

**Implementação de um Modelo de *Data*  
*Warehouse* para o Serviço Nacional de Avisos  
Agrícolas**

por

Miguel Marques Nunes Soares Paulo

Trabalho de Projecto apresentado como requisito parcial para obtenção do grau de

Mestre em Estatística e Gestão de Informação

Especialização em Business Intelligence e Gestão do Conhecimento

pelo

**Instituto Superior de Estatística e Gestão de Informação**

**da**

**Universidade Nova de Lisboa**

**Implementação de um Modelo de Data  
Warehouse para o Serviço Nacional de Avisos  
Agrícolas**

Professor orientador:

Professor Doutor Miguel de Castro Neto

**Novembro, 2011**

## RESUMO

O presente relatório de projecto tem como objectivo demonstrar uma aplicação prática da implementação de um novo sistema tecnológico de gestão de dados numa divisão da DGADR do Ministério da Agricultura - o Serviço Nacional de Avisos Agrícolas (SNAA). Este sistema de gestão de base de dados é denominado por Data Warehouse, um formato de base de dados que organiza os dados dispostos de uma forma concebida para otimizar a sua análise.

O SNAA assume um papel de relevo no contexto agrícola português pois fornece informações sob a forma de avisos, obtidas por previsão e observações periódicas, tendo em vista identificar a melhor oportunidade de tratamento fitossanitário dos meios agrícolas. Estes avisos indicam os produtos fitofarmacêuticos mais adequados a cada situação e conduzem, em regra, a uma diminuição do número de tratamentos fitossanitários.

As inúmeras vantagens decorrentes da aplicação de uma tecnologia de Business Intelligence, associadas ao dinamismo na apresentação da informação, constituem as motivações para a elaboração deste relatório como parte integrante do Mestrado em Estatística e Gestão de Informação, com a vertente de especialização em Business Intelligence e Gestão do Conhecimento.

O presente projecto assenta no princípio que a implementação dum DW sobre a base de dados relacional existente, permitirá um aumento no dinamismo e na forma como se lida com a informação recolhida, tornando-se possível obter diversas vantagens competitivas. Há, assim, que destacar algumas das melhorias que serão obtidas deste projecto:

- i.** Melhoria significativa na consulta e disseminação da informação divulgada pelo SNAA;
- ii.** Melhoria da qualidade dos dados;
- iii.** Potencialidade de criação de sistemas de reporte de informação, que funcionem como sistemas de apoio à decisão aos utilizadores do portal do SNAA;

## **PALAVRAS-CHAVE**

Business Intelligence; Data Warehouse; Serviço Nacional de Avisos Agrícolas; ETL;

## ABSTRACT

The present Project report has as main goal to present a practical application of the implementation of a new database management system in a DGADR division of the Ministry of Agriculture – the Portuguese Agricultural Warning Services (SNAA). This database management system is known as Data Warehouse and has a format that enables to organize data in a way that optimizes analysis.

SNAA has an important role in the context of the Portuguese Agriculture, delivering information in the form of warnings, thus obtained by prevision and periodical observation, it as the purpose of identifying the most effective phyto-sanitary treatment of the agriculture. These warnings point the most suitable phyto-pharmaceutical products to each circumstance and those lead, in general, to a decrease in the number of phyto-sanitary treatments.

The several advantages of the application of Business Intelligence technology related with the dynamism in information presentation constitute the reasons for the elaboration of this report as a part of the Msc in Statistics and Information Management, with the specialization in Business Intelligence and Knowledge Management.

The project settles in the principle that the implementation of a DW, deriving from the present relational database, will allow an increase on the dynamism and in the ways that the retrieved information is managed, thus enabling the possibility to obtain numerous competitive advantages. With this Project the following improvements will be obtained:

- i.** Significant improvement in the search and dissemination of the information released by the SNAA;
- ii.** Improvement on Data Quality;
- iii.** Potential creation of reporting information systems that will act as decision support systems to the users of the SNAA Portal;

## KEYWORDS

Business Intelligence; Data Warehouse; Agricultural Warning Services; ETL;

## ÍNDICE

<b>ÍNDICE DE TABELAS .....</b>	<b>VIII</b>
<b>ÍNDICE DE FIGURAS .....</b>	<b>IX</b>
<b>SIGLAS E ABREVIATURAS .....</b>	<b>X</b>
<b>AGRADECIMENTOS .....</b>	<b>XI</b>
<b>I - INTRODUÇÃO .....</b>	<b>1</b>
I.1. ESTRUTURA DO PROJECTO.....	3
I.2. OBJECTIVO DO PROJECTO .....	4
I.3. VALOR DO PROJECTO .....	5
I.4. CASOS DE SUCESSO ANTERIORES .....	7
I.5. ENQUADRAMENTO DE NEGÓCIO.....	8
<i>I.5.1. Serviço Nacional de Avisos Agrícolas.....</i>	<i>8</i>
<i>I.5.2. Medidas Agro-Ambientais .....</i>	<i>11</i>
<i>I.5.3. Contexto Europeu e Mundial.....</i>	<i>13</i>
<b>II - CONTEXTO.....</b>	<b>15</b>
II.1. BUSINESS INTELLIGENCE.....	15
II.2. DATA WAREHOUSE .....	16
<i>II.2.1. Normalização.....</i>	<i>17</i>
<i>II.2.2. Modelo Dimensional.....</i>	<i>17</i>
<i>II.2.3. Flat Schema .....</i>	<i>18</i>
<i>II.2.4. Terraced Schema .....</i>	<i>20</i>
<i>II.2.5. Star Schema .....</i>	<i>21</i>
<i>II.2.6. Snowflake Schema .....</i>	<i>22</i>
<i>II.2.7. Metadados.....</i>	<i>23</i>
II.3. ETL.....	24
<b>III -EXPLICAÇÃO DOS PROCEDIMENTOS.....</b>	<b>26</b>
III.1. ESTRUTURA DAS BASES DE DADOS .....	27
<i>III.1.1. Requisitos de Negócio.....</i>	<i>27</i>
<i>III.1.2. Granularidade .....</i>	<i>28</i>
<i>III.1.3. Classificação das Entidades .....</i>	<i>28</i>
<i>III.1.4. Identificação das Hierarquias .....</i>	<i>30</i>
<i>III.1.5. Escolha do Modelo Multidimensional .....</i>	<i>31</i>

III.2. STAGING AREA .....	32
III.3. DATA WAREHOUSE .....	38
III.4. EXTRAÇÃO, TRANSFORMAÇÃO E CARREGAMENTO .....	38
III.5. DIMENSÕES E FACTOS .....	42
III.6. MÉTRICAS .....	46
III.7. ANÁLISE DE RESULTADOS .....	48
<b>IV - CONCLUSÕES .....</b>	<b>50</b>
IV.1. LIMITAÇÕES .....	52
IV.2. PROPOSTAS DE TRABALHO FUTURO .....	53
<b>V - REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>56</b>
<b>VI - ANEXOS .....</b>	<b>62</b>
VI.1. ANEXO A - CRONOGRAMA .....	62
VI.2. ANEXO B - DIAGRAMAS .....	64

## ÍNDICE DE TABELAS

Tabela 1 – Definição dos limites no controlo de qualidade dos dados. ....	37
Tabela 2 – Tipologia das Direcções Regionais de Agricultura e Pescas. ....	43
Tabela 3 – Tipologia das Estações de Avisos. ....	43
Tabela 4 – Tipologia das Estações Meteorológicas ....	43
Tabela 5 – Métricas do <i>Data Warehouse</i> . ....	47
Tabela 6 – Modelos a implementar no <i>Data Warehouse</i> . ....	47

## ÍNDICE DE FIGURAS

Figura 1 – Organograma de precedências hierárquicas do SNAA. ....	9
Figura 2 – Localização geográfica das estações de Avisos Públicas e Privadas sobre a alçada do SNAA em Portugal Continental .....	10
Figura 3 – Componentes de um ambiente de <i>Business Intelligence</i> . ....	16
Figura 4 – Exemplo da Metodologia dimensional <i>Flat Schema</i> .....	19
Figura 5 – Exemplo da Metodologia dimensional <i>Terraced Schema</i> .....	20
Figura 6 – Exemplo da Metodologia dimensional <i>Star Schema</i> .....	21
Figura 7 – Exemplo da Metodologia dimensional <i>Snowflake Schema</i> .....	22
Figura 8 – O Meta Modelo <i>Common Warehouse Metamodel</i> . ....	24
Figura 9 – Sistema de <i>Business Intelligence</i> . ....	25
Figura 10 – Exemplo de uma hierarquia. ....	31
Figura 11 – Características dos diferentes modelos dimensionais.....	32
Figura 12 – Diagrama inicial das tabelas da <i>Staging Area</i> . ....	33
Figura 13 – Diagrama da interactividade entre os <i>Data Transformation Services 2 e 3</i> .39	
Figura 14 – Diagrama lógico do <i>Data Transformation Services</i> inicial. ....	39
Figura 15 – Modulo <i>Data Flow Task</i> e ferramentas <i>Conditional Split</i> e <i>Lookup</i> .....	40
Figura 16 – Processo ETL da SA. ....	41
Figura 17 – Processo ETL do DW. ....	42
Figura 18 – Diagrama hierárquico Dimensão Território. ....	44
Figura 19 – Mapa de planeamento de actividades gerais do projecto. ....	62
Figura 20 – Relatório do Microsoft Office Project – Mapa de Gantt. ....	63
Figura 21 – Diagrama da <i>Staging Area</i> . ....	64
Figura 22 – Tabela do <i>Data Warehouse</i> . ....	65
Figura 23 – Tabela da dimensão território do <i>Data Warehouse</i> (DimTerritorio). ....	66
Figura 24 – Tabela da dimensão empregados do <i>Data Warehouse</i> (DimEmpregados). 67	
Figura 25 – Tabela da dimensão Tempo do <i>Data Warehouse</i> (DimTime). ....	68
Figura 26 – Tabela de factos do <i>Data Warehouse</i> (FactDados). ....	69

## **SIGLAS E ABREVIATURAS**

**SNAA** – Serviço Nacional de Avisos Agrícolas

**ETL** – Extração, transformação e carregamento (adaptado de Extract, transform and Load)

**BI** – Business Intelligence

**DW** – Data Warehouse

**SA** – Staging Area

**DTS** - Data Transformation Services

**SQL** - Structure Query Language

**OLTP** – Online transaction Processing

**OLAP** – Online Analytical Processing

**DTSX** – Data Transformation Services para a versão 2005 do Software SQL Server

## AGRADECIMENTOS

**“That which does not kill us,  
only makes us stronger.”**

*Friedrich Nietzsche*

Para que esta tese de mestrado fosse possível desejo expressar os meus sinceros agradecimentos a um conjunto de pessoas que da mais variada forma influenciaram o seu resultado final.

Aos meus pais, Jorge e Noémia que, apesar de todas as adversidades, permitiram que tivesse a oportunidade de avançar nos meus estudos e que tivesse a capacidade e as bases para tornar possível a realização dos meus sonhos. A eles agradeço todo o esforço e dedicação e por sempre acreditarem em mim.

Ao Professor Doutor Miguel Neto, orientador da dissertação, que através das suas críticas construtivas me motivou sempre a dar mais e melhor. Agradeço o facto de ter acreditado em mim e ter tido confiança nas minhas capacidades.

Ao ISEGI, por me ter proporcionado o apoio necessário para proceder com os meus estudos e que me deixou o gosto pela vida académica e pela investigação.

Aos meus familiares e amigos que através das suas oportunas observações me permitiram focar em pormenores que muitas vezes me passaram despercebidos. Obrigado por estarem presentes e por todo o empenho demonstrado.

## I - INTRODUÇÃO

“Good is the enemy of great.”

*Jim Collins*

As tecnologias de informação (TI) estão a alterar a forma como as empresas actuam e estão igualmente a afectar todo o processo produtivo desde o produto ao serviço final. (Porter & Millar, 2001). As TI, permitem assim gerar mais dados consoante a actividade da empresa, possibilitando às mesmas recolher mais dados e capturar mais informação. No entanto, tal como Gardner (1998) afirma, hoje em dia as empresas são cada vez mais inundadas com dados mas conseguem reter e utilizar pouca informação. Para tal, existem ferramentas de *Business Intelligence* (BI), tais como *Data Warehouses* (DW), que permitem colmatar esta lacuna.

Num ambiente em permanente mudança e competitividade, a agricultura tem vindo a assistir a uma evolução acelerada das tecnologias disponíveis para a sua prática (Neto & Fernandes, 2007). Neste contexto, Wang, Zhang e Wang (2006) referem que a evolução que se tem assistido no campo das tecnologias de informação e comunicação, em particular na área da computação móvel e da monitorização remota, tem disponibilizado no mercado capacidades computacionais crescentes. Adicionalmente, afirma que estas têm colocado desafios extremamente interessantes à economia em geral e ao sector agrícola em particular. De acordo com Abdullah, Brobst, Umer e Khan (2004), existem complexas relações entre sistemas informáticos e diversos factores físicos, químicos e biológicos da agricultura que influenciam os mesmos e que necessitam de ser compreendidos. Para compreender estas ligações, são necessários modelos matemáticos e estatísticos aprofundados e com elevada ênfase analítica. A título de exemplo, Wang et al. (2006), enuncia cinco tipos de aplicação de sistemas de Informação (SI) na agricultura, tais como: a) monitorização ambiental; b) agricultura de precisão; c) comunicação integrada entre sistemas; d) automação de sistemas; e) sistemas de controlo e monitorização de alimentos para a produção alimentar.

O presente projecto visa assim envolver estas duas ciências comumente apresentadas em separado, mas com uma tendência cada vez maior de interligação, no aproveitamento dos recursos existentes desde o processo de cultivo até ao combate aos microrganismos que dificultam o mesmo (SNAA, 2010). Neste contexto, surge o

Serviço Nacional de Avisos Agrícolas (SNAA), que tem por objectivo fornecer informações por previsão e por observações periódicas com vista à melhor oportunidade de tratamentos fitossanitários. Adicionalmente, indica os produtos fitofarmacêuticos adequados a cada situação, o que conduz, em regra, a uma diminuição do número de tratamentos. O SNAA, é assim, dotado tecnologicamente de uma infraestrutura composta por estações de avisos que recolhem e armazenam informação meteorológica.

Através da utilização dos recursos fornecidos pelo SNAA, mediante a pessoa do Doutor Miguel de Castro Neto, orientador desta tese de mestrado, pretende-se uniformizar a recolha, promover a integração e implementar novas formas de tratamento dos dados meteorológicos através de um DW.

## **I.1. Estrutura do Projecto**

**Capítulo I** – Capítulo introdutório onde é dado um enquadramento do negócio do cliente, o Sistema Nacional de Avisos Agrícolas, da sua importância a nível nacional, europeu e mundial, assim como os seus objectivos. Adicionalmente, são definidos os objectivos do projecto e o seu valor.

**Capítulo II** – Apresenta os conceitos teóricos e as bases de conhecimento que foram utilizadas neste projecto.

**Capítulo III** - Neste capítulo é explicitada a estrutura das diferentes bases de dados envolvidas no presente projecto e a metodologia de construção da *Staging Area* (SA) e do DW. É também descrito o processo de extração, transformação e carregamento (ETL) implementado, tanto de passagem da base de dados relacional para a SA como da SA para o DW. Este é o capítulo que contém o modelo de projecto que foi implementado e os procedimentos que foram seguidos.

**Capítulo IV** – Expõe as conclusões retiradas da implementação do projecto, as limitações encontradas e propostas de trabalho futuros.

## **I.2. Objectivo do Projecto**

O projecto apresentado tem dois objectivos, um prático e um teórico. O objectivo prático corresponde à implementação de um modelo de *Data Warehousing* para dados meteorológicos, num sistema informático de recolha de informação pouco desenvolvido até ao momento. Como forma de atingir este objectivo, foi construído um modelo de ETL, que permitirá um uso mais eficiente do sistema de gestão de bases de dados, optimizando o processo de consulta dos dados. Todos os mecanismos tecnológicos implementados serão sustentados por bases teóricas e práticas, nomeadamente de casos de sucesso anteriormente desenvolvidos. Por sua vez, o objectivo teórico é o de promover a explicação de todos os benefícios inerentes a um melhor aproveitamento da informação meteorológica disponibilizada pelo SNAA, no âmbito técnico, económico e social. Os utilizadores finais a beneficiar desta implementação são todos os técnicos das estações de avisos agrícolas, as estações de avisos agrícolas, os assinantes das estações e o público em geral.

### I.3. Valor do Projecto

A agricultura é altamente vulnerável às alterações climáticas (Jones, Hansen, Royce & Messina ,2000). Como tal, existe um interesse elevado no aproveitamento da informação meteorológica na agricultura. Tal como Weis, Crowder e Bernardi (1999) afirmam, a informação agrometeorológica é um processo contínuo, que é iniciado com o conhecimento científico e que termina com a análise da informação, destacando que nas fases intermédias existe todo um processo de recolha e tratamento dos dados, assim como a sua posterior disseminação. Igualmente, Pinet et al., (2010) destaca que um DW na agricultura permite analisar e integrar os dados agrícolas produzidos por diferentes fontes, destacando que a informação que é guardada em diferentes bases de dados pode ser agrupada num DW para efectuar análises aprofundadas aos dados. É neste contexto que são utilizados os SI de *Business Intelligence*, nomeadamente os DW's e as ferramentas de *Online Analytical Processing (OLAP)*, tais como *Dashboards*.

Sendo assim, este projecto enquadra-se na junção destas metodologias informáticas numa outra ciência de estudo - a agricultura - permitindo a obtenção de benefícios práticos e construindo a base para o desenvolvimento de outros.

A tecnologia de DW já é amplamente utilizada para diferentes sectores de negócio, no entanto, o conceito de DW para a agricultura é algo pouco elaborado até ao momento (Abdullah & Hussain, 2006). No caso particular de um DW para informação meteorológica e especialmente utilizada para fins agrícolas (agrometeorologia), o número de exemplos é ainda mais reduzido. Haberli e Tombros (2001) afirmam que, excluindo o DW desenvolvido no âmbito do Serviço Nacional de Meteorologia Suíço – MeteoSwiss – não existem outros exemplos de modelos dimensionais utilizados na meteorologia e na climatologia. Esclarece, no entanto, que existem diversos exemplos de modelos relacionais existentes nas organizações meteorológicas dado que a principal ênfase é dada ao armazenamento da informação recolhida e ao seu tratamento, mas não à sua disseminação. Como tal, o presente projecto apresenta claras mais-valias no que concerne à modelação estatística para a previsão dos melhores tratamentos fitofarmacêuticos a adoptar, fornecendo dados meteorológicos mais fiáveis e robustos.

Especificamente, as suas principais vantagens são:

- i. Maior facilidade na consulta de informação meteorológica disponibilizada pelas Estações Meteorológicas (EM), resultando numa melhoria significativa da disseminação da informação divulgada pelo SNAA;
- ii. Construção de uma base mais fiel de informação, mediante o tratamento dos dados com base nos processos mais recentes utilizados no controlo de qualidade de dados meteorológicos;
- iii. Potencialidade de criação de sistemas de reporte de informação mais robustos e interactivos, que funcionem como sistemas de apoio à decisão aos utilizadores do SNAA e na protecção das culturas agrícolas;

O primeiro ponto resulta de um benefício directo da implementação de um DW, ou seja, os utilizadores finais têm a possibilidade de realizar diferentes análises e de uma forma mais eficiente, consumindo menos recursos dos servidores e obtendo um menor tempo de resposta em cada consulta, tendo assim uma maior base de apoio nas suas tomadas de decisão.

O segundo ponto retrata o principal objectivo da produção de informação, a sua posterior difusão aos utilizadores. Esta informação pode, em muitos casos, ser utilizada como factor de melhoria de um conjunto de aspectos e de necessidades que, ou já estavam previstas e eram realizadas com maior esforço, ou nem eram previstas e das quais é possível retirar proveitos. Como caso concreto são de realçar todas as aplicações que influenciam ou influenciaram a agricultura, tendo por base a meteorologia. No âmbito do SNAA, as suas potencialidades prendem-se com o aviso atempado de relatórios fitossanitários e com a prospecção e acompanhamento da evolução de novos inimigos das culturas.

O último ponto evidencia a colocação de plataformas tecnológicas e bases de cálculo informatizadas como suporte a uma forma de previsão atempada para os utilizadores do SNAA. Além disto, outro factor a realçar é a difusão destes avisos mediante um mecanismo rápido e de interacção dinâmica, que permita aumentar o conhecimento informático dos utilizadores excluídos pelo *modus operandi* da tecnologia.

#### I.4. Casos de sucesso anteriores

A referência mais antiga à implementação de um DW na agricultura é dada ao projecto desenvolvido nos EUA, através de uma parceria do Departamento de Agricultura e do Serviço Nacional de Estatísticas Agrícolas (USDA-NASS) em 1997. Este teve como objectivo integrar e padronizar os dados recolhidos pelo NASS que continham dados estatísticos de produção, economia e população agrícola (Yost & Nealon, 1999). Apesar de ter sido, possivelmente, o projecto de DW agrícola mais abrangente, existem ainda outros exemplos. O projecto de DW desenvolvido pelo Instituto de Nacional de Estatística Italiano (ISTAT) - SISSI - que inclui dados de empresas e quintas agrícolas, agregando dados económicos sobre os mesmos num contexto multidimensional (Martino & Santis, 2001) é um outro caso de sucesso. Existe ainda um projecto de DW – *Sydsvensk Radgivning* (SyR) - desenvolvido para os produtores de leite na Suécia (Swensson & Sederblad, 1997).

Mais recentemente, um exemplo de DW para a agricultura, realizado a nível governamental, é o projecto *New Pilot Agriculture Extention Data Warehouse* (PAE DWH) desenvolvido pelo governo do Paquistão, que contém dados relativos a pragas, pesticidas e dados meteorológicos. O PAE DWH teve como principais objectivos o auxílio nos processos de tomada de decisão diários, no que respeita às pragas agrícolas e uma intervenção relevante no planeamento estratégico nacional através da análise de dados históricos (Abdullah & Hussain, 2006). Por último, foi implementado um DW para o sector agrícola do governo Indiano, o *Integrated National Agricultural Resources Information Systems* (INARIS) (Nilakanta, Scheibe & Rai, 2008). O INARIS teve como objectivo agrupar informação dos diferentes distritos indianos referente à produção agrícola, dados meteorológicos, preços de mercado e à dimensão populacional.

## **I.5. Enquadramento de Negócio**

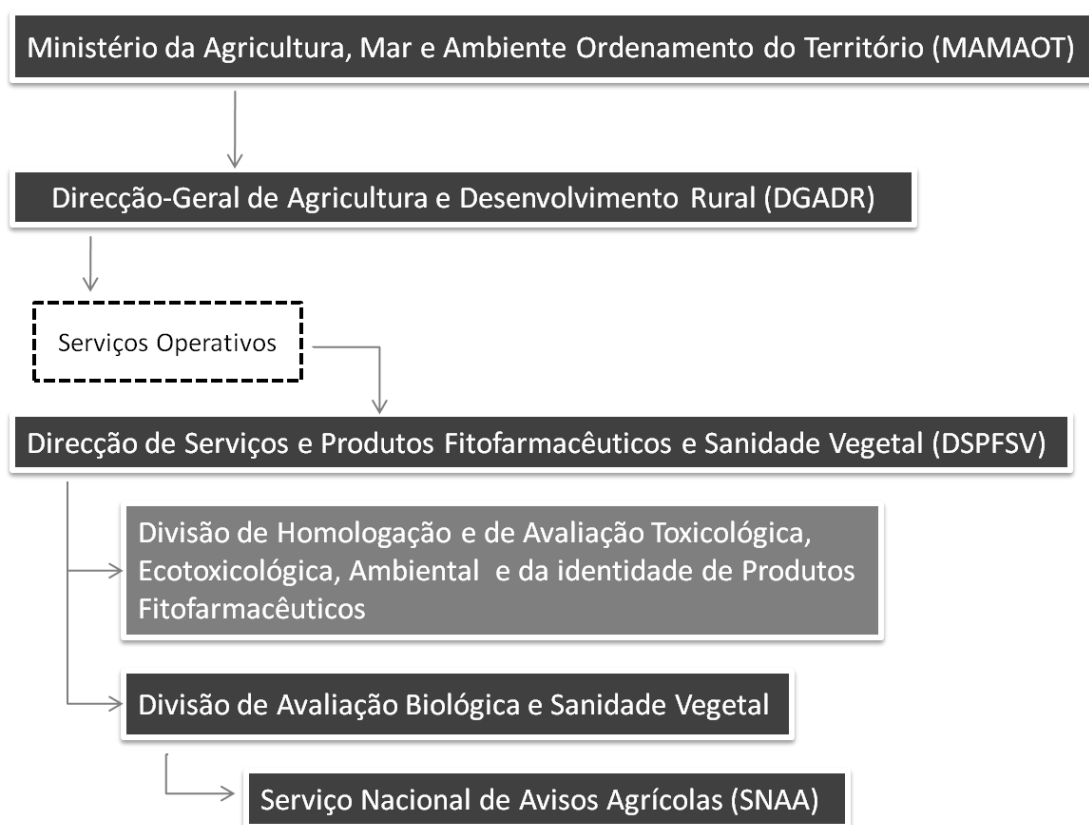
O âmbito deste projecto está delimitado com base num enquadramento dos objectivos, missão e descrição do objecto do negócio. Este projecto terá como objectivo a implementação de um modelo de *Data Warehousing* para o Serviço Nacional de Avisos Agrícolas (SNAA), pelo que se torna relevante definir no que consiste este mesmo organismo.

### **I.5.1. Serviço Nacional de Avisos Agrícolas**

O SNAA é um organismo de âmbito nacional, coordenado pela Divisão de Avaliação Biológica e Sanidade Vegetal. Da lista de competências desta divisão, é possível destacar três que são transmitidas sobre a alçada no SNAA, sendo as seguintes (cf. ponto 5.2 do Despacho nº 9185/2007 de 21 de Maio de 2007):

- i.** Colaborar na coordenação do Serviço Nacional de Avisos Agrícolas (SNAA);
- ii.** Colaborar na validação e implementação de novos métodos de previsão e evolução dos inimigos das culturas de acordo com a boa prática agrícola;
- iii.** Criar e manter actualizada a base de dados com as informações meteorológicas, biológicas e fenológicas fornecidas pelas estações de avisos agrícolas;

A Divisão de Avaliação Biológica e Sanidade Vegetal encontra-se sob a alçada da Direcção de Serviços e Produtos Fitofarmacêuticos e Sanidade Vegetal (DSPFSV), que por sua vez pertence aos serviços operativos da Direcção-Geral de Agricultura e Desenvolvimento Rural (DGADR). A DGADR enquadra-se organicamente no Ministério da Agricultura, Mar e Ambiente Ordenamento do Território (MAMAOT). De uma forma resumida é possível observar o seguinte organigrama de precedências hierárquicas do SNAA, que se encontra abaixo:



**Figura 1** – Organograma de precedências hierárquicas do SNAA.

O SNAA tem por finalidade a emissão de avisos agrícolas através de circulares emitidos pelas EA. Estes avisos contêm informação relativa a (SNAA, 2010):

- i. Previsão das intervenções fitossanitárias, indicando aos agricultores, a nível de cada região as datas mais oportunas e os produtos fitofarmacêuticos mais aconselhados para o combate dos inimigos das culturas, de acordo com os princípios da protecção integrada;
- ii. Determinadas componentes de modo de produção biológico, nomeadamente, a protecção fitossanitária neste modo de produção;
- iii. Algumas práticas culturais, nomeadamente a rega no sentido de transmitir aos agricultores informações que lhes permitam efectuar uma gestão da água mais racional, de acordo com os princípios do modo de produção integrado.

O SNAA é composto estruturalmente por Estações de Avisos Agrícolas (EA) que por sua vez agrupam uma ou mais estações Meteorológicas (EMA ou EM). As EA são na sua maioria públicas e, de acordo com o SNAA no seu *website*, “enquadram-se nas Direcções Regionais de Agricultura e Pescas (DRAP), e nas Organizações de Agricultores (privadas)”, enquadrando-se no seu âmbito “efectuar as tarefas necessárias

para que o Aviso Agrícola chegue ao agricultor em altura oportuna”. O SNAA é ainda dotado tecnologicamente de infraestruturas centralizadas regionalmente - as Estações de Avisos Agrícolas Centrais (EAC) - que recebem os dados meteorológicos enviados pelas EA e, posteriormente, transmitem estas mesmas informações para um servidor central localizado na Sede do SNAA. A informação que transita para as EA é recolhida pelas EMA com uma periodicidade reduzida e enviada no final do dia para cada EA. Por sua vez, as EA enviam a informação para as EAC mediante a utilização de processos informáticos nocturnos. Para uma melhor contextualização da alargada cobertura geográfica do SNAA, observe-se a figura 2:



**Figura 2** – Localização geográfica das estações de Avisos Públicas e Privadas sobre a alçada do SNAA em Portugal Continental

Fonte: SNAA, 2010

As EA servem também de instrumentos de aconselhamento e sensibilização para que seja possível aos agricultores proteger de uma forma mais eficiente as suas culturas. Para este efeito são utilizadas as medidas agro-ambientais de modo de protecção integrada, modo de produção integrado e modo de produção biológico.

Estas medidas agro-ambientais foram adoptadas na reforma da Política Agrícola Comum (PAC), em Junho de 1992, com o objectivo de auxiliar os agricultores na prática de uma agricultura mais respeitadora do Homem e do Ambiente, tendo tido início em Portugal em 1994.

### **I.5.2. Medidas Agro-Ambientais**

O Serviço de Avisos deve proporcionar aos agricultores não só a informação que lhes assegura a melhor oportunidade de tratamento, mas também orientá-los no sentido de praticarem formas de protecção mais racionais, de acordo com os princípios de protecção integrada. De seguida serão definidas cada uma das três medidas defendidas pelo SNAA.

#### **Protecção Integrada**

A protecção integrada é definida segundo a *Food Agriculture Association* (FAO) como um “Sistema de Protecção contra os inimigos das culturas que, tomando em consideração as condições particulares do Ambiente e da dinâmica das populações das espécies em questão, utiliza todos os meios e técnicas apropriadas de modo tão compatível quanto possível”. Esta metodologia tem como objectivo manter as populações dos inimigos das culturas a um nível suficientemente baixo para que os prejuízos ocasionados sejam economicamente toleráveis. Para Amaro & Baggio (1982) a “Protecção Integrada situa-se no âmbito da Protecção das Plantas (Protecção das Culturas ou Fitiatria) que é definida como a ciência que estuda os inimigos das culturas e o conjunto de meios de protecção para os evitar e combater”. Dias (2003), destaca o papel da FAO a nível Europeu “A fórmula utilizada pela *Food Agriculture Association* (FAO) torna abrangente o conceito de modo a que todos os Países utilizem

a mesma terminologia e tenham o mesmo entendimento”. Adicionalmente, faz um enquadramento histórico no que concerne à protecção integrada, demonstrando o papel do SNAA: “a implementação deste modo de Produção foi-se fazendo muito lentamente, primeiro com a criação do Laboratório de Fitofarmacologia, depois com o Sistema Homologação dos produtos fitofarmacêuticos e finalmente, e também muito importante, foi a implementação do Sistema Nacional de Avisos”.

### **Modo de Produção Integrado**

A definição do modo de produção integrado é largamente aceite como “um sistema agrícola de produção de alimentos de alta qualidade que utiliza os recursos naturais e mecanismos de regulação natural em substituição de factores de produção prejudiciais ao ambiente e de modo a assegurar, a longo prazo, uma agricultura viável” (Aguar, Godinho & Costa, 2005). Esta definição foi inicialmente proposta pela *International Organization for Biological and Integrated Control of Noxious Animals and Plants* (IOBC) e pela *West Palaearctic Regional Section* (WPRS), afirmando ainda que “em produção integrada, é essencial a preservação e melhoria da fertilidade do solo e da biodiversidade e a observação de critérios éticos e sociais”. Contextualizando esta definição ao nível português, verifica-se uma evolução das medidas agro-ambientais inicialmente estabelecidas e também a introdução de apoios financeiros compensatórios, de modo a incentivar este modo de produção, tratando-se de um regime de produção menos agressivo e intensivo para o ambiente, baseado nos princípios da defesa do Homem e do meio Ambiente.

Conforme é descrito pela Associação de Jovens Agricultores de Portugal (AJAP) são inúmeras as vantagens desta metodologia. Com a adopção destes modos de produção os produtores passaram a praticar uma agricultura sustentável e colocar no mercado produtos com maior segurança alimentar, adoptando métodos de produção agrícola compatíveis com as exigências de protecção do ambiente e de preservação do espaço rural, usando racionalmente os pesticidas, utilizando-os só quando outras alternativas não se adequavam e estes se tornavam indispensáveis. (AJAP, 2010)

## **Modo de Produção Biológico**

O modo de produção biológico é resultante da produção de produtos alimentares por intermédio da utilização de boas práticas de agricultura biológica. Estes produtos são definidos por Ladeira (2010) como “produtos vegetais e animais, transformados ou não, (...) alimentos para animais, mais recentemente, são aqueles que possuem um conjunto de regras definidas para a sua produção, preparação, importação, rotulagem e o seu próprio controlo, em todos os Estados Membros da União Europeia”. A utilização deste modo de produção pode ser definida através da não utilização de adubos e pesticidas, promotores de crescimento, aditivos e conservantes de síntese, alimentos e de organismos geneticamente modificados.

### **I.5.3. Contexto Europeu e Mundial**

No contexto europeu e inclusive mundial existem diversos organismos que se enquadram na temática da agricultura e de todos os aspectos que a compõem na generalidade. Contudo, contextualizando o SNAA em Portugal, é relevante indicar e descrever os organismos que contêm semelhantes políticas de protecção fitossanitária e de agricultura sustentável.

A nível mundial, a organização com uma expressão mais vincada é a *International Organization for Biological and Integrated Control of Noxious Animals and Plants* (IOBC), que colabora com a *International Union of Biological Sciences* (IUBS), cujo objectivo é defender o modo produção biológico, a gestão integrada de pesticidas e a produção integrada. A IOBC é composta por seis secções regionais, sendo a mais relevante a secção Oeste, a *West Palaearctic Regional Section* (WPRS), que enquadra vinte e quatro países europeus, a região do mediterrâneo e a do médio oriente. Esta secção é caracterizada por ser a mais activa, sendo composta por vinte grupos de trabalho e cinco comissões. As suas principais actividades são a organização e promoção de eventos nos territórios dos seus membros associados e a divulgação e produção de material científico, livros e *newsletters*, incluindo também um boletim científico servindo de referência na produção orgânica e em alimentos orgânicos a nível mundial. A IOBC/WPRS acaba ainda por incentivar e promover o desenvolvimento de

sistemas de produção e protecção biológica integrada, através da investigação e do estudo de casos práticos, assim como da divulgação de informação e de formação especializada.

No contexto europeu existem ainda duas organizações responsáveis pela promoção de agricultura sustentável, a *European Initiative for Sustainable Development in Agriculture* (EISA) e a *European Crop Protection Association* (ECPA), esta última mais focada na utilização e comercialização de produtos fitofarmacêuticos nas plantas.

A EISA é composta por um conjunto de organizações nacionais agrícolas de alguns países europeus, nomeadamente, França, Alemanha, Luxemburgo, Suécia, Reino Unido e Áustria. Esta organização promove, através de um conjunto de boas práticas, o conceito de produção integrada na agricultura como regra para um desenvolvimento sustentável da agricultura europeia.

De modo semelhante, a ECPA tem como missão a promoção dos interesses das associações nacionais dos diferentes países da Comunidade Europeia no que respeita à agricultura tecnológica, no contexto da agricultura sustentável, agindo como embaixadora das políticas europeias neste meio. A ECPA tem como associada portuguesa a *Associação Nacional da Indústria para a Protecção das Plantas* (ANIPLA) cuja representatividade existe desde 1992, e que representa as empresas que desenvolvem e comercializam produtos fitofarmacêuticos utilizados na agricultura. Esta associação congrega mais de 95% do Mercado Nacional de Produtos Fitofarmacêuticos (ANIPLA, 2011)

## II - CONTEXTO

**“Things should be made as simple as possible,  
but not any simpler.”**

*Albert Einstein*

Neste capítulo serão descritas todas as metodologias tecnológicas, assim como estudos relevantes, com vista a contextualizar este relatório de projecto. De modo a sintetizar o conhecimento conceptual sobre as questões abordadas, será utilizada uma forma de revisão de literatura baseada em sínteses conceptuais.

### II.1. Business Intelligence

Segundo Mundy, Thornthwaite e Kimball (2006) a *Business Intelligence* serve como uma ferramenta de auxílio aos 90% de utilizadores de negócio de uma empresa para que estes possam utilizar e extrair a sua informação de uma forma simplificada. Para isso, os utilizadores recorrem às aplicações de BI e aos seus repositórios, *Data Warehouses* (DW) e *Data Marts* (DM). Os mesmos autores referem ainda que o esforço dispendido na construção de aplicações de BI, sem o suporte de repositórios de dados dimensionais e sem processos de extração, transformação e carregamento (ETL), é grande e ineficaz. De forma a abordar correctamente a gestão de dados, a própria criação de repositórios de dados dimensionais é efectuada com o recurso a ferramentas de BI, sendo que a utilização ou construção de ambas é mutuamente exclusiva.

Para Barbieri (2001), a BI permite aceder, explorar e estruturar as informações armazenadas num DW ou DM. Esta capacidade, conjugada com as ferramentas tecnológicas adjacentes, determinou a forma como as tecnologias de informação são encaradas actualmente, ou seja, a BI é, cada vez mais, vista como uma melhoria necessária a implementar em cada organização e como catalisador da competitividade de uma empresa. A figura 3 exhibe as principais componentes de um ambiente de BI.

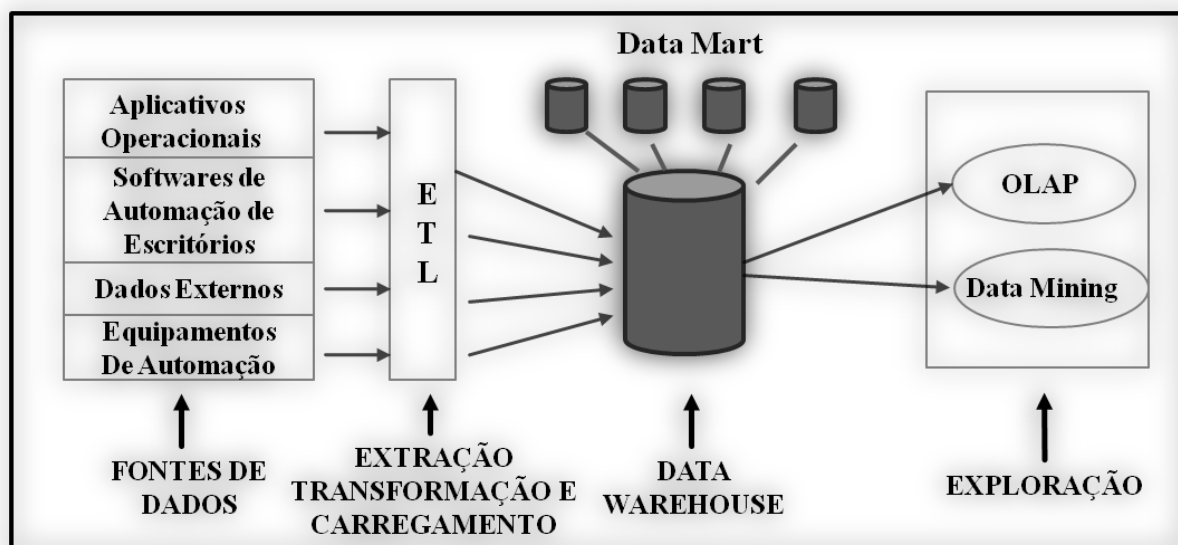


Figura 3 – Componentes de um ambiente de *Business Intelligence*.

Conforme é possível observar, o ambiente de BI é composto por um sistema de bases de dados transaccional, por um processo de ETL que carrega os dados no DW e, por fim, por um conjunto de ferramentas que permitem efectuar a exploração dos dados. Estas ferramentas efectuem análises designadas por *Online Analytical Processing* (OLAP) e utilizam metodologias como *Data Mining* que visam a identificação de padrões, relacionamentos ou modelos, implícitos nos dados armazenados em grandes bases de dados (Han & Kamber, 2001).

Mais recentemente, Rainardi (2008) afirma, correctamente, que a BI é o conjunto de actividades que permite conhecer as situações envolventes ao negócio, mediante a execução de várias análises, tanto em dados internos, como externos. Desta forma, tem como objectivo auxiliar a tomada de decisões estratégicas, táticas e operacionais e aumentar o desempenho da empresa no mercado.

## II.2. Data Warehouse

A definição clássica de um DW introduzida primariamente por Inmon (1992) refere-se a um repositório de dados “organizados por assunto, integrados, variando no tempo e não voláteis que suportam os processos de tomada de decisão dos gestores”. Concordando com esta perspectiva, Gill e Rao (1996) afirmam que, apesar de algumas

divergências, existe um consenso determinando que um DW é uma componente essencial nas soluções de suporte à decisão das empresas. Posteriormente ainda referem que as principais vantagens da utilização de um DW são as seguintes:

- i. “Orientação para o utilizador final” O formato e a organização dos dados estão colocados na perspectiva do utilizador, ou seja, orientados para análise.
- ii. “Possibilidade de guardar informação histórica e de agregar grandes quantidades de informação”. Os dados estão disponíveis, limpos e organizados num formato padrão, sendo continuamente actualizados sem perda de histórico de informação.
- iii. “Possibilidade de integrar informação de variados repositórios de dados”.
- iv. “Aumentar a produtividade empresarial”.

### **II.2.1. Normalização**

A normalização é decorrente do conceito de modelo relacional nas bases de dados, primeiramente abordado por Edgar F. Codd em 1970. Este descreve que a normalização numa base de dados aumenta substancialmente a integridade dos dados nela inseridos, pois reduz o impacto das transacções processadas. Por outro lado, o aumento na normalização de uma base de dados reduz a performance da mesma, ou seja, a rapidez de acesso aos dados. De acordo com os objectivos que se pretendem atingir com a implementação de um DW, a definição standard deste sistema é equivalente à de desnormalização, ou seja, à da eliminação da complexidade nas pesquisas mas colocando entraves à modificação dos dados. De um modo geral, as bases de dados relacionais, utilizadas para o processamento transaccional de dados (OLTP) são tipicamente normalizadas, sendo que as bases de dados utilizadas para processamento analíticos de dados (OLAP) são tendencialmente não normalizadas.

### **II.2.2. Modelo Dimensional**

Um modelo dimensional é uma abordagem metodológica utilizada no desenho de um DW e DM's, representando, na prática, a lógica de modelação da base de dados. (Moody, Kortink & Mark, 2000). Segundo Mundy et al. (2006), um modelo dimensional apresenta como vantagens os seguintes factores:

- i. Apresenta a informação de uma forma simples;
- ii. Retorna os resultados das pesquisas de uma forma mais eficiente;
- iii. Dá informação relevante para o negócio;
- iv. É simples para o utilizador final de perceber o conteúdo estrutural da informação e a forma como esta está agregada;

Pelas razões acima apresentadas, um modelo dimensional é indicado como a escolha mais adequada para a estrutura de um DW. Adicionalmente, é ainda referido que “como a estrutura dimensional está desnormalizada o utilizador não necessita de efectuar junções de tabelas e criar tabelas temporárias nas suas pesquisas”, aumentando desta forma a eficiência no trabalho desenvolvido pelo utilizador final.

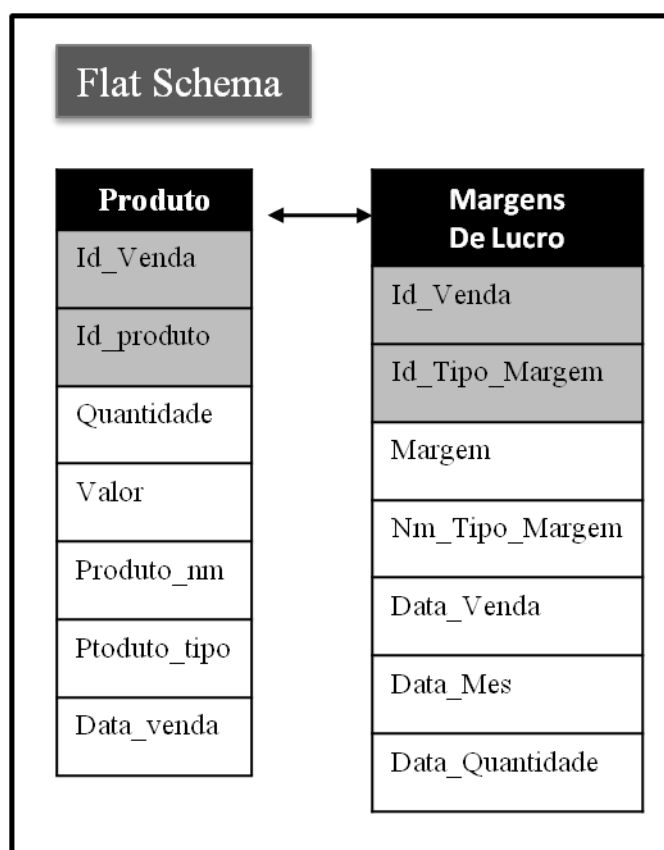
A abordagem mais utilizada e sugerida por Kimball e Ross (1996) consiste num modelo dimensional constituído por tabelas centrais de factos, denominadas *Fact Tables* e pelas suas dimensões associadas. A tabela de factos é constituída por um conjunto de chaves estrangeiras originárias das tabelas de dimensões - *surrogate keys* - conjuntamente com indicadores calculados nos processos de transformação. Realça-se ainda o principal aspecto anexo às tabelas de factos - a granularidade atribuída - que se traduz na definição do mais elevado grau de detalhe que é suportado num DW.

No entanto, existem outras abordagens à modelação dimensional. Abaixo serão descritas as metodologias mais comuns e a sua constituição, o enfoque recai sobre os seguintes: *Flat Schema*, *Terraced Schema*, *Star Schema* e *Snowflake Schema*.

### **II.2.3. Flat Schema**

Este é o modelo multidimensional mais simples, pois permite a construção de um modelo dimensional sem que exista a perda de dados. Esta metodologia é

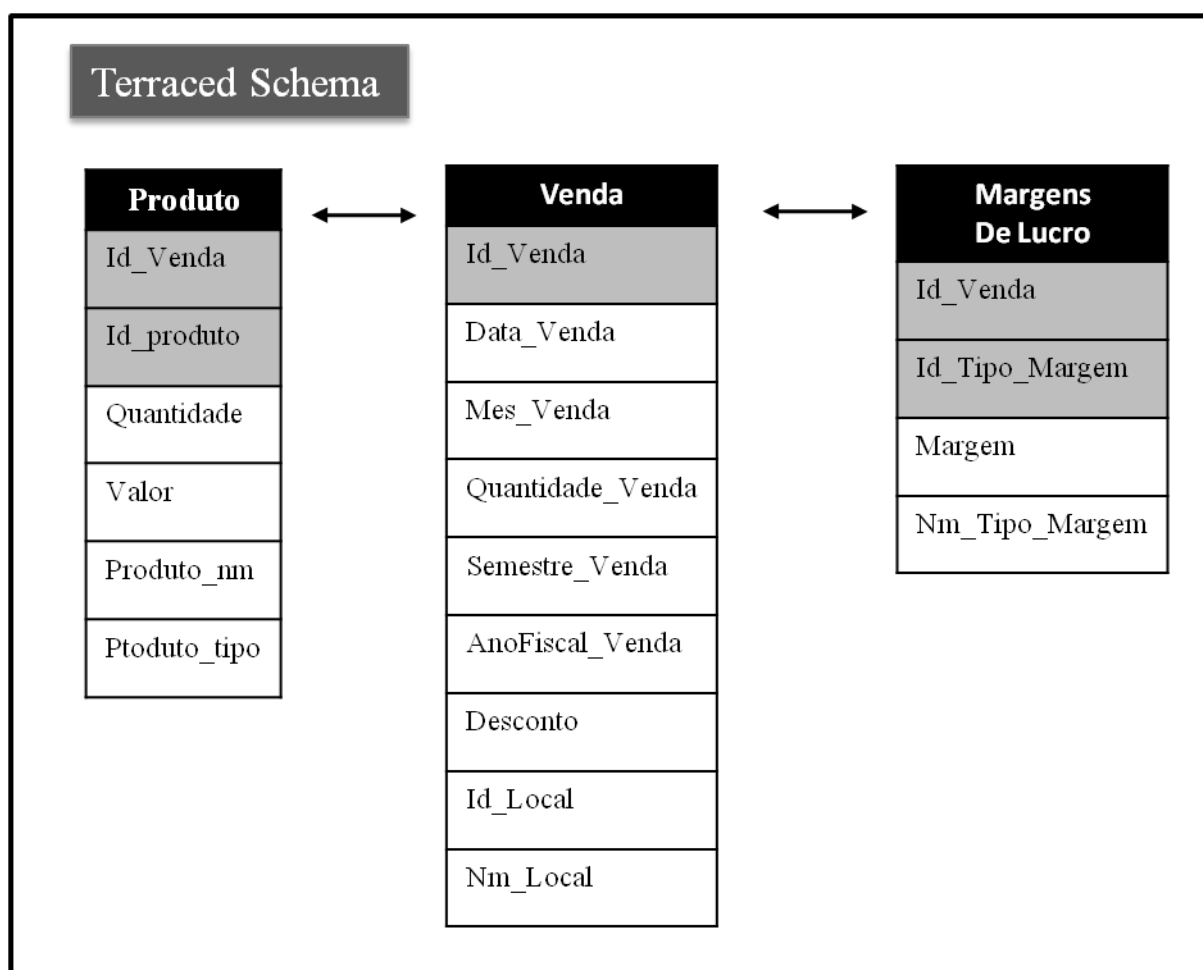
caracterizada pela não utilização de agregações, ou seja, através da minimização de entidades e, por conseguinte, do número de tabelas a criar. No entanto, estas tabelas são geralmente compostas por muitos atributos e pela redundância destes mesmos atributos, ou seja, diminui-se a complexidade relacional mas aumenta-se a complexidade de cada tabela (Moody et al., 2000). Na figura 5 é possível observar um exemplo desta aplicação.



**Figura 4** – Exemplo da Metodologia dimensional *Flat Schema*

## II.2.4. Terraced Schema

Com este modelo obtém-se uma arquitectura composta apenas pelas entidades transaccionais, existindo uma tabela por cada uma destas mesmas entidades (figura 6). As tabelas obtidas possuem as mesmas chaves primárias relativamente ao modelo relacional e não são aplicadas operações de agregação (Moody et al., 2000).



**Figura 5** – Exemplo da Metodologia dimensional *Terraced Schema*

## II.2.5. Star Schema

Este modelo possui uma complexidade superior ao modelo anterior e pode ser descrito pela inclusão de tabelas centrais denominadas tabelas de factos, que por sua vez são compostas por entidades transaccionais. A chave primária desta tabela será a junção das chaves primárias de todas as entidades componentes associadas. Conjuntamente, existem ainda as tabelas de dimensões que são formadas por cada entidade componente, que congregam ainda as entidades de classificação. Sendo assim, as entidades transaccionais estarão ligadas com as entidades componentes e de classificação mediante uma hierarquização das mesmas, permitindo a navegação pelos atributos de uma forma optimizada. Para permitir estas ligações, todos os atributos numéricos nas tabelas de factos que não sejam considerados como métricas deverão estar agregados numa chave primária e ligados às respectivas dimensões. Na figura 7 é possível observar um exemplo de um *Star Schema*.

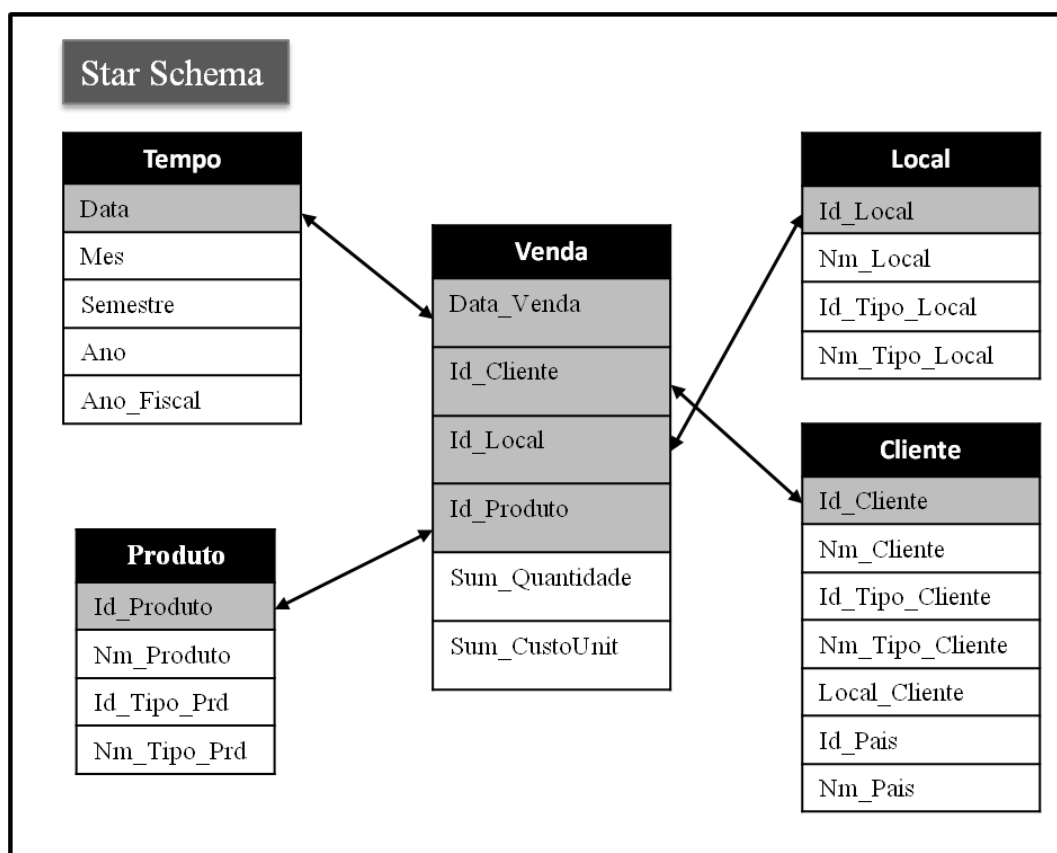


Figura 6 – Exemplo da Metodologia dimensional *Star Schema*

## II.2.6. Snowflake Schema

Conforme foi descrito no modelo anterior, o método *Star Schema* provém de uma desnormalização do modelo relacional em várias tabelas de dimensões e/ou factos. O modelo *Snowflake Schema* é composto pela normalização das hierarquias das dimensões do modelo *Star Schema* (figura 8). É possível então afirmar que é, dentro dos modelos mencionados, o mais semelhante ao modelo relacional.

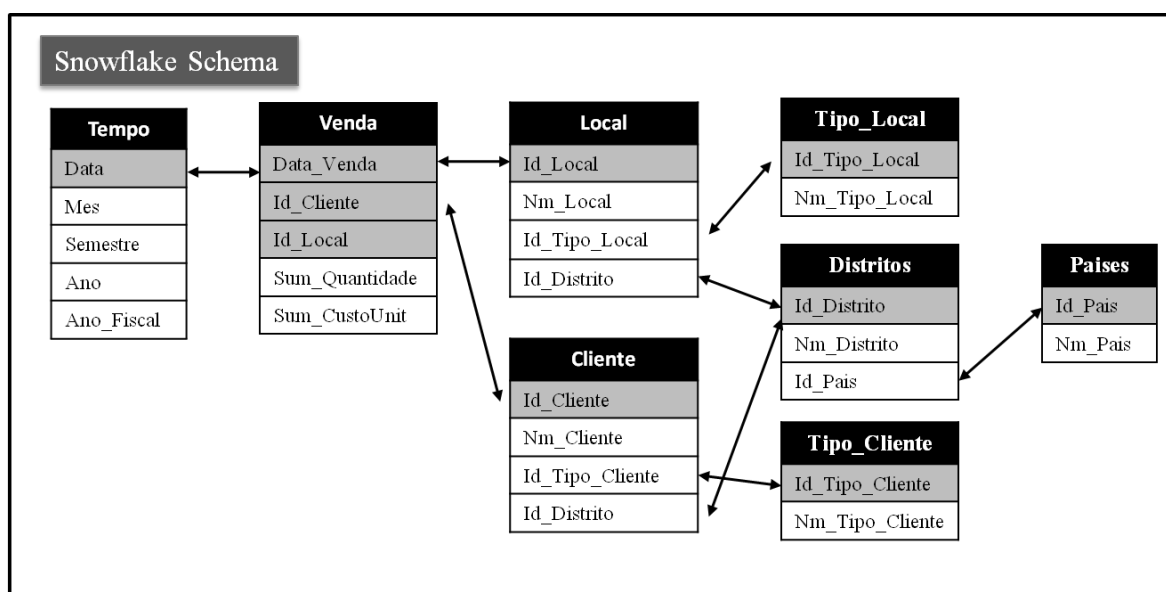


Figura 7 – Exemplo da Metodologia dimensional *Snowflake Schema*

## II.2.7. Metadados

Metadados é uma expressão comumente utilizada para referir “dados sobre dados”, sendo descrita por Gill e Rao (1996) como uma “roadmap to data”, ou seja, uma directiva para os dados. Estes autores acrescentam ainda que metadados são “todos os objectos da base de dados, tudo o que seja pertinente a um DW (tabelas, colunas, *views*, *query's*, *procedures*)”. Marco (2000), leva a definição de metadados a outro patamar, referindo que metadados são “conhecimento” e que são “o conjunto de todos os dados físicos (contidos nos softwares) e o conhecimento (contido nos empregados) de dentro e para dentro da organização, incluindo informação sobre os dados físicos, os processos técnicos e de negócio, regras e restrições dos dados, e estruturas de dados utilizadas por uma organização”. Posteriormente, faz uma divisão entre metadados - “conhecimento sobre o nosso sistema, o nosso negócio e o nosso mercado” - e repositório de metadados “as tabelas físicas utilizadas para armazenar os metadados e que serão disponibilizadas aos utilizadores finais”.

São apresentados por Gill e Rao (1996) como as principais vantagens dos metadados os seguintes factores:

- i. Identificação das fontes operacionais;
- ii. Indicação das nomenclaturas dos objectos;
- iii. Segurança da informação;
- iv. Definições físicas e lógicas das colunas, tabelas e atributos da base de dados;
- v. Implementação de regras e automatização das hierarquias dimensionais;

Mundy et al., (2006) partilham da mesma opinião e acrescentam que os metadados são “toda a informação que descreve os conteúdos, estruturas e operações e um DW”. Pegando ainda neste conceito dividem os metadados mediante três abordagens: técnicos, de negócio e processuais.

Mais recentemente Caldeira (2008) descreve metadados como “um sistema de informação homogéneo”, afirmando que “a metodologia proposta pelo modelo Common Warehouse Metamodel (CWM) é uma das soluções para estas questões de uniformização dos metadados nesta classe de sistemas”. Acrescenta ainda que “o

modelo CWM fornece uma estrutura de suporte comum para a representação dos metadados, das fontes dos dados, do repositório de dados e dos métodos de transformação e análise da informação”. De seguida é apresentado o modelo CWM de modo a contextualizar a estrutura padrão dos metadados:

<b>Gestão</b>	<b>Processos do DW</b>			<b>Operações no DW</b>		
<b>Análise</b>	<b>Transformação</b>		<b>OLAP</b>	<b>Data Mining</b>	<b>Visualização da informação</b>	<b>Nomenclatura</b>
<b>Recursos</b>	<b>Modelo por objectos</b>	<b>Modelo Relacional</b>	<b>Registos Textuais</b>	<b>Modelo Dimensional</b>		<b>XML</b>
<b>Suporte</b>	<b>Informação do Negócio</b>	<b>Tipo de dados</b>	<b>Expressões</b>	<b>Chaves</b>	<b>Análise dos Domínios</b>	<b>Software Deployment</b>
<b>Infra-Estrutura</b>						

**Figura 8** – O MetaModelo *Common Warehouse Metamodel*.  
**Fonte:** Adaptado Poole et al., 2002

### II.3. ETL

A sigla ETL tem uma denominação transversal do processo que descreve. Este processo consiste num procedimento informático que engloba todas as operações realizadas com a localização dos dados, as rotinas de extracção, escolha, limpeza, transformação e carregamento da informação transaccional no DW. Segundo Todman (2001), ETL é utilizado para descrever “o processo de extracção de dados de um sistema fonte e posteriormente de modificação ou transformação desses mesmos dados para um formato que seja mais aceitável para o DW do que o seu formato inicial”. Mundy et al. (2006) vão mais longe e afirmam que “um sistema ETL é o centro da fiabilidade de um sistema de Business Intelligence” e que o seu desenvolvimento é “a actividade que demora mais tempo e que consome mais recursos num projecto de DW/BI”, alertando as dificuldades que existem no desenvolvimento da lógica dimensional de um sistema ETL.

Como é possível observar na figura 10, abaixo apresentada, o sistema ETL apresenta um lugar de destaque relativamente ao conjunto de um projecto informático de BI.

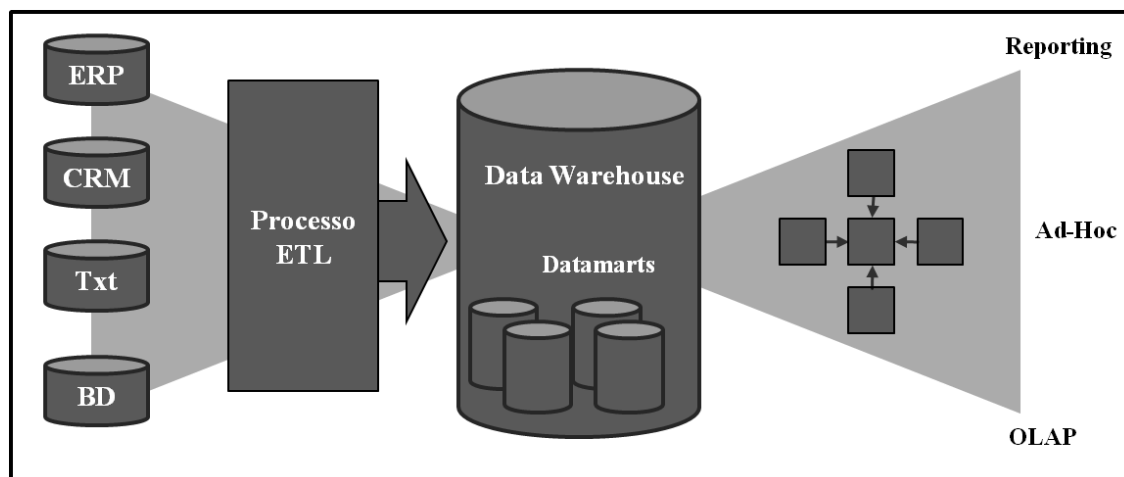


Figura 9 – Sistema de *Business Intelligence*.

Usualmente, o sistema ETL é dividido em três grandes partes: a extracção, transformação e o carregamento. A extracção, conforme foi descrito anteriormente, pressupõe retirar dados de um sistema de bases de dados. Esta etapa implica um especial cuidado no que concerne ao formato dos dados e à selecção apenas dos dados que são necessários para o projecto em curso. Usualmente, de modo a auxiliar este processo é criada por razões de segurança uma área intermédia de camada de dados no procedimento de extracção dos dados, denominada por SA. Todman (2001) descreve os benefícios desta camada do processamento, enunciando que, caso surja um problema no carregamento dos dados, existe uma versão intermédia com dados já extraídos, o que poderá poupar um número significativo de horas no processo ETL.

De seguida tem-se a etapa da transformação dos dados, que inclui tratamento dos dados, alterações no formato, conversões, junções, criação de valores e tratamento de *outliers*<sup>1</sup>. Por último, a etapa de carregamento traduz o carregamento dos dados devidamente transformados num DW. Este processo despoleta um conjunto de acções na base de dados, desde actualizações a verificações referenciais e de integridade dos dados. O carregamento é, assim, a última fase do processo ETL.

<sup>1</sup> *Outlier* - Termo utilizado em estatística que significa uma observação que seja numericamente distante do conjunto dos dados.

### III - EXPLICAÇÃO DOS PROCEDIMENTOS

“What i hear i forget. What i see i remember.

What i do i understand.”

*Confúcio*

Como foi explicado anteriormente, é pretendida com a execução deste projecto a construção de uma base de dados multidimensional, otimizada para pesquisa que suporte as actividades do SNAA. Para permitir o correcto carregamento dos dados que irão fazer parte deste novo modelo de base de dados será necessária a implementação de um processo já descrito anteriormente e designado por ETL. Os dados que irão ser carregados neste novo modelo estão actualmente numa base de dados em modelo relacional que se encontra num servidor. A construção da referida BD estará enquadrada em duas fases:

- i. Numa primeira fase, será implementada uma plataforma de Business Intelligence que interligue toda esta informação, otimizando a recolha os dados meteorológicos, tratando os dados recolhidos e carregando estes dados numa base de dados não relacional, num DW, através de um processo informático ETL.
- ii. Numa segunda fase pretende-se construir um protótipo de um modelo de tratamento de dados que permita a disponibilização de informação mais fiável sobre os dados meteorológicos recolhidos, ou seja, será implementada uma solução de *Data Quality* nos dados.
- iii. Pretende-se a disponibilização da informação numa ferramenta de *Reporting* apelativa e com indicadores relevantes ao negócio.

### **III.1. Estrutura das Bases de dados**

Neste subcapítulo (III.1.) será abordada a temática da construção da estrutura dimensional da SA e do DW e serão igualmente definidos os diferentes tipos de entidades, hierarquias, assim como a granularidade a adoptar.

Conforme foi descrito anteriormente, o objectivo deste projecto prende-se com a construção de um DW. Para que este objectivo seja alcançado é vantajoso seguir um conjunto de procedimentos definidos como boas práticas na implementação de modelos dimensionais (Mundi et al., 2006):

- i.** Estabelecer os requisitos de negócio;
- ii.** Definir a granularidade;
- iii.** Definir as dimensões e as tabelas de factos, através da classificação as entidades e da identificação das hierarquias;

Estes procedimentos permitirão o desenho e posterior implementação do DW e, por conseguinte, da base de dados intermédia, a SA.

#### **III.1.1. Requisitos de Negócio**

Conforme já foi abordado nos objectivos do projecto subjacente, a construção de um DW eficiente, exige o estabelecimento de requisitos de negócio. Estes requisitos ou, processos do negócio podem ser alcançados através da resposta à seguinte pergunta: o que queremos medir? Para este projecto em concreto, o que se pretende medir é a análise das diferentes variáveis meteorológicas para determinados momentos, de modo a produzir estatísticas e modelos de procedimentos de controlo de pragas eficientes e robustos.

### **III.1.2. Granularidade**

A declaração da granularidade na tabela de factos é um aspecto essencial na modelação dimensional. Esta é descrita como a definição do mais elevado grau de detalhe que é suportado num DW (Mundi et al., 2006), ou seja, o máximo de detalhe que será definido. Para a obtenção da granularidade foram colocadas as seguintes questões:

- i. Qual a temperatura média na estação de avisos do Algarve para um dia em específico?
- ii. Probabilidade de existência uma praga, por exemplo traça da azeitona, cochonilha vermelha, numa específica estação de avisos?

### **III.1.3. Classificação das Entidades**

Após a definição da granularidade pretendida, o próximo passo consiste em classificar cada entidade numa das três categorias seguintes:

#### **Entidade Transaccional**

As entidades transaccionais representam detalhes de transacções, são os atributos inerentes ao funcionamento do negócio e que, por isso, representam a base do modelo dimensional. Estas entidades representam um evento que foi realizado num determinado momento e que detém uma quantidade que é possível de ser medida (Moody et al., 2000)

As entidades classificadas nesta categoria são as seguintes:

- i. Dados\_d.

O campo Dados\_d é referente ao indicador, chave primária, da tabela de dados relacional, respeitante aos dados diários.

## Entidade de Componente

As entidades componentes são as que estão directamente ligadas às entidades transaccionais, através de um relacionamento de um-para-muitos. Estas entidades definem os detalhes de cada transacção e respondem às seguintes perguntas: Quem? O quê? Quando? Onde? Como? e Porquê? (Moody et al., 2000). Estas entidades formam a base para a construção de dimensões no modelo *Star Schema*. As entidades classificadas nesta categoria são:

- i. N;
- ii. Data;

O campo N é respeitante à estação de avisos na qual se enquadram as métricas recolhidas, ou seja, correspondente a um indicador geográfico. Por sua vez, o campo Data, como o próprio nome indica significa a data em que foi efectuada a recolha de informação.

## Entidade de Classificação

As entidades de classificação estão ligadas às entidades de componentes da mesma forma que estas estão ligadas às entidades transaccionais, ou seja, de um-para-muitos (Moody et al., 2000). Estas entidades representam hierarquias no modelo dimensional que pertencem a dimensões. As entidades de classificação são as seguintes:

- i. EA\_Tipo;
- ii. EM\_Tipo;
- iii. RH\_Formacao\_Tipo;
- iv. RH\_Tipo\_funcao;

O campo EA\_Tipo permite distinguir se a estação de avisos é central, agrupa várias estações de avisos, ou simples. Similarmente, o campo EM\_Tipo apresenta a mesma distinção, referindo, no entanto, às estações meteorológicas. Por sua vez, o campo RH\_Formacao\_tipo indica qual o enquadramento hierárquico dos funcionários (ex: Eng<sup>o</sup> Agrónomo, Assistente Administrativo, etc..), enquanto que o campo RH\_Tipo\_funcao indica a sua categoria profissional. No que diz respeito à classificação

dos campos nas diferentes entidades, estes podem ser incorporados em várias categorias, sendo por isso necessário proceder à seguinte precedência de classificação:

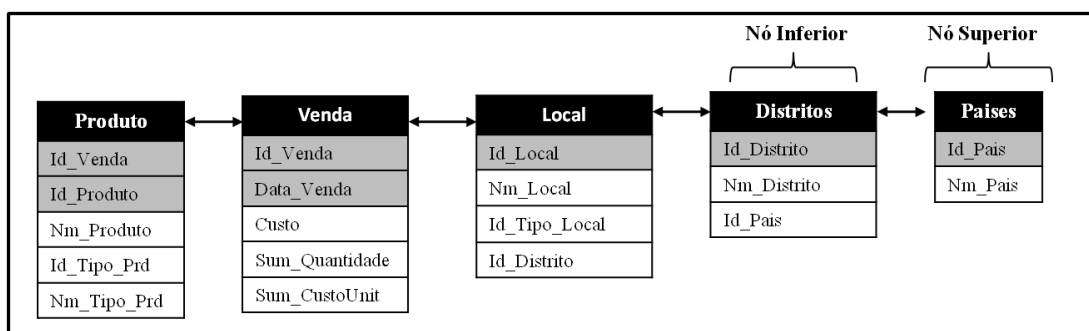
- i. Entidade Transaccional;
- ii. Entidade de Classificação;
- iii. Entidade de Componente;

Alerta-se para o facto de que algumas entidades não foram incluídas no modelo dimensional, pois não se aplicam ao contexto de negócio em vigor.

#### **III.1.4. Identificação das Hierarquias**

Após a classificação das entidades, será necessária a identificação das suas hierarquias correspondentes. De acordo com Moody et al. (2000), as hierarquias poderão ser máximas ou mínimas e estarão interligadas entre si pelas ligações correspondentes e provenientes do modelo relacional. Uma entidade é classificada como mínima quando se encontra no topo de uma hierarquia máxima, ou máxima quando se encontra na hierarquia mínima. As hierarquias são extremamente importantes na modelação dimensional e formam a base para a transformação de um modelo relacional num modelo dimensional. Em suma, uma hierarquia acaba por ser um conjunto de entidades relacionadas.

Existem assim dois tipos de construção de hierarquias: a hierarquia *Top-Down* e a hierarquia *Bottom-up*. A primeira utiliza a metodologia de referência de Bill Inmon (Inmon, 1992), que consiste na construção de um DW normalizado que incorpora vários DM da organização. A segunda é defendida por Ralph Kimball (Kimball & Ross, 1996), na qual os DM são criados e definidos em primeiro lugar. Estes DM contêm as dimensões e as tabelas de factos e dizem respeito a uma área do negócio. Seguidamente, caso exista a necessidade de interligar os vários DM's construídos é utilizada a metodologia de integração apelidada de *data warehouse bus architecture*. Na figura 12 é apresentado um exemplo de uma hierarquia.



**Figura 10** – Exemplo de uma hierarquia.

Assim, foram identificadas as hierarquias descritas de seguida:

- i. Hierarquia Território;
- ii. Hierarquia Empregados;
- iii. Hierarquia Tempo;

### III.1.5. Escolha do Modelo Multidimensional

Esta fase diz respeito à escolha do modelo multidimensional a ser utilizado. Relativamente ao que é explicitado acima, aquando da definição dos vários modelos dimensionais, estes modelos podem ser separados pela sua complexidade e redundância, conforme é possível observar na figura 4. Verifica-se assim que o modelo dimensional *Flat Schema* apresenta uma menor complexidade e, por conseguinte, uma maior redundância na informação. Por outro lado, o modelo *Snowflake Schema* apresenta uma maior complexidade e uma menor redundância. É possível também apontar uma maior normalização ao modelo *Snowflake Schema* e uma maior desnormalização ao modelo *Flat Schema*.



**Figura 11** – Características dos diferentes modelos dimensionais

Atendendo a estes factos, o modelo dimensional que mais se ajusta ao projecto é o modelo *Star Schema*. Este modelo abrange a complexidade desejada e, através da divisão do modelo dimensional em factos e dimensões, permite uma análise otimizada ao objectivo de negócio. Conforme Kimbal e Ross (1996) referem, o modelo *Star Schema* é o que melhor suporta os sistemas de tomada de decisão devido à sua natureza simplificada.

### III.2. Staging Area

A SA é descrita por Kimbal e Ross (2002), como sendo mais do que um repositório de dados, sendo adicionalmente um conjunto de processos que resultam numa solução de ETL, ou seja, a SA é o elo de ligação entre as bases de dados operacionais e o DW. Um dos pontos essenciais a garantir é a salvaguarda da SA no que respeita a acessos e utilização de dados por parte dos utilizadores finais. A SA irá, por isso, estar restrita ao carregamento por parte de dados transaccionais provenientes da base de dados de origem, assim como a consultas e alterações de metadados. Adicionalmente, Haberli e Tombros (2001) afirmam que a SA deve conter todas as ferramentas necessárias para integrar, carregar, controlar a qualidade e tratar os dados meteorológicos.

O processo de construção da SA passou por uma recolha do conjunto de tabelas que servirão de base para as futuras tabelas de dimensões e de factos, tendo sido seguida

a lógica de implementação tecnológica e o conjunto de boas práticas informáticas. A figura abaixo apresenta o esboço lógico inicial das ligações das tabelas para a SA.

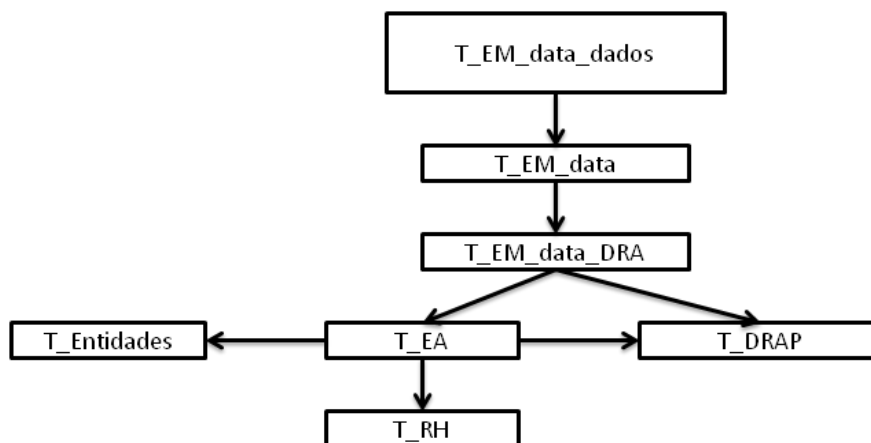


Figura 12 – Diagrama inicial das tabelas da *Staging Area*.

No entanto, este modelo foi aperfeiçoado para um modelo mais completo e adequado às dimensões e à tabela de factos (ver anexo VI.2.1.). Como se verifica, o objectivo na construção do modelo estrutural da base de dados da SA acabou por ser a melhor adequação dos dados para o carregamento final do DW. Além disto, também foi optimizada a sua transformação, ou seja, foi seguida uma estratégia de simplificação do modelo relacional actualmente em vigor, focado somente nas tabelas que viriam a ser utilizadas na SA e que por sua vez iriam ser transformadas para a obtenção de dimensões e métricas.

Tendo em conta esta transformação, devido ao volume de dados colocado na SA e de modo a ultrapassar os constrangimentos necessários à transferência dos dados, tornou-se necessário tirar partido das funcionalidades de performance incluídas no *Business Intelligence Development Studio*. As metodologias utilizadas para aumentar a performance foram as seguintes (Thompson, 2006; Duffy, 2008; Kejser & Lee, 2008; Ali, 2009; Microsoft, 2011):

- i. Eliminação da chave primária e da indexação automática nas tabelas de destino, ou seja, da SA. Estes índices e chaves serão posteriormente criados assim que seja terminada a transacção. A implementação de índices nas

tabelas que contêm maior volume de dados têm como objectivo agrupar os dados para que a sua extracção, ou visualização, seja efectuada de uma forma mais optimizada.

- ii. Não utilização de *querys* à base de dados através do pacote *Data Flow Task* de “Select \*” (*select all*) mas sim através da escolha exacta das colunas que são pretendidas. Esta metodologia permite prevenir a utilização de memória intermédia das conexões à base de dados de forma desnecessária e caso não seja seguida é despoletado um aviso aquando da compilação do projecto.
- iii. No *package data flow*, utilização da opção *fast column* e sem a opção de *check constraints, keep nulls e keep identity*. A restrição *check constraints* está definida por defeito pois permite garantir que nenhuma *constraint* é violada. Como neste caso foram retiradas todas as *constraints* da tabela de destino, de modo a aumentar a performance, não existe este problema.
- iv. Nos *stored procedures*, que suportam as transacções efectuadas na base de dados e que são utilizados no ETL, foram colocados valores máximos por cada transacção nos variados ciclos de validação, de modo a conseguir um canal fluido de comunicação por cada transacção.
- v. Alteração na base de dados da SA do modelo de recuperação da base de dados para *Simple*. Esta opção em detrimento de *full* permite reduzir drasticamente os *transaction logs* durante uma transacção, ou seja, estes não são actualizados por cada transacção que é efectuada na base de dados mas apenas após *backup* dessa mesma base de dados.

Conjugado com as boas práticas de implementação e manutenção de projectos informáticos no *Business Intelligence Development Studio*, foi incluído um conjunto de regras de transformação a aplicar às métricas recolhidas dos dados fonte.

## Regras de Transformação

Nada tem um maior potencial de deteriorar o desempenho e o valor organizacional de um DW do que um procedimento de qualidade dos dados inapropriado, não entendível ou não efectuado. (Ballou & Tayi, 1999). Conforme Lenzerini (2002) refere, a qualidade dos dados é essencial para o desenvolvimento e implementação de um DW. Desta forma, para Nilakanta, Scheibe e Rai (2008), o controlo de qualidade da informação recolhida é uma prioridade, pois faz com que os dados sejam considerados fiáveis para os utilizadores. Por conseguinte, a confiança dos utilizadores irá depender da qualidade da informação disponibilizada no DW. Musa et al. (2003) afirma que a automatização da recolha de dados meteorológicos constitui um desafio, tanto para o processamento, como para a gestão dos dados. Esta automatização na recolha dos dados é propícia a erros e leva à necessidade de um controlo robusto na qualidade dos dados. No que concerne a dados meteorológicos, os potenciais erros nos dados a ocorrer são os seguintes (Zahumensky, 2004; Neto, Pinheiro, Cavaco & Mendes, 2007):

- i. Erros aleatórios, de natureza aleatória e que podem ser positivos ou negativos. De modo geral, estes erros têm uma distribuição simétrica próxima de zero.
- ii. Erros sistemáticos, usualmente apresentam uma distribuição assimétrica próxima de zero. Estes erros apresentam um valor positivo ou negativo enviesado do valor real e podem ocorrer devido a uma falha na calibração do sensor da EM.
- iii. Erros elevados, estes erros são usualmente causados devido a problemas no sensor ou induzidos por erros no tratamento dos dados. Estes erros são tipicamente fáceis de detectar.

No que diz respeito ao controlo de qualidade de dados meteorológicos, em 2004, Zahumensky reuniu as linhas mestres para estes tratamentos de dados (Zahumensky, 2004; Salvati & Brambilla, 2008) e que viriam a ser adoptadas pela *World Meteorological Organization* (WMO). Estas consistem no controlo dos dados aquando a sua recolha e posteriormente, após a sua inserção no sistema de base de dados.

Aquando da sua recolha das EM, os dados deverão ser categorizados e validados de acordo com um conjunto de limites proposto. Passando esta etapa de validação, os mesmos deverão ser submetidos a validações de carácter estatístico que averiguem a sua consistência temporal e interna, assim como desvios-padrões e variâncias. O controlo de qualidade dos dados proposto por Zahumensky e outros autores (Eischeid, Baker, Karl & Diaz, 1995) pode ser assim dividido em: a) Valores Plausíveis, que se encontram dentro dos limites máximos e mínimos propostos; b) Consistência Temporal, de modo a avaliar o grau de alteração das medidas recolhidas em tempo real e entre duas recolhas; c) Consistência Interna, para avaliar a relação entre os parâmetros medidos por mais que uma EM, ou seja, é feita a comparação entre os valores medidos por uma EM com as suas estações vizinhas.

No presente projecto, os mecanismos de detecção da qualidade dos dados meteorológicos recaem essencialmente na componente de ETL, tendo em conta que outros procedimentos já foram implementados na recolha destes dados junto das estações meteorológicas e na inserção dos dados na base de dados relacional. Conforme Neto, Pinheiro, Cavaco e Mendes (2007) descrevem, foram designados dois procedimentos de validação dos dados recolhidos junto das EM.

O primeiro define parâmetros de classificação, através da atribuição de limites máximos e mínimos, os quais permitem identificar e tratar os erros de maior dimensão. Estes poderão ser: a) Plausíveis, aparentemente correctos; b) Suspeitos, que carecem de validação; c) Incorrectos.

Posteriormente, os dados catalogados como suspeitos ou incorrectos terão de ser analisados com base no seu histórico recolhido. Para esta análise estatística Neto et al. (2007) utilizou a metodologia estatística presente na tabela 1.

Valores observados	Limites
Valores Plausíveis	Mínimo Histórico $\leq$ Valores Plausíveis $\leq$ Máximo Histórico
Valores Incorrectos	Mínimo Histórico $>$ Valores Incorrectos $>$ Máximo Histórico
Variação de Valores observados	Limites
Valores Plausíveis	(Média Período $- 2 \times$ Desvio Padrão) $\leq$ Valores Plausíveis $\leq$ (Média Período $+ 2 \times$ Desvio Padrão)
Valores suspeitos	(Média Período $+ 2 \times$ Desvio Padrão) $>$ Valores Suspeitos $\leq$ Máximo Histórico
	(Média Período $- 2 \times$ Desvio Padrão) $<$ Valores Suspeitos $\geq$ Mínimo Histórico
Valores Incorrectos	(Média Período $- 2 \times$ Desvio Padrão) $\leq$ Valores Plausíveis $\leq$ (Média Período $+ 2 \times$ Desvio Padrão)

**Tabela 1** – Definição dos limites no controlo de qualidade dos dados.

**Fonte:** Adaptado Neto et al., (2007)

Conforme foi anteriormente referido, as regras de transformação foram efectuadas nos dados diários, tendo em conta que estes eram aqueles cuja aplicabilidade do ponto de vista do negócio faria sentido. Estas regras replicaram todos os processos de controlo de qualidade efectuados na base de dados relacional para o DW pois como Haberli & Tombros (2001) referem, os dados devem ser sempre submetidos a controlos de qualidade constantes visto que novas e mais robustas metodologias permitem uma melhor detecção de erros.

Tendo em conta este pressuposto, foi implementado um novo processo de validação dos dados que afectou tanto os dados históricos, previamente recolhidos, como todos os dados que são recolhidos diariamente. Este processo consiste no seguinte:

- i.** Limpeza no momento em que o campo *N*, a localização, e o campo *Data* forem nulos;
- ii.** Limpeza dos dados com o campo *N* e *Data* igual e o campo *ID* diferente;
- iii.** Selecção de um histórico diário de 20 anos;

Estes procedimentos não estão relacionados directamente com os dados meteorológicos mas influenciam a fiabilidade dos mesmos.

### III.3. Data Warehouse

No modelo dimensional *Star Schema*, definido no capítulo anterior, as dimensões e as tabelas de factos são tabelas pertencentes ao DW. Estas tabelas irão ser carregadas com informação proveniente da SA através de uma metodologia simples de carregamento por diferenças. Destaca-se que todos os processos de transformação, limpeza e modificação dos dados são feitos aquando do carregamento dos dados para a SA. Seguidamente, no capítulo respeitante às Dimensões, encontram-se todas as relações entre as tabelas da SA e as respectivas dimensões. Adicionalmente, em anexo encontra-se o diagrama corresponde à modelação dimensional do DW.

### III.4. Extração, Transformação e Carregamento

Com o objectivo de simplificar o processo de modelação no SSIS foram construídos três *Data Transformation Services* (DTSX). Estes DTSX dividem-se em três categorias:

- i. O carregamento inicial, em que é feita a passagem e carregamento dos dados para a SA e posteriormente para o DW, sendo que neste carregamento é feita a passagem de todos os dados históricos.
- ii. O carregamento recursivo da SA, em que os novos *inputs* diários introduzidos na base de dados operacional são transferidos para a SA de acordo com o conjunto de regras de transferência definidos no capítulo anterior.
- iii. O carregamento recursivo do DW, em que os novos *inputs* diários introduzidos na SA são transferidos para o DW.

Desta forma, o modelo de funcionamento diário recai sobre o DTSX 2 e 3, sendo que o primeiro DTSX, o inicial, apenas deverá ser processado uma única vez e como uma solução de instalação do DW. Abaixo é apresentado o diagrama de funcionamento do DTSX 2 e 3:

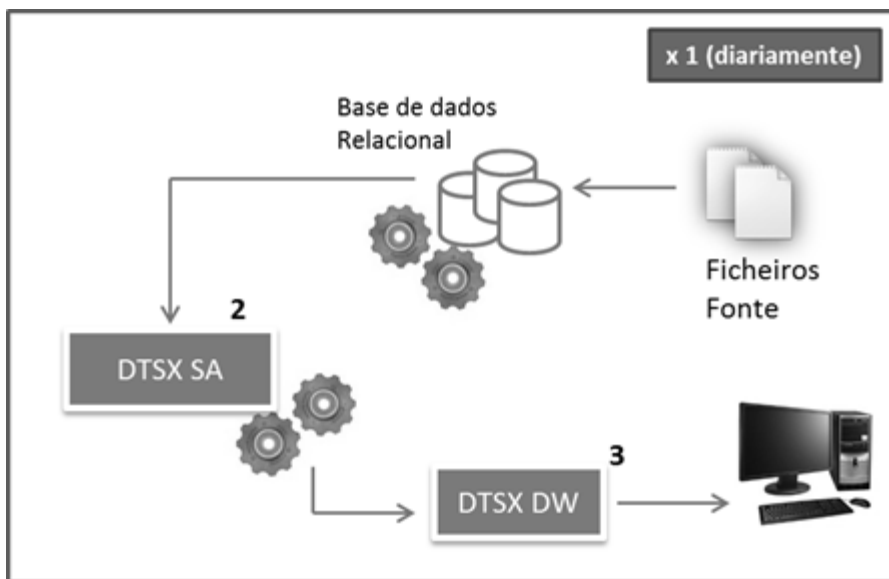


Figura 13 – Diagrama da interactividade entre os *Data Transformation Services* 2 e 3.

Como o DTSX inicial corre apenas uma vez, este necessita de abranger todo o processo ETL de acordo com o diagrama abaixo:

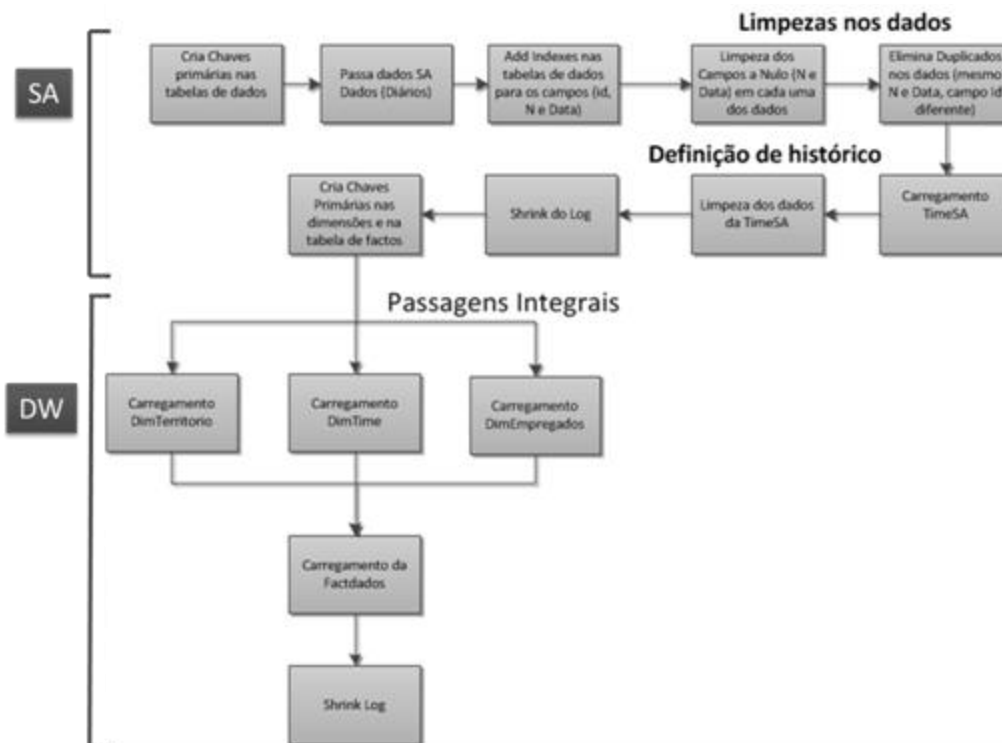


Figura 14 – Diagrama lógico do *Data Transformation Services* inicial.

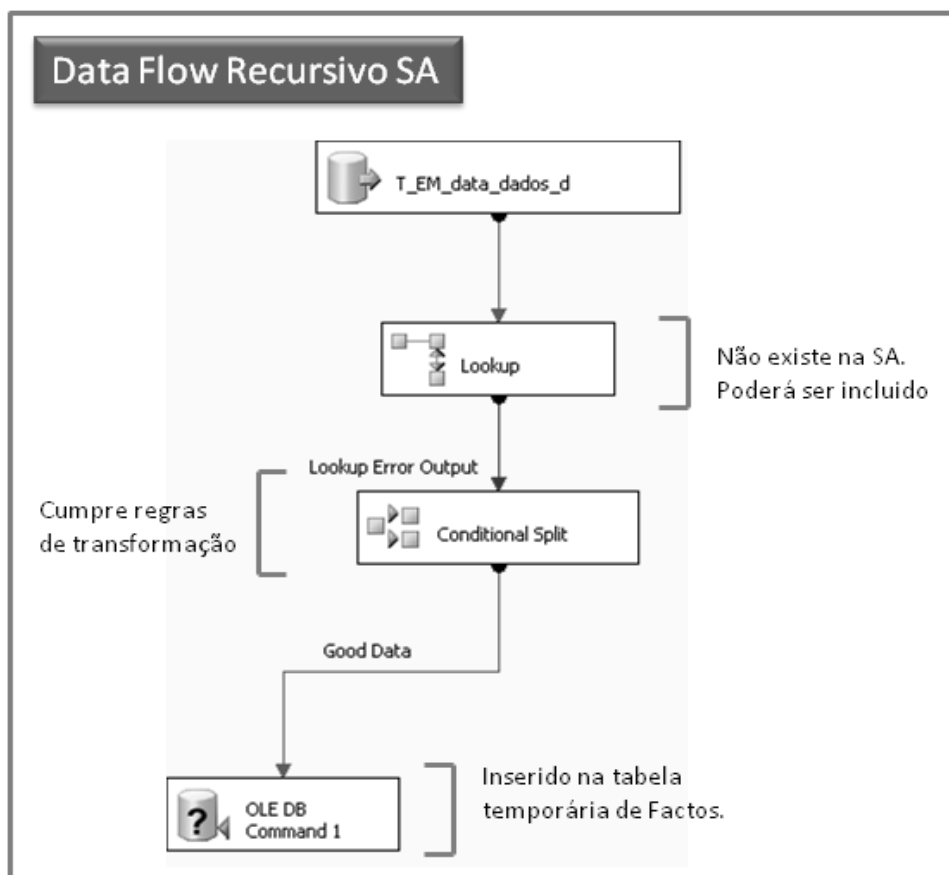
Relativamente aos DTSX 2 e 3, estes processos ETL foram efectuados recorrendo essencialmente a *scripts* que executam *stored procedures* e aos módulos *Data Flow Task* (figura 16). Estes módulos permitem essencialmente efectuar a extracção de uma fonte de dados, executar a sua transformação, limpeza e modificação e, por fim, o seu carregamento. Todo este processamento é efectuado recorrendo às diferentes ferramentas do *SQL Server Integration Services (SSIS)*, nomeadamente ao *Lookup* e ao *Conditional Split*.

Dando um enquadramento técnico, a ferramenta *Lookup* permite comparar a *Data Source* com uma *query* em específico, ou uma tabela, mediante um conjunto de campos seleccionáveis. Por sua vez, caso os dados cruzem, ou não, poderá ser configurada uma acção. Neste caso, a acção a efectuar em caso de sucesso será a de executar a ferramenta *Conditional Split* que permitirá, por sua vez, definir um conjunto de restrições e qual a acção seguinte a executar.



**Figura 15** – Módulo *Data Flow Task* e ferramentas *Conditional Split* e *Lookup*.

Estas ferramentas permitem constituir um processo ETL mais robusto e eficiente, resultando na figura abaixo:



**Figura 16** – Processo ETL da SA.

Conforme se verifica, o erro do procedimento *lookup* resulta nos novos registos a inserir e o módulo *Conditional Split* contém as condições a aplicar. Por sua vez, o comando final insere os registos que cumpriram os requisitos numa tabela de factos temporária da SA, a qual terá aplicado um procedimento SQL de validação das métricas. No caso do processo ETL do DW, e conforme já foi explicitado no subcapítulo anterior (III.3), existe apenas uma passagem da nova informação da SA para o DW de acordo com a figura abaixo:

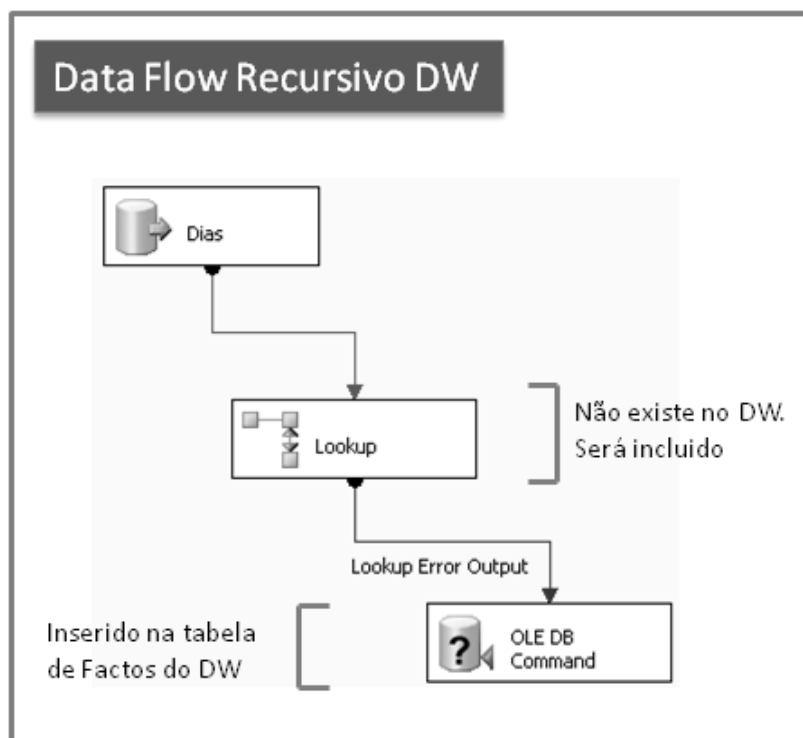


Figura 17 – Processo ETL do DW.

### III.5. Dimensões e Factos

As dimensões de tempo e localização, são as principais dimensões para qualquer DW agrícola (Nilakanta et al., 2008). As dimensões e factos resultantes da análise de negócio foram as seguintes: DimTerritorio, DimEmpregados, DimTempo e FactDados. Alerta-se para o facto de que a DimEmpregados foi apenas construída e não foi incorporada no DW dado que não existe a presente necessidade de analisar os dados meteorológicos para um determinadas EM e saber quais são os seus funcionários.

## DimTerritorio

Esta tabela de dimensão resulta da junção das tabelas T\_EM\_data, T\_EM\_data\_DRA, T\_DRAP, T\_EA, T\_ENTIDADES e T\_EA\_TIPO. Esta dimensão representa os locais de onde é recolhida a informação meteorológica. Esta pode ser recolhida consoante as seguintes tipologias:

- i. Direcções Regionais de Agricultura e Pescas que podem ser públicas ou privadas.

Nome	Tipo	Sigla
Direcção Regional de Agricultura e Pescas do Norte	Pública	DRAPNorte
Direcção Regional de Agricultura e Pescas do Centro	Pública	DRAPCentro
Direcção Regional de Agricultura e Pescas de Lisboa e Vale do Tejo	Pública	DRAPLVT
Direcção Regional de Agricultura e Pescas do Alentejo	Pública	DRAPAL
Direcção Regional de Agricultura e Pescas do Algarve	Pública	DRAPALG
Direcção-Geral de Agricultura e Desenvolvimento Rural	Pública	DGADR
Associação Regional da Agricultura das Terras de Montenegro	Privada	ARATM
Orivárzea - Orizicultores do Ribatejo, S.A.	Privada	Orivárzea
Centro Operativo e Tecnológico Hortofrutícola Nacional	Privada	COTHN
Associação para a Produção e Protecção Integrada de Citrinos	Privada	APPIC

**Tabela 2** – Tipologia das Direcções Regionais de Agricultura e Pescas.

- ii. Estações de avisos simples ou centrais.

Nome	Sigla
Serviço Central Coordenador	SCC
Estação de Avisos Central	EAC
Estação de Avisos	EA
Estação de Avisos Privada	EAP

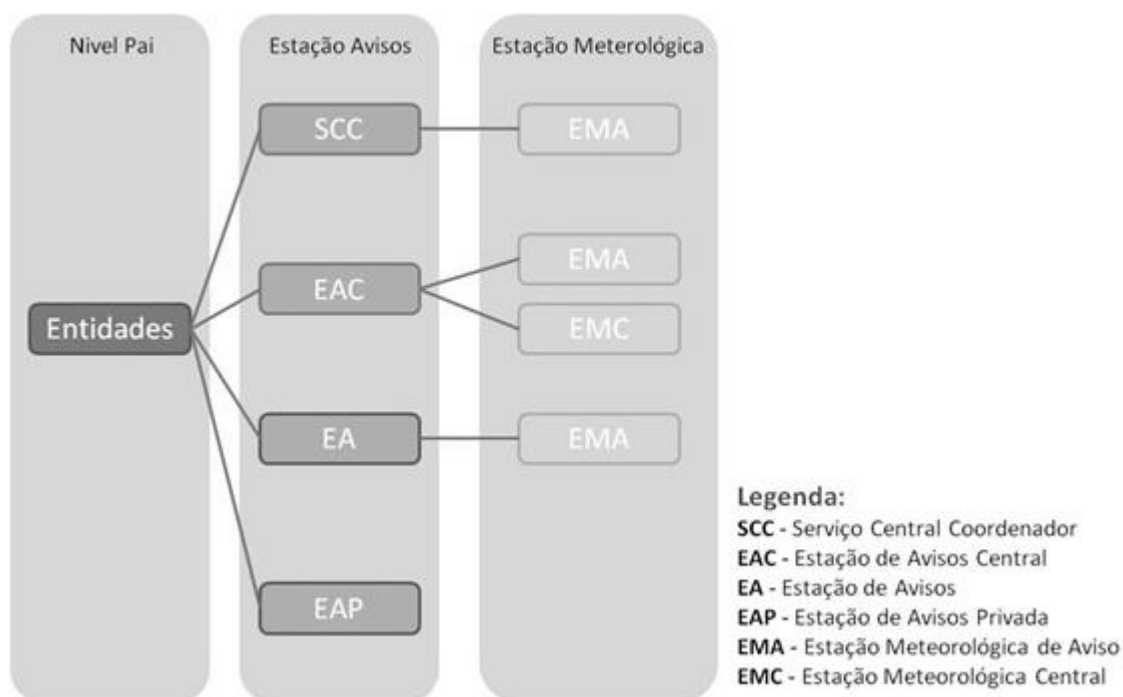
**Tabela 3** – Tipologia das Estações de Avisos.

- iii. Estações meteorológicas de avisos ou centrais

Nome	Sigla
Estação Meteorológica de Aviso	EMA
Estação Meteorológica Central	EMC

**Tabela 4** – Tipologia das Estações Meteorológicas

Abaixo é apresentado o diagrama que demonstra a distinção hierárquica dentro desta dimensão:



**Figura 18** – Diagrama hierárquico Dimensão Território.

## **DimEmpregados**

A tabela DimEmpregados resulta da junção das tabelas T\_RH, T\_RH\_Tipo\_Funcao e T\_RH\_Formacao. Esta dimensão contém todos os empregados responsáveis pelas estações de avisos, assim como a informação relativa à sua posição hierárquica e contratual como empregados da administração central.

## **DimTime**

A tabela DimTime resulta da sintetização de todas as datas, distintas por tipologia das diferentes tabelas de dados, associadas a um ID por posterior inserção. Esta dimensão permite um nível de análise relativamente alargado e composto pelos seguintes atributos:

- i. Tipologia;
- ii. Ano;
- iii. Mês;
- iv. Dia;
- v. Hora;
- vi. Minutos;

As características principais desta dimensão são o código interno (ID\_tempo) e a tipologia. Este campo define qual a periodicidade dos dados, permitindo que no futuro, caso se pretenda analisar mais do que os dados diários exista essa possibilidade de integração. Neste momento este campo detem o valor fixo de *d*.

## **FactDados**

A tabela de factos contém todas as métricas cujas posteriores análises irão incidir e tal como referido por Moody et al. (2000) tem como chave primária o conjunto de chaves primárias das tabelas de dimensões, sendo neste caso a dimensão território e a dimensão tempo. Esta tipologia de chaves primárias é apelidada de *Surrogate Keys*, ou seja, campos inteiros sequenciais que foram carregados aquando da introdução de dados nas tabelas de dimensões e que têm como objectivo a junção das tabelas de dimensões à tabela de factos (Kimbal & Ross, 1996).

### III.6. Métricas

Todas as métricas utilizadas são originárias de dados fonte, ou seja, estes atributos foram recolhidos pelas estações meteorológicas e pelas estações de avisos não tendo sido calculados no decorrer do processo de transformação do ETL. Contudo, o conjunto de atributos recolhidos dos processos actualmente em produção são dispersos e alguns não têm aplicação prática nos modelos implementados. Desta forma, foram seleccionadas as métricas com maior ênfase no contexto negocial. Na Tabela 4 é apresentado o conjunto de métricas e a sua descrição:

<b>Métrica</b>	<b>Unidade</b>	<b>Descrição</b>
TA_Med	C°	Temperatura média
TAx_Max	C°	Temperatura máxima
Tam_Min	C°	Temperatura mínima
HR_Med	%	Humidade relativa média
HRx_Max	%	Humidade relativa máxima
HRm_Min	%	Humidade relativa mínima
hr_sup90	%	Humidade Relativa > 90 %
hr8090	%	Humidade Relativa 80% -90%
hr_inf40	%	Humidade Relativa < 40 %
RG_Int	KJ/m <sup>2</sup>	Radiação global
RG	W/m <sup>2</sup>	Radiação global
VV_kh	k/h	Velocidade do vento média
VV_Med	m/s	Velocidade do vento média
VVx_Max	m/s	Velocidade do vento máxima
DV	°	Média geométrica diária da direcção do vento a 2metros de altura
dv1	°	Média geométrica diária da direcção do vento a 1 metro de altura
PR	Mm	Precipitação
pr_tot	Mm	Precipitação total
ET0	Mm	Evapotranspiração de referência
HFx_max	%	Humidade da folha molhada máxima
HFm_min	%	Humidade da folha molhada mínima
HF_med	%	Humidade da folha molhada média
HF_10	Nh	Número de horas em que a humidade da folha molhada (> 10%)
HF75	Nh	Nº de horas de folha molhada (>75%)
HFTR	Nh	Condição: HF>75 e Tar>15 e R>120
Humec_tot	%	Humidade da folha molhada total

hs_med	%	Humidade do solo média
tr0_med	C°	Temperatura média da relva ou à superfície
tr_med	°C	Temperatura média à superfície (da relva)
trx_max	°C	Temperatura máxima à superfície (da relva)
trm_min	°C	Temperatura mínima à superfície (da relva)
TS0_Med	C°	Temperatura média do solo a 0cm de profundidade
TS0x_Max	C°	Temperatura máxima do solo a 0cm de profundidade
TS0m_Min	C°	Temperatura mínima do solo a 0cm de profundidade
TS5_Med	C°	Temperatura média do solo a 5cm de profundidade
TS5x_Max	C°	Temperatura máxima do solo a 5cm de profundidade
TS5m_Min	C°	Temperatura mínima do solo a 5cm de profundidade
ts10_med	C°	Temperatura média do solo a 10cm de profundidade
ts10x_max	C°	Temperatura máxima do solo a 10cm de profundidade
ts10m_min	C°	Temperatura mínima do solo a 10cm de profundidade
TS15_Med	C°	Temperatura média do solo a 15cm de profundidade
TS15x_Max	C°	Temperatura máxima do solo a 15cm de profundidade
TS15m_Min	C°	Temperatura mínima do solo a 15cm de profundidade

**Tabela 5** – Métricas do *Data Warehouse*.

Estas métricas foram escolhidas pois são utilizadas pelo conjunto de modelos actualmente em produção e à potencialidade de desenvolvimento de novos modelos com base nestas mesmas métricas. Os modelos em questão são os seguintes:

<b>Modelos</b>
EPI Mildio
Estenfiliose
Oídio
Gafa
Olho de Pavão
Bichado da fruta
Aranhão vermelho – Desenvolvimento
Cochonilha de S. José – Desenvolvimento
Cochonilha Vermelha – Desenvolvimento
Traça de Azeitona

**Tabela 6** – Modelos a implementar no *Data Warehouse*.

Desta forma, apenas foi escolhida como fonte de dados a respeitante aos dados diários, pois são os únicos cuja usabilidade é pertinente. Estes dados foram tratados através de vários processos de limpeza conforme descrito acima. As métricas escolhidas são provenientes das fontes de dados simples, no entanto, esta não é a única forma de utilização de métricas.

### III.7. Análise de Resultados

Implementar um DW centralizado, sobre dados agrícolas não é uma tarefa trivial. É essencial um bom planeamento inicial (Nilakanta et al., 2008). Além deste facto, todas as instituições participantes na construção do DW deverão estar cientes dos seus benefícios inerentes à construção do DW. No entanto, para que tal aconteça é necessário permitir aos utilizadores uma manipulação interactiva, dinâmica e fácil dos dados presentes no DW. Esta exploração analítica de um conjunto vasto de dados meteorológicos pode ser suportada por uma correcta aplicação de um DW e de tecnologias OLAP. Um DW disponibiliza, assim, a flexibilidade suficiente a um conjunto grande de dados, enquanto as técnicas OLAP fornecem mecanismos para efectuar análises *ad-hoc* em profundidade (Abdullah & Hussain, 2004). Estas técnicas, apesar de não terem sido desenvolvidas no decorrer do projecto, constituem ferramentas essenciais na rentabilização de um DW.

A agregação de informação em cubos OLAP permite efectuar análises multidimensionais com um elevado grau de autonomia e disponibilidade, através de funcionalidades de *Slice*, *Dice*, *Drill-up* e *Drill-Down*. Estas funcionalidades permitem um detalhe mais pormenorizado das métricas em análise, comercialmente conhecidas por *Key Performance Indicators* (KPI), e resultam em interfaces dinâmicos para o utilizador, tais como *Dashboards*<sup>2</sup> e gráficos. Actualmente, estas funcionalidades já são disponibilizadas por um largo conjunto de casas de software, como por exemplo: Microsoft Analysis Services, IBM Cognos, Oracle Hyperion, Pentaho, Teradata, Microstragey e mais recentemente Excel 2010.

---

<sup>2</sup> Um *dashboard* permite disponibilizar informação de uma forma intuitiva e dinâmica e são semelhantes ao cockpit de um avião ou de um automóvel.

Os mais recentes desenvolvimentos tecnológicos nesta área trouxeram ainda a integração com informação geográfica através de plataformas de dados espaciais (ex. ESRI) e que integram os dados provenientes dos cubos com a localização geográfica (ex: *Google Maps* e *Reporting Services 3.0* da *Microsoft*). Neste caso, torna-se especialmente relevante dado que uma das principais dimensões em análise dos dados meteorológicos é a localização territorial das EM. Por fim, existe ainda a possibilidade de efectuar análises mais profundas aos dados, tentando descobrir tendências futuras nos modelos de prevenção de pragas através do *Data Mining*, que em comparação com as análises de dependência e correlação tratadas por métodos estatísticos, permite encontrar hipóteses mais complexas que incluem, tanto condições numéricas, como lógicas (Pokorny, 2005).

## IV - CONCLUSÕES

**“Knowledge has to be improved, challenged,  
and increased constantly, or it vanishes.”**

*Peter F. Drucker*

Cada vez mais, com o desenvolver de todos os processos e revoluções tecnológicas, a frase proferida por Benjamin Franklin em 1776 “Time is money” se torna intemporal. Tal como Waddington e Hayler (2011) alertam, o factor tempo é cada vez mais um factor relevante nas transacções negociais e que a análise destas devem, e deverão ser aproveitadas para posteriores vantagens competitivas. No caso concreto de um DW para a agrometeorologia e no contexto do SNAA, as vantagens competitivas são delineadas a nível nacional no âmbito das estratégias políticas agrícolas, no entanto, estas tornam-se bastante relevantes no que concerne ao tratamento atempado de pragas com o potencial de arruinar imensidões de terrenos agrícolas.

Tendo em conta estes objectivos, é seguro afirmar que o projecto realizado cumpriu todos os critérios delineados inicialmente. No decorrer deste projecto, a metodologia de implementação informática apenas sofreu pequenas alterações relacionadas com a usabilidade dos dados, não comprometendo o resultado final. Sendo assim, destacam-se os principais objectivos alcançados da realização deste projecto:

- i. Desenho e implementação de uma SA e de DW eficiente e robusto;
- ii. Desenho, desenvolvimento e implementação de um processo de ETL auditável e que permite uma gestão eficiente;

O primeiro e segundo objectivos foram desenvolvidos em simultâneo em razão das várias fases do desenvolvimento informático e ao impacto que cada um deles tinha no outro. É seguro afirmar que ambos os objectivos foram cumpridos na totalidade, salvaguardando que, como em qualquer projecto informático, existe sempre um conjunto de procedimentos que podem ser efectuados como forma de complementar os sistemas actualmente em vigor, ou de os actualizar para uma versão mais eficiente. Neste caso, tendo-se em conta os rápidos desenvolvimentos no sector tecnológico, ambas as hipóteses são igualmente exequíveis. Fazendo uma divisão de âmbitos, os objectivos acima enumerados poderão ser considerados como os objectivos tangíveis,

pois são aqueles que são mensuráveis e que se qualificam como a base do projecto. Agregados a estes objectivos existem ainda aqueles que representam o potencial de desenvolvimento de novos projectos com base no DW implementado, e que se classificam como objectivos intangíveis alcançados.

Com base no projecto efectuado, existe um alto potencial de desenvolvimento de modelos analíticos, com vista à análise dos dados recolhidos e já tratados, correspondendo estes às mais recentes tecnologias de tratamento de informação. Foram também garantidos os mais recentes mecanismos de indexação e de interligação de tabelas, de modo a permitir uma eficiente utilização por parte dos utilizadores finais. Este projecto lança as bases tecnológicas que irão permitir, em suma, o trabalho que é proposto no subcapítulo IV.3 - *Trabalho Futuro*.

#### IV.1. Limitações

As limitações do projecto prendem-se essencialmente nas fundações tecnológicas necessárias para a realização do mesmo. No entanto, existem limitações gerais impostas pela realização de projectos informáticos.

Neste trabalho de projecto foi usada a tecnologia Microsoft mediante a utilização do software de gestão de base de dados *Microsoft SQL Server* e das componentes inseridas na sua *framework*. Utilizou-se a versão de 2005 e a componente *SQL Server Integration Services (SSIS)* conjuntamente com o *SQL Management Studio* (ferramenta de gestão de bases de dados). As limitações tecnológicas resultantes da utilização deste *software* provêm da não adaptabilidade do modelo de tratamento de dados a outras bases de dados que não a inicialmente desenvolvida (Mundi et al, 2006; Kimbal & Ross, 1996). Este factor resulta, assim, na inexistência da possibilidade de utilização do mesmo modelo para outro projecto de extracção ou construção de um DW semelhante (Ong, 1999; Stonebraker, Bear & Cetintemel, 2007).

Previamente à construção do modelo informático, procedeu-se à construção do modelo dimensional. Na construção da estrutura organizacional de um DW foram encontradas as seguintes limitações:

- i. O espaço de armazenamento. O espaço que teve de ser reservado à SA e ao DW foi semelhante ao que estava reservado para a base de dados relacional o que implicou a triplicação do volume de dados a reter em disco;
- ii. A ocupação de recursos de sistema originados pelo carregamento do DW. Esta ocupação dos recursos é oriunda da necessidade de carregamento do DW de tempo a tempo, sendo que neste caso o carregamento é diário tendo em conta a volumetria de dados produzida diariamente;
- iii. O facto de permitir pouca volatilidade da estrutura de dados e da metadata;

Adicionalmente, existem ainda duas questões que são susceptíveis de criar algumas limitações, nomeadamente, o número de utilizadores a aceder ao servidor e a segurança dos dados. Ambos os componentes são vistos como limitações, pois como

Rosenthal e Sciore (2000) afirmam existe a presente dificuldade de gerir os sistemas de segurança de um DW com coerência, sendo que as permissões atribuídas deverão satisfazer os requisitos de negócio e ser actualizadas consoante os outros sistemas tecnológicos da organização.

## **IV.2. Propostas de Trabalho Futuro**

O presente projecto encontra-se actualmente em fase de implementação da infra-estrutura tecnológica, tendo como objectivo a instalação do modelo dimensional no servidor de produção, assim como do processo ETL. Este terá invariavelmente de ser adaptado à base de dados relacional actual, visto que as bases de dados utilizadas para a construção do modelo actual são de teste e por conseguinte possuem uma nomenclatura diferente.

Além dos procedimentos de implementação, no decorrer de todo o projecto foram encontrados alguns procedimentos funcionais com algum potencial de melhoria e que não se prendem com o âmbito deste projecto:

### **Processamentos técnicos**

- i.** Melhoria do processo de carregamento diário dos dados provenientes das estações meteorológicas, nomeadamente a redução do tempo de passagem dos dados e com a limpeza das tabelas intermédias de carregamento;
- ii.** Alterações na base de dados relacional. Estas alterações deverão ser feitas ao nível do número de tabelas, número de *views* e de *stored procedures* em excesso e não utilizáveis;
- iii.** A capacidade de manutenção adequada à escalabilidade dos dados, tendo em conta o grau de crescimento inerente ao DW. Neste caso, ainda não foram avaliados os graus de crescimento dos dados, no entanto, deverão ser implementados mecanismos de manutenção e de criação de histórico para armazenamento de dados mais antigos com base nos requisitos dos utilizadores finais;

- iv. Implementação de uma ferramenta de *reporting* para os futuros desenvolvimento a nível de modelação analítica – OLAP, com vista a desempenhar os objectivos descritos no subcapítulo III.7 – Análise de Resultados;

Tal como é referido por Gardner (1998), a manutenção e melhoria de um DW coordena um conjunto de tarefas operacionais que compreendem a disponibilidade e a performance do DW, a expansão da solução para DM's e para outras áreas de negócio, assim como para o suporte de outras aplicações de negócio. Isto permite que exista um desenvolvimento parcial na infra-estrutura do DW alargando o seu âmbito por exemplo ao tratamento e disponibilização dos dados horários.

### **Website**

Na plataforma interna, desde a autenticação por parte dos utilizadores, indo à validação de dados meteorológicos, sugere-se que as diferentes validações temporais sejam as seguintes:

- i. Validação diária ir até 30 dias atrás;
- ii. Validação horária ir até 5 dias atrás;
- iii. Validação de 30 minutos ir até 2 dias atrás;
- iv. Validação de 15 minutos ir até 2 dias atrás;

Adicionalmente, existe agora a oportunidade de basear o website do SNAA, em especial toda a parte de apresentação de modelos de simulação de existência de pragas na componente do DW.

Nilakanta et al., (2008) referem, correctamente, que todas as instituições que retirem vantagens da construção de um DW deverão estar cientes dos benefícios inerentes à implementação do mesmo. Aliado a este facto, existe a necessidade de avaliação da solução de qualidade dos dados implementada, tendo em conta que a qualidade dos dados não poderá ser medida sem que, na sua génese total, esta seja submetida a aprovação por parte dos utilizadores finais (Strong, Lee & Wang, 1997). Como tal, torna-se relevante a avaliação do grau de satisfação dos utilizadores finais,

relativamente ao DW implementado. Esta medição do grau de satisfação poderá ser medida mediante a metodologia abordada por Chen, Soliman, Mao & Frolick (2000), na qual são tidos em conta a qualidade dos dados, o suporte aos utilizadores, o formato da informação e a sua precisão. O autor sugere assim que a recolha da informação seja efectuada no *website* do SNAA visto que este é o ponto de contacto mais abrangente com os seus utilizadores.

## **Gerais**

Tendo em conta os propósitos espelhados na missão e objectivos do SNAA e concentrando este organismo toda a informação e divulgação relativa a avisos agrícolas, propõe-se que sejam criadas sinergias e se conjuguem esforços no sentido de evitar que tanto o SNAA, organismo criado para este efeito, como cada uma das diferentes Direcções Regionais de Agricultura e Pescas (DRAP) procedam à divulgação de avisos de teor semelhante. Destaca-se ainda que se tal situação não for regularizada e os âmbitos de cada organismo não forem acertados, mesmo presumindo a lateralização de competências, existe a forte possibilidade de divulgação de informação sobre a mesma temática com conteúdo diferente.

## V - REFERÊNCIAS BIBLIOGRÁFICAS

Abdullah, A., Brobst, S., Umer, M., Khan, M. F. (2004). *The Case for an Agri Data Warehouse: Enabling Analytical Exploration of Integrated Agricultural Data*. In: Proceedings of the IASTED international Conference on Databases and Applications. Innsbruck, Austria: 139-144

Abdullah, A., Hussain, A. (2006). *Data Mining a New Pilot Agriculture Extension Data Warehouse*. Journal of Research and Practice in Information Technology. Vol. 38. No.3: 229-249

Aguiar, A., Godinho, M.C., Costa, C.A. . (2005). *Produção Integrada*. Porto: Principia, Publicações Universitárias e Científicas.

AJAP. *A Importância da Protecção e Produção Integrada na assistência técnica aos agricultores*. Retrieved 17-12-2010, from [http://www.ajap.pt/sippi/recursos/docs/pi\\_rev67\\_importancia\\_pi.pdf](http://www.ajap.pt/sippi/recursos/docs/pi_rev67_importancia_pi.pdf)

Ali, A. (2009). "SQL server Integration Services SSIS Best Practices." Retrieved 22-09-2011, from [www.mssqltips.com/tip.asp?tip=1840](http://www.mssqltips.com/tip.asp?tip=1840).

Amaro, P. Protecção Integrada ou “Boa” Prática Fitossanitária. Retrieved 10-11-2010, from <http://www.agroportal.pt/a/pamaro.htm>.

ANIPLA. "Associação Nacional da Indústria para a Protecção das Plantas." Retrieved 22-09-2011, from <http://www.anipla.com/>.

Ballou, D. P., Tayi, G. K. (1999). *Enhancing Data Quality in Data Warehouse Environments*. Communications of the ACM. Vol.42, No.1: 73-78.

Barbieri, C. (2011). *BI – Business Intelligence – Modelagem & Tecnologia*. Excel Books do Brasil Editora. Rio de Janeiro, Brasil.

Caldeira, C. P. (2008). *Data Warehousing - Conceitos e Modelos* (1st ed.). Lisboa: Edições Sílado, Lda.

Chen, L. D., Soliman, K. S., Mao, E., Frolick, M.N. (2000). *Measuring user satisfaction with data warehouses: an exploratory study*. Info Manage. 37: 103-110.

DGADR. Organização e competências. Retrieved 10-11-2010, from <http://www.dgadr.pt/>.

Dias, Â. (2003). Protecção Integrada - Boletim Informativo. Retrieved 17-12-2010, from <http://www.cna.pt/inicio.htm>.

Duffy, B. (2008). *SSIS 2008 Performance Best Practices*. In: Irish SQL Academy. Level 300.

EISA. European Initiative for sustainable development in agriculture. Retrieved 12-12-2010, from <http://www.sustainable-agriculture.org/news.php>.

Eischeid, J. K., Baker, C. B., Karl, T. R., Diaz. (1995). *The Quality Control of Long-Term Climatological Data Using Objective Data Analysis*. Journal of applied meteorology: 2787 – 2795.

ETL-Tools.info. *Data warehouse*. Retrieved 16-12-2010, from [http://etl-tools.info/en/bi/datawarehouse\\_concepts.htm](http://etl-tools.info/en/bi/datawarehouse_concepts.htm).

Gardner, S. (1998) *Building the Data Warehouse*. Communications of the ACM Vol41, No.9. 53-60.

Han, J., Kamber, M. (2001). *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers.

IFAP. Medidas Agro-Ambientais. Retrieved 10-11-2010, from <http://www.inga.min-agricultura.pt/ajudas/agroamb.html>.

Inmon, H. W. (1992). *Building the Data Warehouse*: QED Technical Publishing Group.

IOBC-WPRS. International Organization for Biological and Integrated Control of Noxious Animals and Plants, West Palaearctic Regional Section. Retrieved 12-12-2010, from <http://www.iobc-wprs.org/>.

Jones, J. W., Hansen, J. W., Royce, F. S., Messina, C. D. (2000). *Potential benefits of climate forecasting to agriculture*. *Agriculture, Ecosystems and Environment* 82: 169-184.

Kejser, T., Lee, D. (2008) "Top 10 SQL Server Integration Services Best Practices". Retrived: 22-10-2011, from: <http://sqlcat.com>.

Kimball, R., Ross, M. (1996). *The Data Warehouse Toolkit*. John Wiley & Sons, Inc. Indianapolis, Indiana: Wiley Computer Publishing, Inc.

Ladeira, M., F. Projovem - modo de produção biológico. Retrieved 17-12-2010, from [http://projovem.drapc.min.agricultura.pt/base/documentos/agricultura\\_biologica.htm](http://projovem.drapc.min.agricultura.pt/base/documentos/agricultura_biologica.htm).

Lenzerini, M. (2002). *Data Integration: a theoretical perspective*. In: Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. Madison, Wisconsin: 233-246.

Marco, D. (2000). *Building and Managing the Meta Data Repository: A Full Lifecycle Guide*. EUA: John Wiley & Sons, Inc.

Martino, L., Santis, A.D. (2001). *Improvement of quality in statistics: an integrated system for managing surveys*. International Conference on Quality in Official Statistics. Estocolmo, Suécia: 1-31.

Microsoft, “Tuning Your SSIS Package Data Flow in the Enterprise (SQL Server Video)”. Retrieved: 22-10-2011, from: <http://msdn.microsoft.com/en-us/library>.

Moody, D. L., Kortink, Mark A.R. (2000). From Enterprise Models to Dimensional Models: Methodology for Data Warehouse and Data Mart Design. In: Proceedings of the International Workshop on Design and Management of Data Warehouses. Estocolmo, Suécia: 1-12.

Mundi, J., Thorntwaite, W. and Kimball, R. (2006). *The Microsoft Data Warehouse Toolkit*. Indianapolis, Indiana: Wiley Publishing, Inc.

Neto, M. C., Fernandes, L. M. (2007). *i-Farm: A exploração agrícola inteligente da sociedade da informação e do conhecimento*. Actas do I Congresso Ibérico IV Congresso Espanhol de AgroEngenharia. Albacete, Espanha: 4-6.

Neto, M.C., Pinheiro, A. C., Cavaco, M. e Mendes, F. (2007). *Bringing the Portuguese Agricultural Warning Services to a new Media*. In: Proceedings of the Joint Conference - The 6th Conference of the European Federation for Information Technology in Agriculture, Food and Environment and The 4rd World Congress on Computers in Agriculture and Natural Resources. Glasgow, Scotland: 2-5.

Nilakanta, S., Scheibe, K., Rai, A. (2007). *Dimensional issues in agricultural data warehouse designs*. Computers and Electronics in Agriculture, 60: 263-278.

Ong, H. (1999). *The Evolution Of A Data Warehouse Architecture - One Size Fits All ?*. Oracle Open World. Singapura: 1-7.

Pinet, F., Miralles, A., Bimonte, S., Vernier, F., Carluer, N., Gouy, V., Bernard, S. (2010). *The use of UML to design agricultural data warehouses*. International Conference on Agricultural Engineering: 1-10.

Pokorny, J. (2005). *Database architectures: current trends and their relationships to environmental data management*. In: Proceedings of the 19th Conference Informatics for Environmental Protection. Brno, República Checa: 24-28.

Poole, J., Chang, D., Tolbert, D., Mellor, D. (2002). *Common Warehouse Metamodel*. John Wiley & Sons, Inc. New York: Wiley Computer Publishing, Inc. 37

Porter, M. E., Millar, V. E. (1985). *How Information Gives you competitive advantage*. Harvard Business Review, Julho-Agosto: 149-160

Rainardi, V. (2008). *Building a Data Warehouse: With Examples in SQL Server*. New York: Apress.

Rosenthal, A., Sciore, E. (2000). *View Security as the Basis for Data Warehouse Security*. Proceedings of the International Workshop on Design and Management of Data Warehouses. Estocolmo, Suécia: 1-8.

Salvabati, M., Brambilla, E. (2008). *Data quality control procedures in Alpine meteorological services*. Università degli Studi di Trento. Trento, Itália: 1-28.

Stonebraker, M., Bear, C., Cetintemel, U. (2007). *“One Size Fits All”: An Idea Whose Time Has Come and Gone*. CIDR'07.USA: 1-10.

Strong, D. M., Lee, Y. W., Wang, R. Y (1997). *Data Quality in Context*. Communications of the ACM Vol. 40, No.5: 103-110.

Swensson, C., Sederblad, B. (1997). *Data warehouse – A new tool in extension service to Swedish Milkproducers*. First European Conference for Information Technology in Agriculture. Copenhagen, Dinamarca: 15-18.

SNAA. Avisos Fitossanitários. Retrieved 10-11-2010, from <http://snaa.dgadr.pt/>.

Thompson, J. (2006) “SSIS: Suggested Best Practices and naming conventions”. Retrived: 22-10-2011, from: <http://consultingblogs.emc.com/jamiethomson>.

Todman, C. (2001). *Designing a Data Warehouse*. New Jersey: Prentice Hall PTR.

Waddington, D., Hayler, A. (2011). *Time is Money*. The Information Difference Company Ltd. 4

Wang, N., Zhang, N., Wang, M. (2006). *Wireless sensors in agriculture and food industry – Recent development and future perspective*. Computers and Electronics in Agriculture 50: 1-14.

Weiss, A., Crowder, L. V., Bernardi, M. (1999). *Communicating agrometeorological information*. WMO – World Meteorological Organization. Vol. 48, No.4: 368-374.

Yost, M., Nealon, J. (1999). *Using a dimensional data warehouse to standardize survey and census metadata*. National Agricultural Statistics Service, U.S. Department of Agriculture, Fall: 1-6.

Zahumensky, I. (2004). *Guidelines on Quality Control Procedures for Data from Automatic Weather Stations*. World Meteorological Organization, Commission for basic Systems, Expert team on requirements for data from automatic weather stations. Third Session. Geneva, Suíça: 1-9.

## VI - ANEXOS

### VI.1. Anexo A - Cronograma

Para as actividades incluídas neste projecto previu-se inicialmente uma duração total de 7 meses, de acordo com o seguinte planeamento:

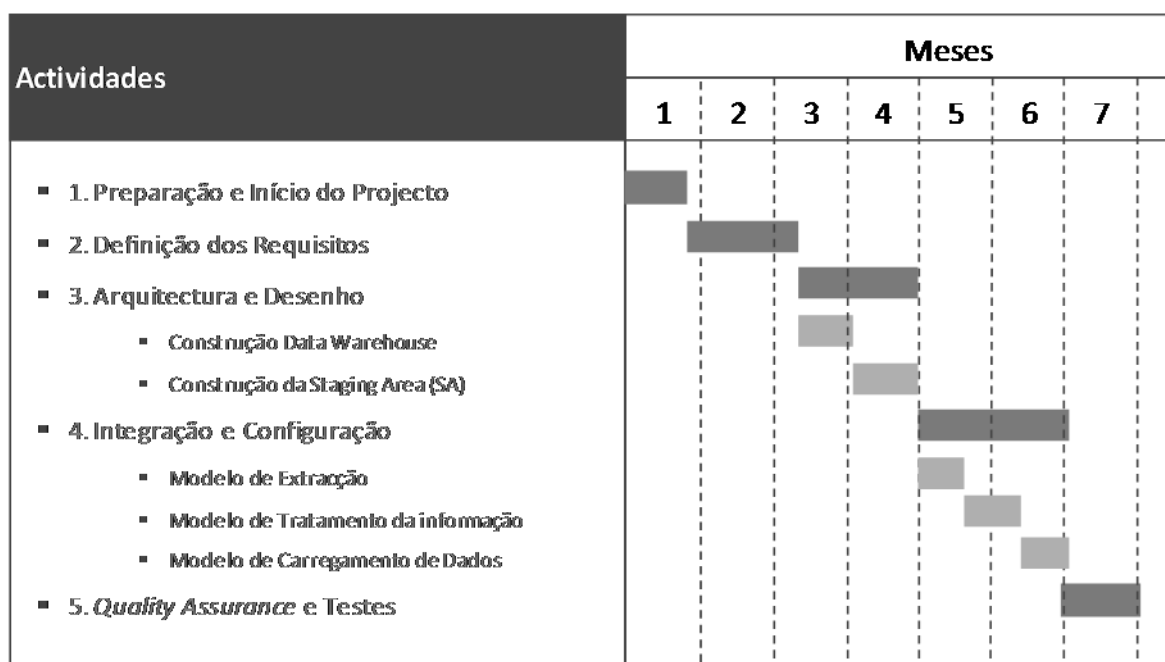


Figura 19 – Mapa de planeamento de actividades gerais do projecto.

O objectivo foi o de iniciar o projecto e o seu relatório em Janeiro e o de o finalizar em Julho, ficando o mês de Agosto reservado para ajustes de documentação final, eventuais testes *ad-hoc* ao modelo ETL e para a entrega e fecho do projecto, ou seja, a entrega da tese de Mestrado. No entanto, existiram derrapagens face á previsão inicial, que é demonstrada no Mapa de Gantt abaixo (figura 21), e o encerramento do projecto só foi possível de efectuar no decorrer do mês de Outubro. O mês de Novembro ficou assim reservado para todas as actividades previstas inicialmente para o mês de Agosto.

ID		Task Name	Duration	Start	Finish
1		Preparação e Início de Projecto	12 days	Mon 10-01-11	Tue 25-01-11
2		Definição de Requisitos	28 days	Wed 26-01-11	Fri 04-03-11
3		<b>Arquitectura e Desenho</b>	<b>41 days</b>	<b>Mon 07-03-11</b>	<b>Sat 30-04-11</b>
4		<b>Construção do DW</b>	<b>41 days</b>	<b>Mon 07-03-11</b>	<b>Sat 30-04-11</b>
5		Construção Data Warehouse	21 days	Mon 07-03-11	Mon 04-04-11
6		Construção da Staging Area (SA)	20 days	Tue 05-04-11	Sat 30-04-11
7		<b>Integração e Configuração</b>	<b>43 days</b>	<b>Tue 03-05-11</b>	<b>Thu 30-06-11</b>
8		<b>Construção do Modelo ETL</b>	<b>43 days</b>	<b>Tue 03-05-11</b>	<b>Thu 30-06-11</b>
9		Modelo de Extração	13 days	Tue 03-05-11	Thu 19-05-11
10		Modelo de Tratamento da informação	16 days	Fri 20-05-11	Fri 10-06-11
11		Modelo de Carregamento de Dados	14 days	Mon 13-06-11	Thu 30-06-11
12		Quality Assurance e Testes	21 days	Fri 01-07-11	Sun 31-07-11
13		Preparação de Documentação Final	10 days	Mon 01-08-11	Sun 14-08-11
14		Entrega e Fecho do Projecto	13 days	Mon 15-08-11	Wed 31-08-11

**Figura 20** – Relatório do Microsoft Office Project – Mapa de Gantt.

## VI.2. Anexo B - Diagramas

### Diagrama Staging Area

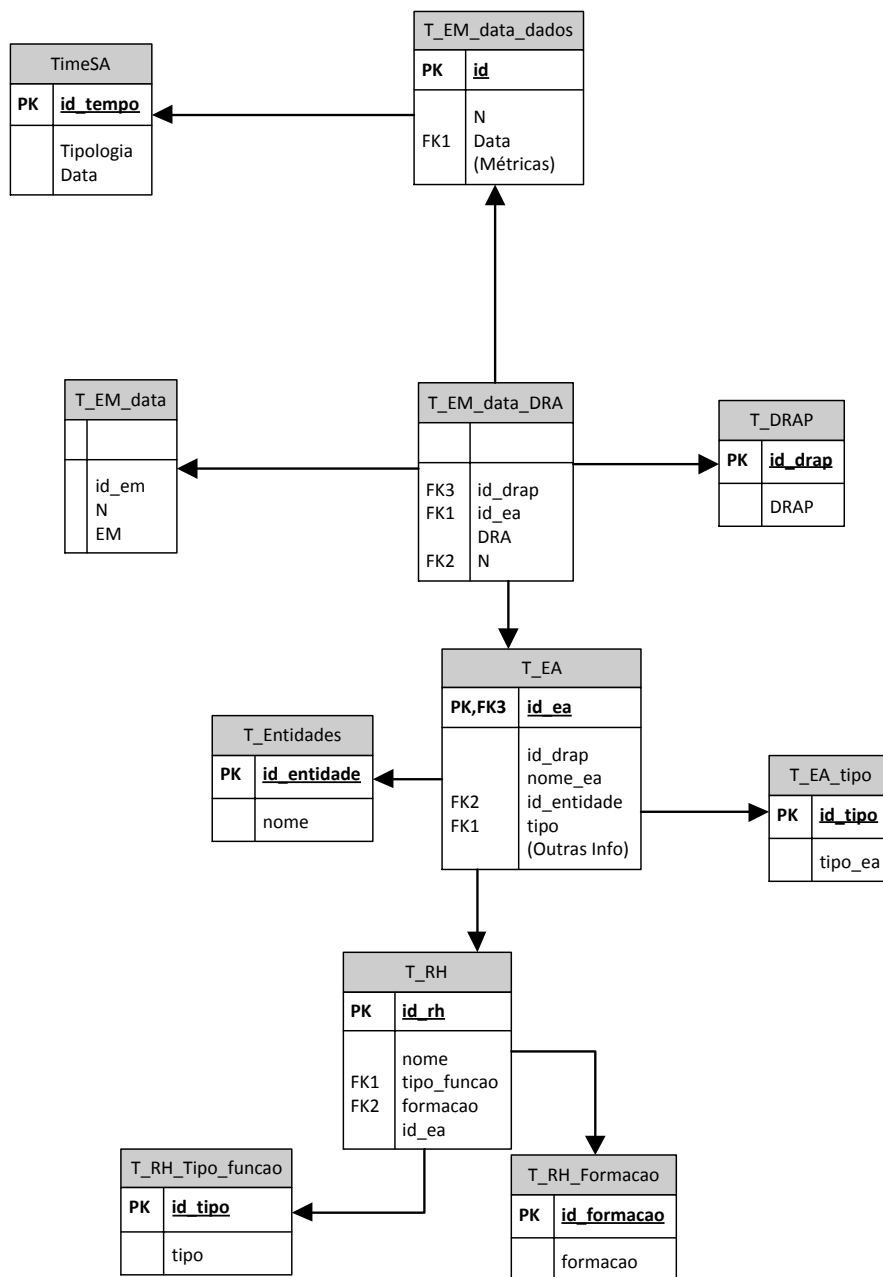


Figura 21 – Diagrama da Staging Area.

## Diagrama Data Warehouse

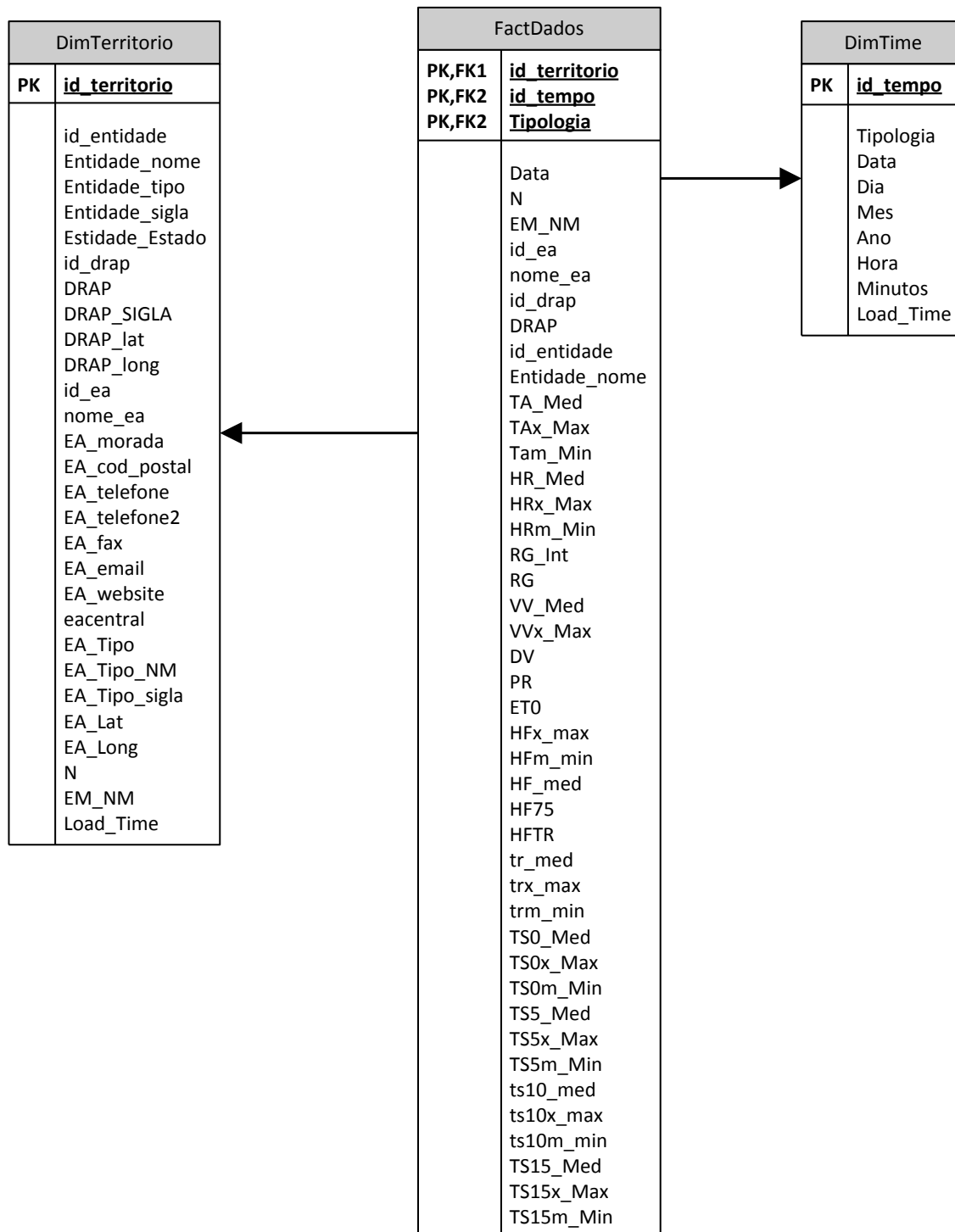


Figura 22 – Tabela do Data Warehouse.

## Dimensão Território

DimTerritorio	
PK	<u>id_territorio</u>
	id_entidade
	Entidade_nome
	Entidade_tipo
	Entidade_sigla
	Estidade_estado
	id_drap
	DRAP
	DRAP_SIGLA
	DRAP_lat
	DRAP_long
	id_ea
	nome_ea
	EA_morada
	EA_cod_postal
	EA_telefone
	EA_telefone2
	EA_fax
	EA_email
	EA_website
	eacentral
	EA_Tipo
	EA_Tipo_NM
	EA_Tipo_sigla
	EA_Lat
	EA_Long
	N
	EM_NM
	Load_Time

**Figura 23** – Tabela da dimensão território do *Data Warehouse* (DimTerritorio).

## Dimensão Empregados

DimEmpregados	
PK	<u>id_empregado</u>
	id_rh nome Interno id_tipo_funcao Hierarquia id_formacao formacao morada cod_postal email telefone foto id_ea estado nivel username password Load_Time

**Figura 24** – Tabela da dimensão empregados do *Data Warehouse* (DimEmpregados).

## Dimensão Tempo

DimTime	
PK	<u>id_tempo</u>
	Tipologia Data Dia Mes Ano Hora Minutos Load_Time

**Figura 25** – Tabela da dimensão Tempo do *Data Warehouse* (DimTime).

## Factos Dados

FactDados	
PK	<u>id_territorio</u>
PK	<u>id_tempo</u>
PK	<u>Tipologia</u>
	Data N EM_NM id_ea nome_ea id_drap DRAP id_entidade Entidade_nome TA_Med TAx_Max Tam_Min HR_Med HRx_Max HRm_Min RG_Int RG VV_Med VVx_Max DV PR ET0 HFx_max HFm_min HF_med HF75 HFTR tr_med trx_max trm_min TS0_Med TS0x_Max TS0m_Min TS5_Med TS5x_Max TS5m_Min ts10_med ts10x_max ts10m_min TS15_Med TS15x_Max TS15m_Min

**Figura 26** – Tabela de factos do *Data Warehouse* (FactDados).