existentes para esse efeito, Notas de Encomenda. Relativamente aos processos dívidas de clientes e dívidas a fornecedores, estes advêm dos mesmos documentos utilizados nos processos vendas e compras, respectivamente, mas que por algum motivo ainda não foram pagos. O processo de ETL terá que percorrer cada linha desses documentos inserindo as medidas presentes em cada uma delas no Data Warehouse ligando essas medidas a cada uma das respectivas dimensões.

Deste modo, podemos resumir toda a estrutura do processo ETL através da Figura 5.

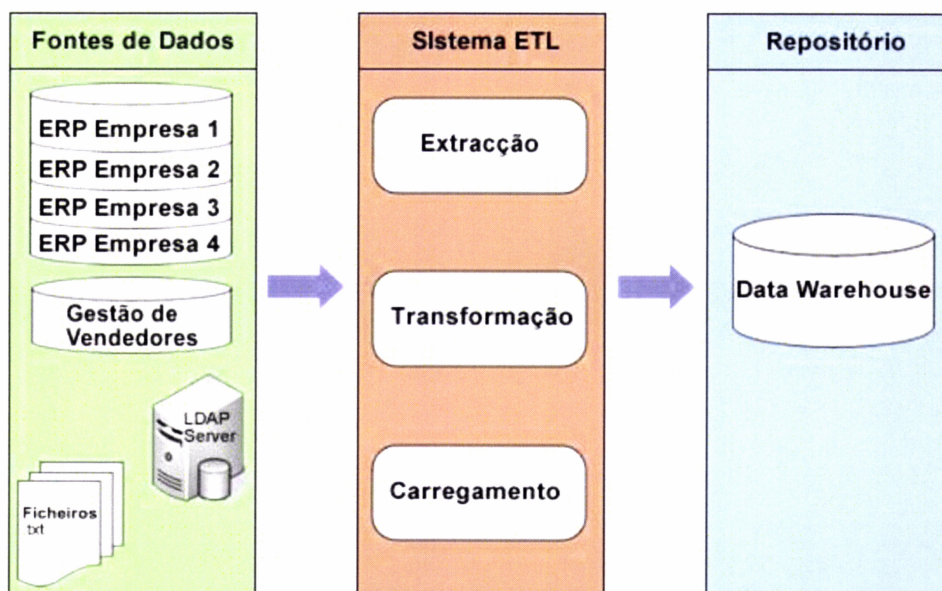


Figura 5 – Arquitectura do processo ETL do caso prático

Como fonte de dados temos quatro bases de dados dos sistemas transaccionais das empresas do grupo, mais uma aplicação de gestão de vendedores, a Active Directory com informação dos utilizadores e os ficheiros texto com informação geográfica. A tarefa de extração do sistema ETL terá

que ir ler os dados a cada uma destas fontes, depois a fase de transformação, como o próprio nome indica, tem como objectivo limpar e transformar os dados tornando-os consistentes e sem erros e por fim, a fase de carregamento que coloca os dados no nosso sistema analítico.

9.5.Cálculo de Medidas

Os factos ou medidas presentes nos esquemas em estrela permitem retirar conclusões e efectuar análises sobre valores totais de vendas ou quantidade total de compras segundo determinadas dimensões. Por exemplo, o valor total de vendas e o valor total do custo dos produtos vendidos por cada família de artigos em cada país num determinado mês do ano. Mas se o objectivo da análise se prender com a Margem Bruta $[(\text{valor da venda} - \text{custo do produto})/\text{valor da venda}]$, essa análise não é possível de uma forma directa, uma vez que este cálculo é uma percentagem, logo é um facto não aditivo. Os dados podem ser agrupados de diferentes maneiras segundo as dimensões do esquema em estrela, as medidas presentes no esquema em estrela são factos aditivos e são somados na agregação, ao contrário dos factos não aditivos que pelas suas características não podem ser somados. Desta forma, estes dados não podem estar presentes no esquema em estrela, mas são imprescindíveis para diversas análises. É nesta fase que as ferramentas de análises são importantes, tal como o SQL Server Analysis Services utilizado no projecto. Esta ferramenta permite desenvolver estes campos calculados. Através de uma linguagem de programação multidimensional, o MDX, permite efectuar este tipo de cálculos a partir de factos aditivos. Desta forma, é possível criar a medida Margem Bruta, para tal apenas é necessário escrever a fórmula que a ferramenta encarrega-se de efectuar o cálculo sempre segundo as dimensões

que forem escolhidas, de uma forma transparente para o utilizador. Assim, se o cálculo for relativo a cada ano de cada empresa, a ferramenta irá efectuar o cálculo segundo essas dimensões, caso seja por cada mês de um determinado ano para todo o grupo, a ferramenta calcula perante essa situação.

A Figura 6 mostra o funcionamento dos cálculos através do Analysis Services com o exemplo da Margem Bruta, como podemos verificar através da ferramenta este cálculo torna-se bastante simples.

The screenshot displays the configuration interface for a measure in Analysis Services. It includes sections for Name, Parent Properties, Expression, and Additional Properties.

Name: [Margem Bruta]

Parent Properties:
Parent hierarchy: MEASURES
Parent member: [Empty field] [Change]

Expression:
$$\frac{([Measures].[Valor da Venda] - [Measures].[Custo do Produto])}{[Measures].[Valor da Venda]}$$

Additional Properties:
Format string: "Percent"
Visible: True
Non-empty behavior: Valor da Venda, Custo do Produto

Figura 6 – Cálculo de Medidas em MDX (Cálculo da Margem Bruta)

Nos anexos 6 e 7 podemos verificar o funcionamento deste tipo de factos. Assim, no Anexo 6 temos uma análise do total de vendas, custo e Margem Bruta por empresa no ano 2009. Como podemos verificar, as medidas Valor da Venda e Custo do Produto são somados no que diz respeito ao total das empresas. Quanto ao campo Margem Bruta, é calculado para cada empresa, mas o total não é a soma das margens brutas de cada empresa mas sim através do cálculo dos totais $[(\text{Total do Valor da Venda} - \text{Total do Custo do Produto}) / \text{Total do Valor da venda}]$. O anexo 7 é um outro exemplo onde podemos comprovar o funcionamento do cálculo. O campo é calculado em relação ao grupo (todas as empresas), por Ano e Mês. Assim, temos o cálculo

para cada mês do ano 2009, o total para o ano 2009 e para o ano 2010 e por fim, o total para estes dois anos. Podemos concluir mais uma vez que o resultado nunca é um somatório, mas novamente um cálculo.

A medida apresenta-se ao utilizador como uma medida dita normal. O campo Margem Bruta aparece nos relatórios tal como, por exemplo, o Valor da Venda, não causando nenhum tipo de confusão ao utilizador sobre a sua utilização. Deste modo, este tipo de ferramentas facilita a criação de medidas calculadas, que de outra forma seriam muito difíceis de serem criadas.

9.6.Requisitos de Segurança

A segurança da informação é imprescindível em qualquer sistema e os sistemas analíticos não são excepção. Uma vez que estamos a consolidar dados de diversas empresas esta análise teve de ser muito cuidadosa. A segurança da informação é bastante importante neste Data Warehouse, é necessário definir quem tem acesso de leitura e/ou escrita dos dados. Mesmo apenas no acesso de leitura aos dados é necessário perceber quem pode ver o quê e foi efectuada uma análise no sentido de perceber e resolver esta questão, chegando-se ao organograma da Figura 7.



Figura 7 – Organograma do Grupo Empresarial

Com base na estrutura organizacional anterior, os administradores do grupo terão acesso aos dados do grupo, isto é, de todas as empresas sem excepção. Consequentemente terá que haver a garantia de que os gestores de

cada empresa apenas possam consultar e analisar os dados dessa mesma empresa. Por outras palavras, o Data Warehouse deve conter perfis de acesso aos dados, para assim se identificar que utilizadores podem aceder a determinados dados.

9.7. Análise de Resultados

A disponibilização do Data Warehouse aos utilizadores é um processo delicado que pode por em causa todo o projecto até agora desenvolvido. Antes de cada disponibilização de cada processo de negócio é imprescindível testar o correcto funcionamento da ferramenta e a consistência dos dados. O Data Warehouse foi exaustivamente testado para que não existam falhas e o resultado não podia ser melhor. Foram também efectuados testes de aceitação junto dos utilizadores à medida que estes aprendiam a servir-se do sistema.

A utilização deste sistema analítico permite aos utilizadores serem autónomos, ao contrário do que se sucedia até aqui, onde a maioria das vezes os utilizadores tinham que recorrer a alguém para os ajudar a efectuar as suas consultas. Actualmente qualquer utilizador consegue através desta ferramenta efectuar as suas consultas de uma forma bastante rápida, em poucos minutos consegue-se obter um relatório com toda a informação pretendida. Comparando este sistema com a forma antiquada que era utilizada pelo grupo, o tempo de espera para gerar um relatório é agora muito menor. Criar um relatório a partir do Data Warehouse levará, dependendo das necessidades, cerca de 5 minutos a ser concluído. Mas, sem a utilização desta aplicação, chega a demorar horas. Outro aspecto bastante interessante é a actualização dos dados após a criação de um relatório. Actualizar os dados de um relatório

através da forma que o grupo utilizava demoraria horas, por outro lado, ao se utilizar o Data Warehouse um relatório será actualizado em apenas alguns segundos.

Do Anexo 8 ao Anexo 17 existem diversos mapas ou análises de exemplo que os administradores/gestores irão realizar para apoiar as suas decisões. Estas são apenas algumas das imensas formas de agrupar a informação e melhor se saber o actual estado do grupo e a evolução do seu negócio. Através destes mapas exemplo os decisores obtêm as suas respostas para melhor poderem tomar decisões. Deste modo, pode-se adiantar que o Data Warehouse responde às questões para o qual foi projectado.

10. Conclusões

Os Data Warehouse podem ser uma mais-valia para uma organização, uma vez que permitem análises sobre dados da organização, facilitando e apoiando a tomada de decisões dos utilizadores. Sendo os gestores e os administradores os que mais utilizam estes sistemas, procurando bases para fundamentar e apoiar as suas decisões. Este projecto não foge à regra e também ele é uma mais-valia para o grupo.

O Data Warehouse assume um papel fundamental no estudo comercial do grupo, permitindo identificar pontos fortes e/ou fracos, contribuindo para o estudo de soluções que possam aumentar a competitividade das empresas. Este projecto de Data Warehouse satisfaz os requisitos e objectivos propostos inicialmente na sua totalidade, contribuindo para uma evolução positiva do negócio de cada empresa e, consequentemente para o grupo.

O desenvolvimento de um Data Warehouse não é uma tarefa trivial, bem pelo contrário, é um processo deveras complexo pois obedece a uma sequência de tarefas que têm que ser bem planeadas e que não devem ser simplesmente ignoradas. Foram algumas as dificuldades encontradas ao longo de todo o projecto, sendo a tarefa mais complicada a análise do processo de negócio. Na fase de análise do processo de negócio, em conjunto com os respectivos utilizadores, sempre que surgiu alguma dúvida e foi colocada uma questão as respostas foram sempre da mesma forma, sobre esse tema nós usamos este *layout*. O *layout* não é mais do que um relatório com determinados campos e respectivos valores, pelo que foi bastante complicado chegar à definição dos processos de negócio e como eles se encontram dentro

do grupo. A outra grande dificuldade prendeu-se com o pouco conhecimento dos sistemas operacionais tornando a análise de dados bastante complexa. O não conhecimento da estrutura das bases de dados transaccionais tornou qualquer tentativa de compreender os dados muito complicada. O apoio de um especialista e gestor dos sistemas operacionais foi, neste ponto, fundamental para o conhecimento das fontes de dados.

Todos os utilizadores encontravam-se familiarizados com a interface de análise de dados, o Excel. Pelo que não existiu uma grande adaptação por parte dos utilizadores e apenas lhes foi explicado como ligar o Excel ao sistema OLAP e, posteriormente, de uma forma breve, como gerar um relatório. Durante o período de uma semana, tempo escolhido pelo grupo, os utilizadores começaram a realizar os seus mapas de análise, sendo-lhes prestado apoio para que tirassem qualquer dúvida que surgisse. Findo este tempo o resultado foi bastante positivo, na maioria dos casos os utilizadores ficaram muito agradados com os resultados que obtiveram. Unicamente um utilizador com menos conhecimento do Excel e nenhum conhecimento da funcionalidade de tabelas e gráficos dinâmicos necessitou de auxílio, mas que se adaptou depressa a esta funcionalidade.

O Data Warehouse veio facilitar em muito a análise do negócio do grupo. O tempo de resposta dos Data Warehouse comparativamente com os sistemas OLTP é substancialmente menor, diminuindo também o tempo para gerar um relatório. Os relatórios eram todos eles gerados directamente dos sistemas operacionais individualmente, pelo que quando surgia a necessidade de obter um determinado relatório com dados de todas as empresas, o utilizador teria que extrair um relatório de cada uma das empresas e depois juntar esses relatórios num só. Esta é uma tarefa que demora o seu tempo até

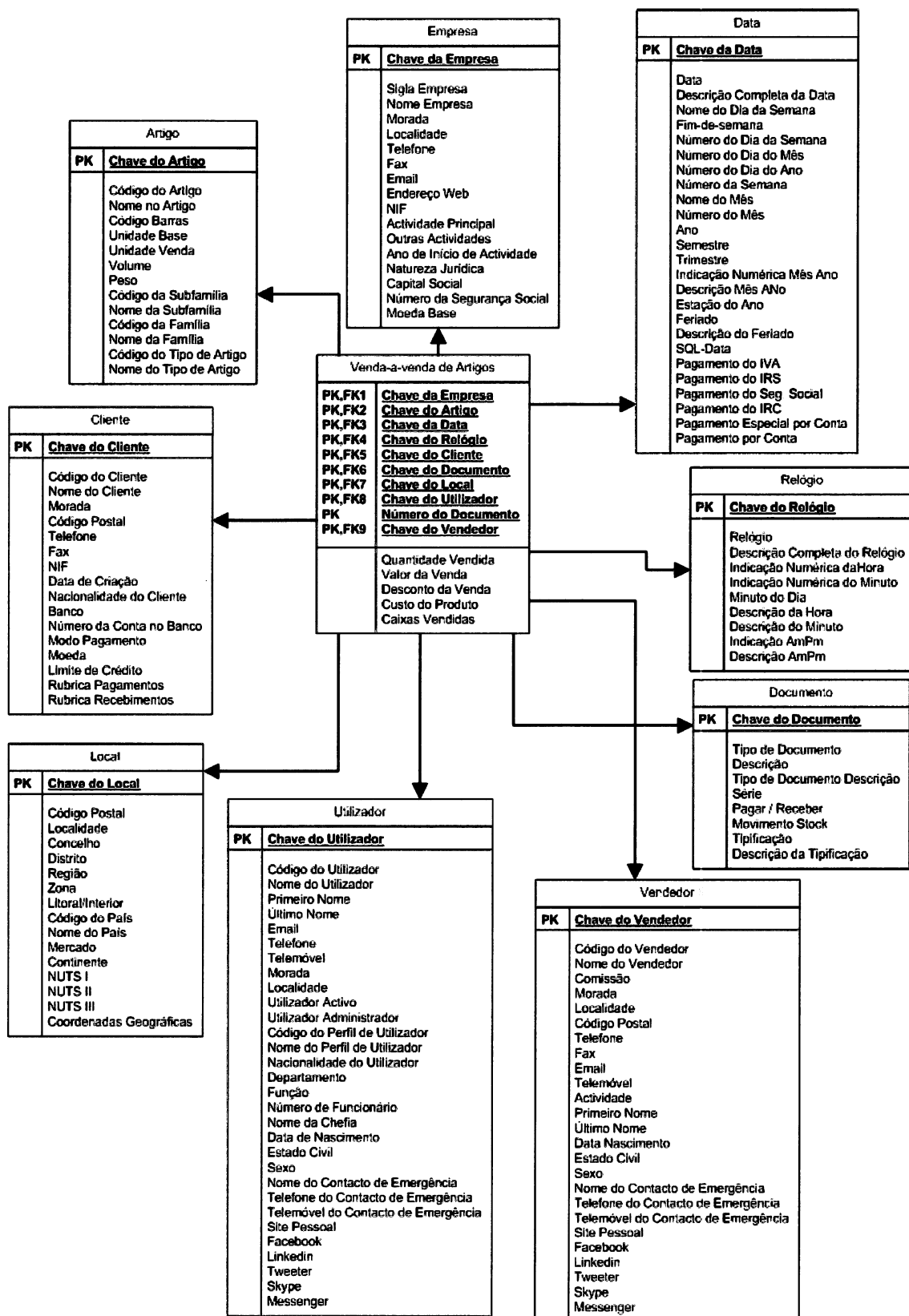
estar pronta para ser alvo de análises. Por vezes, dependendo do tipo de análise que se pretende este método não era fiável ou levaria bastante tempo até se conseguir um relatório. Como solução para este problema os utilizadores começaram a pedir a um especialista em SQL para criar consultas para as suas necessidades, ficando assim dependentes da disponibilidade dessa pessoa. O Data Warehouse veio por um lado, fazer com que os utilizadores apenas dependam de si próprios para a realização de mapas e, por outro, não se verifica qualquer problema em gerar relatórios para uma empresa específica ou para todo o grupo, ou ainda realizar análises apenas para algumas dessas empresas. A realização dos relatórios fica completamente a cargo dos utilizadores de uma forma simples e directa.

10.1. Trabalho Futuro

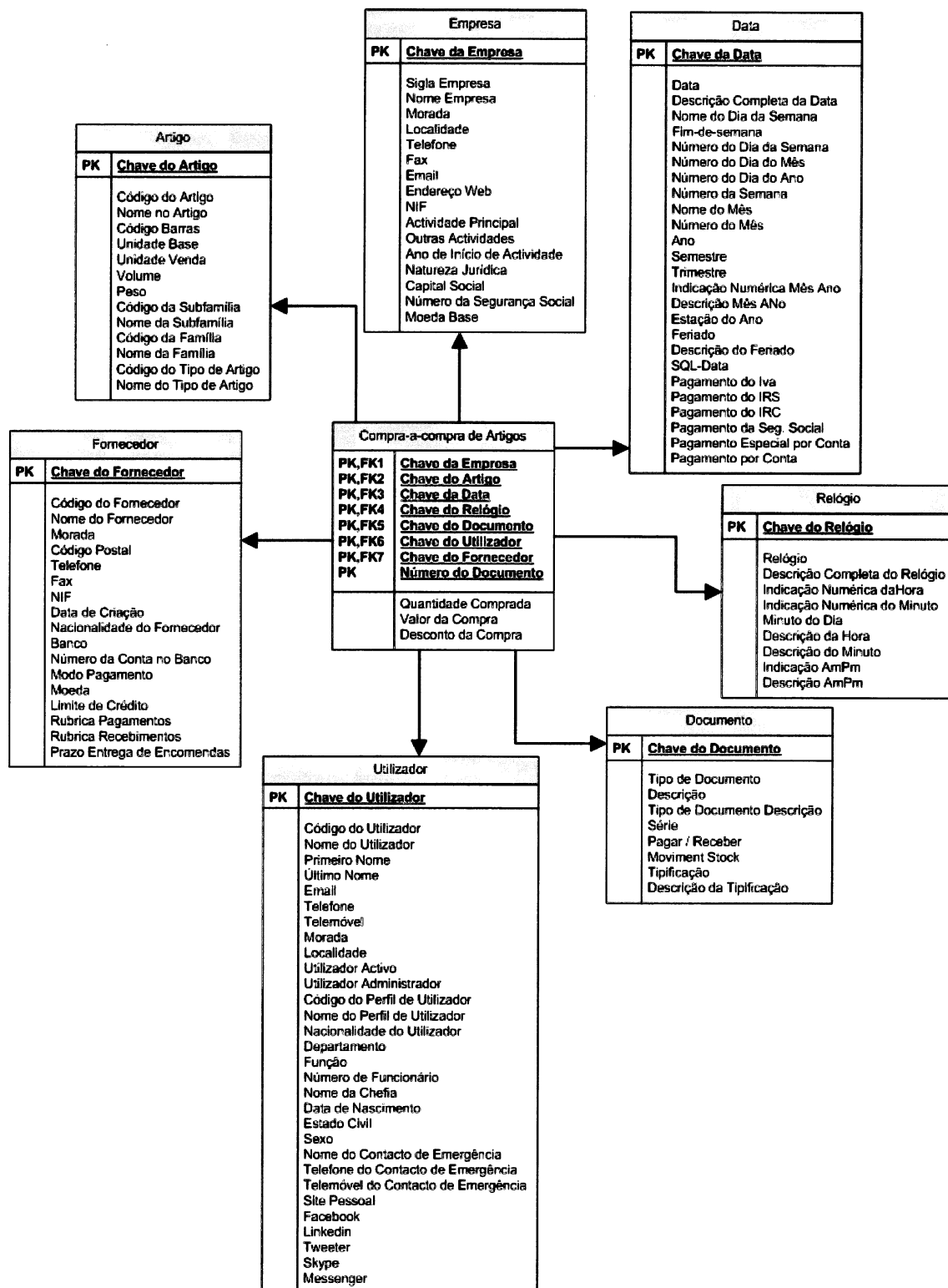
O grupo adquiriu recentemente uma aplicação que permite efectuar previsões de vendas. A integração destes dados no Data Warehouse será muito importante, não só pelas análises previsionais mas também por análises que permitam a comparação entre o previsto e o real. Está prevista a sua utilização para o próximo ano, pelo que a análise desta ferramenta e posterior integração com o Data Warehouse está neste momento em fase de avaliação.

Anexos

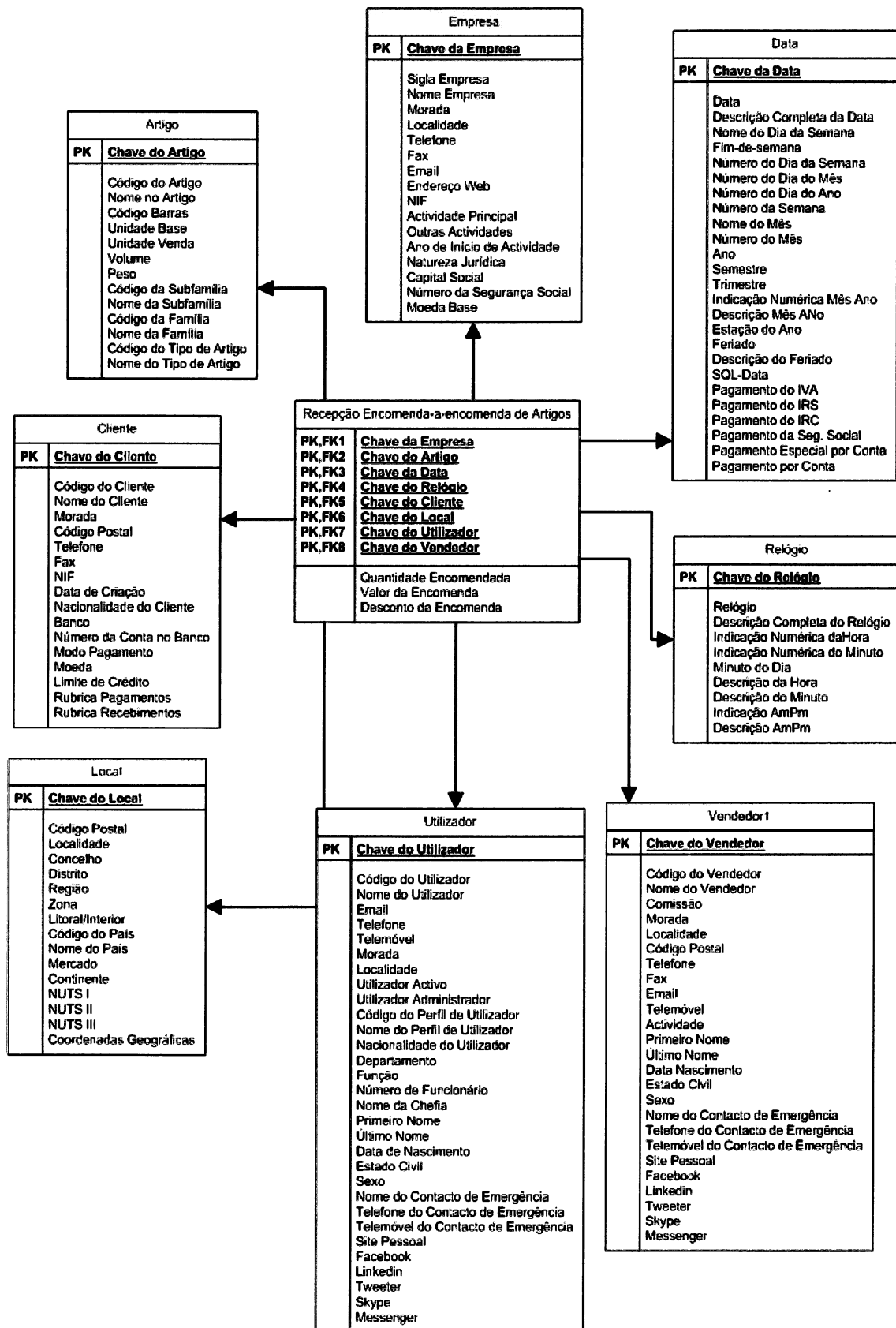
Anexo 1: Esquema em Estrela do Processo Vendas



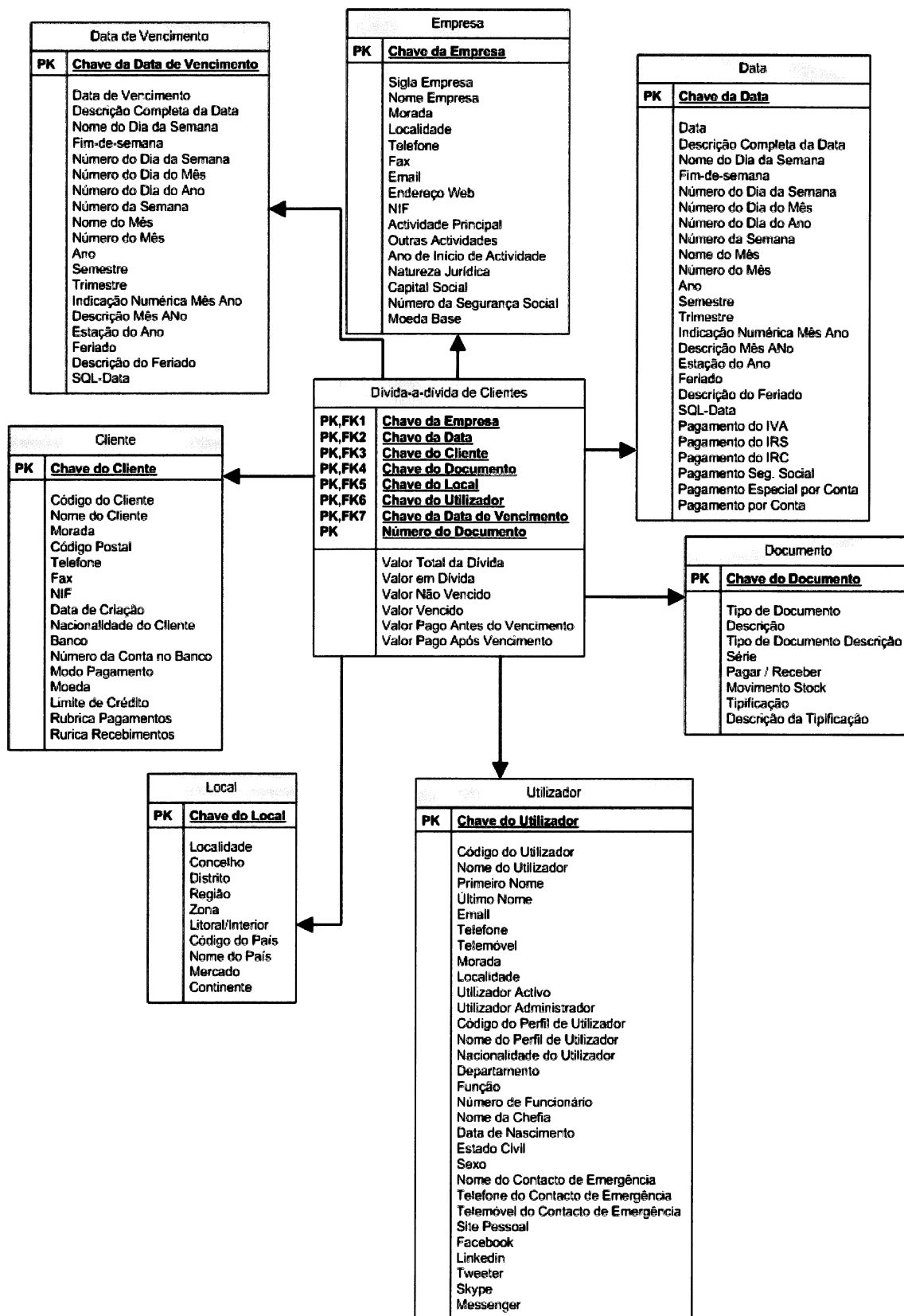
Anexo 2: Esquema em Estrela do Processo Compras



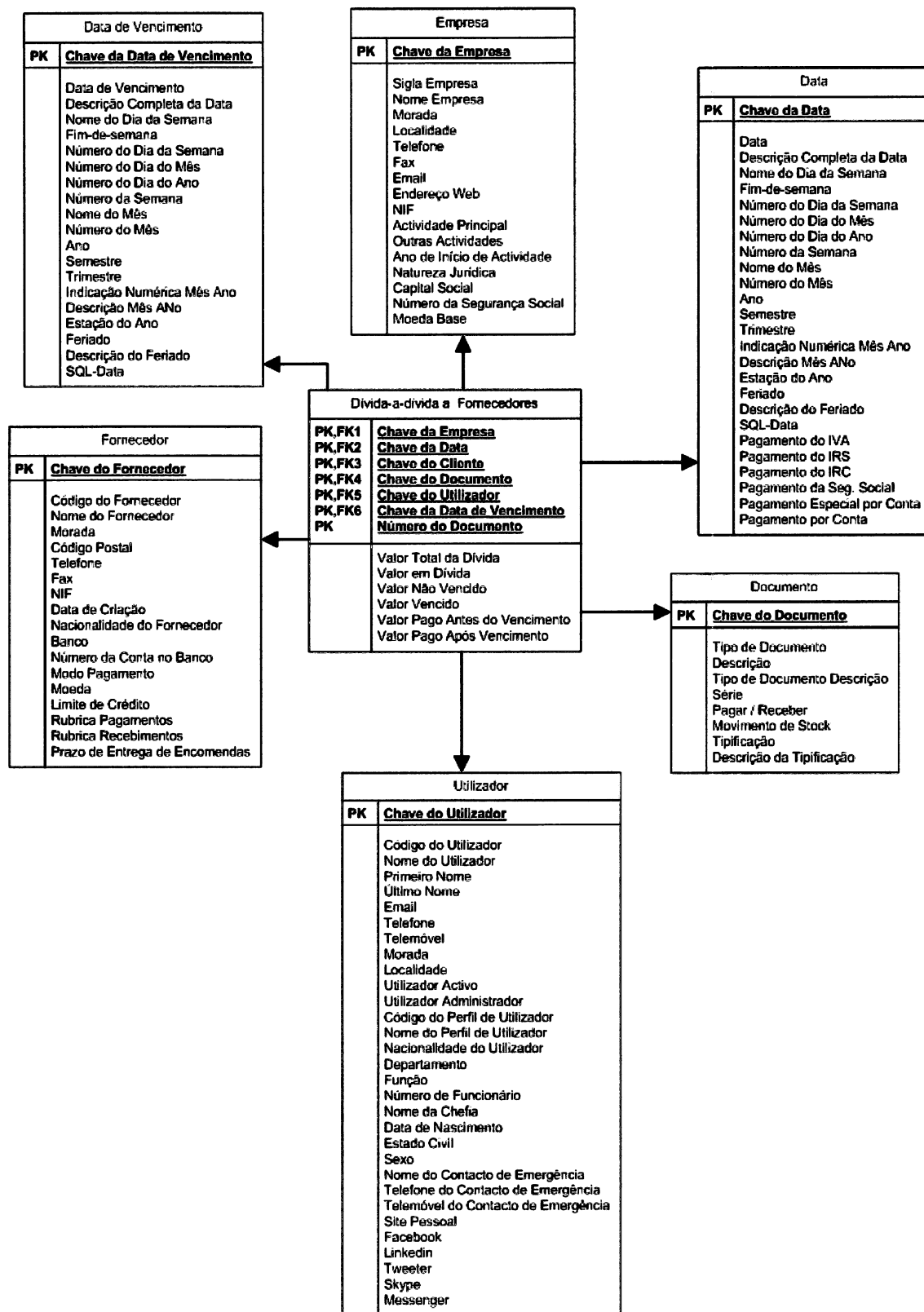
Anexo 3: Esquema em Estrela do Processo Recepção de Encomendas



Anexo 4: Esquema em Estrela do Processo Dívidas de Clientes



Anexo 5: Esquema em Estrela do Processo Dívidas a Fornecedores



Anexo 6: Demonstração de cálculos multidimensionais utilizando o Analysis

Services. Exemplo do cálculo da Margem Bruta por Empresa para o Ano 2009

	Ano ▼		
	2009		
Nome Empresa ▼	Valor da Venda	Custo do Produto	Margem Bruta %
Empresa 1	2.318.566,08 €	1.308.721,96 €	43,55%
Empresa 2	294.928,23 €	156.676,37 €	46,88%
Empresa 3	189.552,60 €	98.510,81 €	48,03%
Grand Total	2.803.046,91 €	1.563.909,14 €	44,21%

Anexo 7: Demonstração de cálculos multidimensionais utilizando o Analysis

Services. Exemplo do cálculo da Margem Bruta por Ano e Mês para o Grupo

Ano ▼	Nome Mes ▼	Valor da Venda	Custo do Produto	Margem Bruta %
2008	Janeiro	227.160,30 €	130.549,10 €	42,53%
	Fevereiro	292.243,78 €	178.139,16 €	39,04%
	Março	287.270,64 €	164.816,28 €	42,63%
	Abril	270.964,88 €	160.700,25 €	40,69%
	Maio	362.388,49 €	229.516,92 €	36,67%
	Junho	288.379,41 €	192.565,67 €	33,22%
	Julho	301.617,79 €	196.332,53 €	34,91%
	Agosto	333.266,36 €	200.138,14 €	39,95%
	Setembro	380.510,50 €	222.832,27 €	41,44%
	Outubro	315.846,61 €	199.598,74 €	36,81%
	Novembro	289.235,93 €	158.452,43 €	45,22%
	Dezembro	268.922,98 €	147.159,25 €	45,28%
	Total	3.617.807,66 €	2.180.800,74 €	39,72%
2009	Janeiro	175.472,19 €	100.099,73 €	42,95%
	Fevereiro	220.429,32 €	123.184,59 €	44,12%
	Março	226.194,91 €	123.453,53 €	45,42%
	Abril	181.275,89 €	106.788,34 €	41,09%
	Maio	198.380,91 €	125.532,36 €	36,72%
	Junho	239.305,81 €	143.088,79 €	40,21%
	Julho	219.805,53 €	125.074,32 €	43,10%
	Agosto	272.437,36 €	151.288,16 €	44,47%
	Setembro	294.068,61 €	160.103,23 €	45,56%
	Outubro	226.979,67 €	108.286,52 €	52,29%
	Novembro	223.999,86 €	124.324,63 €	44,50%
	Dezembro	324.696,84 €	172.684,93 €	46,82%
	Total	2.803.046,91 €	1.563.909,14 €	44,21%
2010		455.395,83 €	254.301,31 €	44,16%
Grand Total		6.876.250,40 €	3.999.011,19 €	41,84%

Anexo 8: Mapa de Vendas

Data.Ano	2009			
	Quantidade Vendida	Valor da Venda	Caixas Vendidas	Margem Bruta %
Azeites	4.688.335	411.697,80 €	91.344	41,54%
Diversos	4.565.417	140.009,34 €	11.093	97,58%
DOP	64.519	125.160,21 €	31.799	16,59%
Extra Virgem	47.784	119.642,73 €	37.982	7,43%
Virgem	10.615	26.885,52 €	10.469	17,60%
Espirituosas	50.832	235.028,64 €	24.047	63,62%
Aguardentes	22.088	67.021,34 €	10.700	59,16%
Amêndoa Amarga	317	2.301,51 €	55	87,99%
Anis	1.608	12.943,16 €	632	72,62%
Vodka	1.398	6.215,93 €	498	65,94%
Whisky	25.421	146.546,72 €	12.163	64,38%
Espumantes	86.427	186.012,50 €	41.753	44,73%
Espumante Branco	36.218	93.748,21 €	16.845	49,22%
Espumante Reserva Branco	35.473	64.561,58 €	17.841	36,47%
Espumante Tinto	14.736	27.702,70 €	7.067	48,78%
Licores	18.951	57.999,66 €	8.044	52,45%
Licor de Amora	1.916	5.659,50 €	703	51,01%
Licor de Café	2.990	10.741,00 €	1.364	55,84%
Licor de Ginja	3.111	8.308,83 €	1.258	43,36%
Licor de Medronho	1.721	5.904,75 €	746	60,97%
Licor de Mirtilo	3.017	10.724,67 €	1.356	47,73%
Licor de Nozes	3.584	7.679,60 €	1.486	56,67%
Licor de Whisky	2.612	8.981,32 €	1.131	54,16%
Merchandising	24.227	37.632,94 €	2.491	62,70%
Merchandising	24.227	37.632,94 €	2.491	62,70%
Não Disponível	-59.821	155,21 €	0	100,00%
Vinagre	4.973	7.433,19 €	3.304	20,51%
Vinagre	4.973	7.433,19 €	3.304	20,51%
Vinhos	1.558.752	1.867.086,98 €	745.415	41,76%
Branco	73.554	47.058,68 €	36.891	22,24%
Diversos	38.082	10.851,35 €	4.225	19,05%
DOC Branco	15.976	25.908,22 €	8.119	47,41%
DOC Tinto	70.288	149.467,25 €	35.180	53,82%
Reserva Branco	156.059	200.392,17 €	78.029	37,20%
Reserva Tinto	623.737	1.054.583,52 €	311.829	48,21%
Rosé	5.479	17.520,06 €	5.341	28,28%
Seleção Branco	129.648	45.783,97 €	43.247	16,35%
Seleção Tinto	83.524	54.556,08 €	41.329	16,04%
Tinto	362.405	260.965,68 €	181.225	26,94%
Total Geral	6.372.676	2.803.046,91 €	916.398	44,21%

Anexo 9: Evolução de Vendas do Vendedor João Nunes

Nome	João Nunes					
	2008	2009		2010		
	Valor da Venda	Caixas Vendidas	Valor da Venda	Caixas Vendidas	Valor da Venda	Caixas Vendidas
Exportação	168.863,93 €	43.233	0,00 €	0	0,00 €	0
Espanha	150.218,81 €	37.155	0,00 €	0	0,00 €	0
França	18.645,12 €	6.078	0,00 €	0	0,00 €	0
Nacional	141.894,17 €	31.190	147.956,02 €	36.249	23.735,62 €	5.879
Portugal	141.894,17 €	31.190	147.956,02 €	36.249	23.735,62 €	5.879
Total Geral	310.758,11 €	74.423	147.956,02 €	36.249	23.735,62 €	5.879

Anexo 10: Mapa de Variação de Vendas em relação ao ano anterior

Rótulos de Linha	2009		2010	
	Variação de Caixas LY	Variação de Vendas LY	Variação de Caixas LY	Variação de Vendas LY
Azeites	3.003,89	24.420,74 €	-13.578,65	-803,11 €
Diversos	2.525,62	28.340,01 €	-5.222,84	24.263,51 €
DOP	-945,50	-5.985,02 €	-1.567,36	-4.408,92 €
Extra Virgem	1.455,58	2.251,16 €	-6.345,41	-19.320,71 €
Virgem	-31,81	-185,40 €	-443,05	-1.336,99 €
Espirituosas	-5.188,42	-42.894,79 €	-154,51	-2.237,18 €
Aguardentes	-818,35	-4.413,65 €	104,40	737,76 €
Amêndoa Amarga	-26,17	-881,30 €	-8,83	-239,26 €
Anis	-187,32	-3.315,22 €	6,54	201,40 €
Vodka	-6,49	-18,59 €	40,32	559,37 €
Whisky	-4.150,09	-34.266,04 €	-296,94	-3.496,45 €
Espumantes	-99,08	-328,30 €	-2.107,09	-8.923,70 €
Espumante Branco	-450,81	-2.112,16 €	-737,44	-3.971,26 €
Espumante Reserva Branco	378,36	1.688,62 €	-869,72	-3.013,46 €
Espumante Tinto	-26,62	95,24 €	-499,93	-1.938,98 €
Licores	-868,95	-5.249,60 €	-678,85	-5.309,08 €
Licor de Amora	-133,77	-1.054,30 €	7,12	47,05 €
Licor de Café	13,93	108,54 €	-184,35	-1.441,75 €
Licor de Ginja	-130,75	-982,20 €	-260,25	-1.768,25 €
Licor de Medronho	-145,13	-1.128,77 €	-4,25	-28,66 €
Licor de Mirtilo	-14,34	-141,62 €	-304,75	-2.401,72 €
Licor de Nozes	-381,38	-1.437,08 €	100,18	541,02 €
Licor de Whisky	-77,51	-614,18 €	-32,56	-256,78 €

Merchandising	-152,84	-1.322,53 €	-55,49	-1.089,07 €
Merchandising	-152,84	-1.322,53 €	-55,49	-1.089,07 €
Vinagre	-887,06	-2.389,55 €	-281,28	-762,71 €
Vinagre	-887,06	-2.389,55 €	-281,28	-762,71 €
Vinhos	-22.740,50	-95.738,53 €	-26.211,36	-107.681,41 €
Branco	1.556,44	1.874,33 €	-342,86	-555,68 €
Diversos	175,34	-2.823,53 €	-338,01	-719,15 €
DOC Branco	906,50	-995,65 €	-441,50	-1.372,75 €
DOC Tinto	-1.740,04	-6.073,43 €	-4.134,50	-17.560,88 €
Reserva Branco	-1.588,84	-4.134,82 €	-668,66	-1.368,21 €
Reserva Tinto	-30.575,32	-94.170,13 €	-29.127,83	-99.524,16 €
Rosé	-330,00	-725,20 €	-65,33	-210,08 €
Seleção Branco	3.303,67	3.469,97 €	181,67	1.135,08 €
Seleção Tinto	1.070,50	1.405,14 €	578,50	763,40 €
Tinto	4.481,25	6.434,78 €	8.147,17	11.731,01 €
Total Geral	-26.932,96	-123.502,57 €	-43.067,23	-126.806,24 €

Anexo 11: Mapa de Compras

Data.Ano

2009

	Valor da Compra			
	Janeiro	Fevereiro	Março	Abril
Administrativo	131,15 €	131,15 €	151,66 €	212,40 €
Seguros	131,15 €	131,15 €	151,66 €	212,40 €
Azeites	17.284,52 €	6.846,79 €	12.585,26 €	4.019,68 €
Diversos	8.482,80 €	641,63 €	3.837,16 €	808,72 €
DOP	2.040,34 €	1.618,34 €	2.864,46 €	2.774,97 €
Extra Virgem	5.908,66 €	4.016,71 €	3.828,70 €	
Virgem	852,72 €	570,10 €	2.054,95 €	435,98 €
Comercial	537,50 €			
Diversos	537,50 €			
Fornecimentos e Serviços Externos	4.920,24 €	3.529,08 €	4.972,22 €	3.659,29 €
Água	8,77 €	12,35 €	19,64 €	9,19 €
Combustível	914,89 €	284,27 €	894,26 €	465,99 €
Conservação e Recuperação	65,24 €	241,53 €	44,85 €	135,39 €
Deslocações	606,16 €	288,84 €	116,38 €	317,04 €
Despesas de Representação	114,00 €			
Diversos TI	482,40 €	498,66 €	465,80 €	532,04 €
Electricidade	657,10 €	599,86 €	613,52 €	18,83 €
Ferramentas e Utensílios	159,54 €	58,96 €	77,72 €	105,34 €
Limpeza	50,21 €	34,86 €	76,58 €	82,30 €

Material de Escritório	82,52 €	14,10 €	33,82 €	78,13 €
Outros FSE	452,35 €	293,42 €	662,04 €	404,30 €
Rendas e Alugueres	747,11 €	747,11 €	1.541,41 €	942,73 €
Transportes de Pessoal	490,58 €	445,98 €	423,68 €	468,28 €
Vigilância	89,36 €	9,13 €	2,53 €	99,74 €
Logística	1.489,55 €	2.139,88 €	1.268,22 €	1.885,08 €
Certificados	10,63 €	84,66 €	45,36 €	66,36 €
Diversos	161,25 €	5,00 €	1,00 €	2,00 €
Transportes	1.317,67 €	2.050,21 €	1.221,86 €	1.816,72 €
Marketing	4,50 €	3.562,05 €	933,95 €	3.672,95 €
Marketing	4,50 €	3.562,05 €	933,95 €	3.672,95 €
Merchandising	25.517,63 €	34.243,08 €	29.422,76 €	39.837,68 €
Merchandising	25.517,63 €	34.243,08 €	29.422,76 €	39.837,68 €
Não Disponível		13,43 €	15,64 €	
Não Disponível		13,43 €	15,64 €	
Outros	6.174,66 €	11.090,64 €	43.024,09 €	13.649,90 €
Diversos	6.174,66 €	11.090,64 €	43.024,09 €	13.649,90 €
Produção	25.318,00 €	42.535,98 €	9.699,50 €	53.887,89 €
Diversos	10.107,88 €	37.713,44 €	8.036,54 €	47.330,55 €
Existências	13.026,28 €	3.204,75 €	60,15 €	4.332,04 €
LAB	30,81 €	16,15 €	38,81 €	85,26 €
Linhas de Montagem	744,47 €	120,83 €	777,82 €	263,09 €
Produtos	468,60 €	555,71 €	17,57 €	388,01 €
Qualidade				
Reparação	939,96 €	925,09 €	768,61 €	1.488,94 €
Produtos Manutenção		18,98 €	1.075,43 €	3.028,37 €
Consumíveis				156,70 €
Equipamento		18,98 €		7,02 €
Outros			1.075,43 €	2.864,65 €
Tecnologias de Informação			150,00 €	
Consumíveis				
Diversos TI			150,00 €	
Hardware				
Serviços de Suporte e Manutenção				
Vinagre	535,92 €			332,64 €
Vinagre	535,92 €			332,64 €
Vinhos	18.928,77 €	23.782,04 €	146,25 €	16.596,90 €
Diversos	18.928,77 €	23.782,04 €	146,25 €	16.596,90 €
Total Geral	100.842,43 €	127.893,09 €	103.444,97 €	140.782,77 €

Anexo 12: Mapa de Vendas vs Compras

Trimestre	2009		
	Valor da Venda	Valor da Compra	Vendas vs Compras
1	622.096,42 €	332.180,49 €	289.915,93 €
Janeiro	175.472,19 €	100.842,43 €	74.629,76 €
Fevereiro	220.429,32 €	127.893,09 €	92.536,23 €
Março	226.194,91 €	103.444,97 €	122.749,94 €
2	618.962,61 €	372.544,88 €	246.417,73 €
Abril	181.275,89 €	140.782,77 €	40.493,12 €
Maio	198.380,91 €	113.639,42 €	84.741,49 €
Junho	239.305,81 €	118.122,69 €	121.183,12 €
3	786.311,50 €	405.558,02 €	380.753,48 €
Julho	219.805,53 €	106.201,93 €	113.603,59 €
Agosto	272.437,36 €	70.700,75 €	201.736,62 €
Setembro	294.068,61 €	228.655,34 €	65.413,27 €
4	775.676,38 €	477.510,20 €	298.166,17 €
Outubro	226.979,67 €	142.491,39 €	84.488,28 €
Novembro	223.999,86 €	169.790,36 €	54.209,51 €
Dezembro	324.696,84 €	165.228,45 €	159.468,39 €
Total Geral	2.803.046,91 €	1.587.793,59 €	1.215.253,32 €

Anexo 13: Mapa de Encomendas e Variação de Encomendas face ao ano anterior

Data.Ano	2009	
	Valor da Encomenda	Variação de Encomendas LY
Empresa 1	2.312.913,66 €	51.486,25 €
Exportação	826.824,34 €	-75.833,64 €
Azeites	63.997,47 €	1.407,53 €
Espirituosas	159.177,85 €	-44.326,58 €
Espumantes	25.064,53 €	-1.425,31 €
Licores	25.539,89 €	-310,20 €
Merchandising	5.012,78 €	3.220,08 €
Vinagre	494,36 €	-174,16 €
Vinhos	547.537,45 €	-34.224,99 €
Nacional	1.486.089,31 €	127.319,89 €
Azeites	1.377,92 €	-781,80 €
Espirituosas	176.665,37 €	4.461,49 €
Espumantes	119.700,99 €	20.486,52 €
Licores	40.123,84 €	-3.362,33 €
Merchandising	1.371,26 €	119,40 €

Vinagre	10,79 €	-137,82 €
Vinhos	1.146.839,15 €	106.534,43 €
Empresa 2	196.756,81 €	-32.695,06 €
Nacional	196.756,81 €	-32.695,06 €
Azeites	186.156,71 €	-29.306,07 €
Merchandising	7.101,23 €	-716,55 €
Vinagre	3.498,87 €	-2.672,43 €
Empresa 3	56.407,69 €	56.407,69 €
Exportação	13.113,13 €	13.113,13 €
Azeites	1.237,85 €	1.237,85 €
Espirituosas	2.039,08 €	2.039,08 €
Espumantes	219,35 €	219,35 €
Licores	1.337,46 €	1.337,46 €
Merchandising	357,90 €	357,90 €
Vinagre	45,05 €	45,05 €
Vinhos	7.876,44 €	7.876,44 €
Nacional	43.294,56 €	43.294,56 €
Azeites	21,02 €	21,02 €
Espirituosas	227,13 €	227,13 €
Espumantes	482,52 €	482,52 €
Licores	1.230,43 €	1.230,43 €
Merchandising	0,45 €	0,45 €
Vinhos	41.333,01 €	41.333,01 €
Total Geral	2.566.078,15 €	75.198,88 €

Anexo 14: Mapa de Vendas vs Encomendas

Vendedor	Valor da Venda			Valor da Encomenda		
	2008	2009	2010	2008	2009	2010
Ana Andrade	128.768,51 €	7.211,47 €	226,39 €	157.655,71 €	239.103,31 €	28.004,49 €
Andreia Silva	266.950,00 €	7.257,03 €	-851,23 €	238.417,90 €	293.187,37 €	59.150,67 €
António Manuel	281.721,54 €	-6.390,23 €	-27,04 €	306.283,82 €	204.342,71 €	23.308,94 €
Fábio Faría	330,03 €	0,00 €	0,00 €	208,44 €	0,00 €	0,00 €
João Manuel	440,14 €	106,67 €	19,11 €	288,30 €	6,24 €	0,00 €
João Nunes	310.758,11 €	147.956,02 €	23.735,62 €	108.262,48 €	76.934,86 €	5.693,62 €
Loja	677,62 €	341,15 €	60,70 €	102,68 €	124,59 €	0,00 €
Loja 1	6.509,11 €	168,00 €	6,44 €	9.234,88 €	47,95 €	0,00 €
Miguel Vítor	38,98 €	28,68 €	34,17 €	30,85 €	0,00 €	0,00 €
Não Disponível	2.501.160,48 €	2.650.375,16 €	421.478,26 €	1.551.015,08 €	1.700.430,34 €	218.912,91 €
Nuno Ribeiro	120.453,13 €	-4.007,03 €	175,50 €	119.379,12 €	51.900,77 €	0,00 €
Total Geral	3.617.807,66 €	2.803.046,91 €	444.857,91 €	2.490.879,27 €	2.566.078,15 €	335.070,62 €

Anexo 15: Mapa de Dívidas de Clientes

Data.Ano 2009

	Valor Total da Dívida de Clientes	Valor Em Dívida de Clientes
Empresa 1	49.837,24 €	48.059,13 €
Julieta	45.897,60 €	44.119,49 €
Maria Miguel	3.939,64 €	3.939,64 €
Empresa 2	72.447,16 €	27.026,84 €
Arminda	58.470,40 €	13.050,08 €
Lúcio	13.976,76 €	13.976,76 €
Empresa 3	20.630,70 €	20.512,83 €
Barbara Gabriela	17.673,57 €	17.673,57 €
Beatriz	1.602,98 €	1.602,98 €
Mauro	240,45 €	240,45 €
Nicolau	1.113,71 €	995,83 €
Empresa 4	45.720,00 €	45.720,00 €
Valmor	45.720,00 €	45.720,00 €
Total Geral	188.635,10 €	141.318,80 €

Anexo 16: Mapa de Dívidas a Fornecedores

Data.Ano 2010

Data.Nome Mes Março

Data.Numero Dia Mes (Itens múltiplos)

	Valor Total da Dívida a Fornecedores	Valor Em Dívida a Fornecedores
Empresa 1	-51.992,02 €	-51.992,02 €
Amílcar	-2.376,00 €	-2.376,00 €
Gabriel Victor	-1.393,20 €	-1.393,20 €
Hernâni	-6.569,19 €	-6.569,19 €
Iva	-3.397,05 €	-3.397,05 €
Jacinta	-5.112,00 €	-5.112,00 €
José João	-1.066,93 €	-1.066,93 €
Leonardo	-1.205,36 €	-1.205,36 €
Romilda	-2.083,33 €	-2.083,33 €
Rosa	-10.649,85 €	-10.649,85 €
Sandoval	-14.758,05 €	-14.758,05 €
Sidnei	-2.151,06 €	-2.151,06 €
Talassa	-1.230,00 €	-1.230,00 €
Empresa 2	-23.689,27 €	-23.689,27 €
Esmeraldo	-1.256,64 €	-1.256,64 €
Iva	-1.257,49 €	-1.257,49 €

Liana	-2.298,00 €	-2.298,00 €
Maria Raquel	-2.280,00 €	-2.280,00 €
Marina	-2.006,66 €	-2.006,66 €
Odin	-6.238,01 €	-6.238,01 €
Said	-1.215,00 €	-1.215,00 €
Sidnei	-7.137,47 €	-7.137,47 €
Empresa 3	-3.990,00 €	-3.990,00 €
Oliver	-3.990,00 €	-3.990,00 €
Total Geral	-79.671,29 €	-79.671,29 €

Anexo 17: Mapa de Dívidas a Fornecedores vs Dívidas de Clientes

	Valor Total da Dívida a Fornecedores	Valor Total da Dívida de Clientes
Empresa 1	-227.844,55 €	118.195,16 €
2008	-15.370,36 €	66.646,46 €
2009	-24.498,64 €	49.383,47 €
2010	-187.975,56 €	2.165,23 €
Empresa 2	-33.404,08 €	192.221,87 €
2008	0,00 €	13.249,33 €
2009	1.014,10 €	72.530,31 €
2010	-34.418,18 €	106.442,23 €
Empresa 3	-13.397,98 €	59.339,58 €
2008	0,00 €	1.225,07 €
2009	1.681,92 €	19.466,37 €
2010	-15.079,90 €	38.648,14 €
Empresa 4	0,00 €	48.480,00 €
2008	0,00 €	0,00 €
2009	0,00 €	0,00 €
2010	0,00 €	2.760,00 €
Total Geral	-274.646,61 €	418.236,61 €

Referências

- Becker, B. 2003. Kimball Design Tip #46: Another Look At Degenerate Dimensions. [Acedido: 8 de Julho de 2010]. Disponível em: <http://www.rkimball.com/html/designtipsPDF/DesignTips2003/KimballDT46AnotherLook.pdf>.
- Caldeira, C. 2008. *Data Warehousing: Conceitos e Modelos*. Edições Sílabo, Lisboa.
- Clemes, M. 2001. *Data Warehouse como Suporte ao Sistema de Informações Gerenciais em uma Instituição de Ensino Superior: Estudo de Caso na UFSC*. Florianópolis.
- Comes, G. 2001. *Contribuição ao Estudo da Implementação de Data Warehousing: Um Caso no Setor de Telecomunicações*. São Paulo.
- Inmon, W. 2002. *Building the Data Warehouse (Third Edition)*. John Wiley & Sons, New York
- Jacobson, R. e Misner, S. 2006. *Microsoft SQL Server 2005 Analysis Services – Step by Step*. Microsoft Press.
- Kelley, C., Oates, J., Rehm, C. 2003. What is a degenerate dimension and how is it used. [Acedido: 8 de Julho de 2010]. Disponível em: <http://www.information-management.com/news/7844-1.html>.
- Kimball, R. 1999. A Practical Method for Planning a Data Warehouse – The Matrix. [Acedido: 3 de Julho de 2010]. Disponível em: http://intelligent-enterprise.informationweek.com/db_area/archives/1999/990712/webhouse.jhtml.

Kimball, R. 2008. *Fact Tables – Dimensional Perspectives*. [Acedido: 4 de Julho de 2010]. Disponível em: http://www.information-management.com/issues/2007_54/10002185-1.html.

Kimball, R. e Caserta, J. 2004. *The Data Warehouse ETL Toolkit – Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. John Wiley & Sons, Indianapolis.

Kimball, R. e Ross, M. 2002. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling* (Second Edition). John Wiley & Sons, New York.

Ross, M. 2002. *Kimball Design Tip #41: Drill Down Into A More Detailed Bus Matrix*. [Acedido: 3 de Julho de 2010]. Disponível em: <http://www.rkimball.com/html/designtipsPDF/DesignTips2002/KimballDT41DrillDown.pdf>.

Ross, M. 2005. *The Matrix: Revisited*. [Acedido: 3 de Julho de 2010]. Disponível em: <http://intelligent-enterprise.informationweek.com/showArticle.jhtml?articleID=174300383>.

Ross, M. 2009. *Kimball University: The 10 Essential Rules of Dimensional Modeling*. [Acedido: 4 de Julho de 2010]. Disponível em: <http://intelligent-enterprise.informationweek.com/showArticle.jhtml;jsessionid=42NILWYWGUGBQE1GHPCKH4ATMY32JVN?articleID=217700810>.

Sá, Jorge. 2009. *Metodologia de Sistemas de Data Warehouse*. Tese de Doutoramento, Universidade do Minho.

Santos, F. 2010. Modelo Dimensional para Data Warehousing. [Acedido: 4 de Julho de 2010]. Disponível em: <http://fmsantos.infotuga.com/?p=219#more-219>.

Silvers, F. 2008. *Building and Maintaining a Data Warehouse*. Auerbach Publications, Florida.

Sousa, R. 2010. Bus Matrix, uma breve explicação. [Acedido: 3 de Julho de 2010]. Disponível em: <http://www.infotuga.com/?p=217>.



