

**DEPARTAMENTO DE INFORMÁTICA**

**Mestrado em Engenharia Informática**

Dissertação

Sistema de Referenciação Geográfica

**João Pedro Pereira Perdigão nº 19011**

**ORIENTADOR:** Luis Arriaga da Cunha  
**COORDENADOR EXTERNO:** Jorge Basílio

**DEPARTAMENTO DE INFORMÁTICA**

**Mestrado em Engenharia Informática**

Dissertação

Sistema de Referenciação Geográfica



177956

**João Pedro Pereira Perdigão nº 19011**

**ORIENTADOR:** Luis Arriaga da Cunha  
**COORDENADOR EXTERNO:** Jorge Basílio

## Agradecimentos

Primeiramente gostaria de agradecer a toda a minha família, amigos e namorada o apoio e incentivo que me prestaram e ao facto de estarem sempre presentes, especialmente aos meus pais porque sem eles esta oportunidade nunca teria surgido.

Agradeço também ao professor Luis Arriaga da Cunha, da Universidade de Évora, pela orientação que me facultou ao longo destes meses durante o projecto em que estive envolvido. Os seus conselhos e indicações foram determinantes.

Mais um agradecimento especial ao Dr. Jorge Basílio pela forma como me integrou na equipa, pelo apoio e pela oportunidade de poder participar na realização deste projecto.

Para finalizar a todos os meus colegas que me acompanharam ao longo da vida académica, em particular a Edgar Marino, André Teixeira, Valter Carvalho, Francisco Matias e Pedro Guerreiro.

## **Sistema de Referência Geográfica – A sua influência nas tomadas de decisão de uma empresa de telecomunicações**

Nos dias de hoje o acesso à informação por parte das empresas é vital para o bom desempenho das suas funções. As empresas de telecomunicações não fogem à regra, a sua posição no mercado está dependente das decisões que são tomadas com base na avaliação dessa informação.

Para suportar os processos de apoio à decisão é coerente recorrer-se a Data Warehouses que permitem integrar informação de diversas fontes, verificando a sua qualidade, actualização e coerência, organizando-a para um fácil acesso e consulta de vários pontos de vista.

Numa empresa de telecomunicações móvel, um Data Mart geográfico baseado na informação de tráfego da companhia que pode identificar as localizações preferenciais dos utilizadores na rede é muito importante porque fornece indicadores muito úteis para o departamento de marketing e negócio da empresa de maneira a que se saiba onde e como actuar para permitir que esta se desenvolva e ganhe vantagem no mercado.

## **Geographic Referencing System - The influence in a telecommunications company decision-making**

Today the access to information by enterprises is vital for the companies performance. Telecommunications companies are no exception. Their position in the market is dependent on the decisions that are taken based on the evaluation of such information.

To support the decision making process a Data Warehouse is today an extremely useful tool; it integrates information from different sources, checking on its validity, quality and update, coherence, organizing it for an easy access and search from various perspectives.

In a mobile telecommunications company a geographical Data Mart-based traffic information that can identify the preferential locations of users on the network is very important. It provides useful indicators to the Department of Marketing and Business there by allowing you to know where and how to act and boosting the development of the company.

# Índice

<b>1. Apresentação</b> .....	<b>1</b>
1.1. Contexto Académico .....	2
1.2. Contexto Científico .....	2
1.3. Contexto Institucional.....	3
1.4. Objectivos da Dissertação.....	3
<b>2. Enquadramento do Projecto</b> .....	<b>4</b>
2.1. Business Intelligence.....	6
2.1.1. Conceito e enquadramento histórico.....	6
2.1.2. Business Intelligence nas empresas de Telecomunicações.....	6
2.1.3. Arquitectura de um sistema de Business Intelligence.....	7
2.2. Data Warehouse .....	7
2.2.1. Enquadramento e motivação.....	7
2.2.2. Definição de Data Warehouse.....	8
2.2.3. Arquitectura de um Data Warehouse.....	9
2.2.3.1. <i>Arquitectura tradicional</i> .....	9
2.2.3.2. <i>Arquitectura em duas camadas</i> .....	10
2.2.3.3. <i>Arquitectura em três camadas</i> .....	10
2.2.3.4. <i>Arquitectura com uma base dados integrada</i> .....	10
2.2.4. Modelo de dados de um Data Warehouse .....	11
2.2.4.1. <i>Modelo Estrela</i> .....	11
2.2.4.2. <i>Modelo Floco de Neve</i> .....	12
2.2.5. Processos Básicos .....	12
2.2.5.1. <i>Extracção</i> .....	13
2.2.5.2. <i>Transformação</i> .....	13
2.2.5.3. <i>Carregamento e Indexação</i> .....	13

<u>2.2.5.4. Garantia de Qualidade</u> .....	13
<u>2.2.5.5. Publicação</u> .....	13
<u>2.2.5.6. Consulta/Atualização</u> .....	14
<u>2.2.5.7. Segurança</u> .....	14
<u>2.2.5.8. Recuperação e Backup</u> .....	14
<b>2.3. Geo-Referenciação</b> .....	<b>14</b>
<b>2.3.1. Conceito</b> .....	<b>14</b>
<b>2.3.2. Sistema de Informação nas Empresas de Telecomunicações</b> .....	<b>15</b>
<b>2.3.3. Sistema de Informação Geográfica</b> .....	<b>15</b>
<b>2.3.4. Location Intelligence</b> .....	<b>16</b>
<b>2.3.5. Geomarketing</b> .....	<b>17</b>
<b>2.4 Outros Conceitos</b> .....	<b>18</b>
<b>2.4.1. Data Mart</b> .....	<b>18</b>
<u>2.4.1.1. Data Mart versus Data Warehouse</u> .....	18
<u>2.4.1.2. Definições de Data Mart Geográfico</u> .....	18
<b>2.4.2. ETL</b> .....	<b>21</b>
<b>2.4.3. OLAP</b> .....	<b>21</b>
<u>2.4.3.1. OLAP vs OLTP</u> .....	22
<b>2.4.4. ODS</b> .....	<b>22</b>
<b>3. Metodologias Utilizadas</b> .....	<b>24</b>
<b>3.1. Ferramentas</b> .....	<b>25</b>
<b>3.1.1. DataStage</b> .....	<b>25</b>
<b>3.1.2. Toad</b> .....	<b>26</b>
<b>3.1.3. ZOC</b> .....	<b>27</b>
<b>3.1.4. FileZilla</b> .....	<b>28</b>
<b>3.1.5. ArcGIS</b> .....	<b>29</b>
<u>3.1.5.1. ArcView</u> .....	29
<u>3.1.5.2. ArcEditor</u> .....	30
<u>3.1.5.3. ArcInfo</u> .....	30

<b>3.2. Linguagens .....</b>	<b>31</b>
<b>3.2.1. PL-SQL.....</b>	<b>31</b>
<b>3.2.2. Shell Script.....</b>	<b>32</b>
<b>4. Descrição do Desenvolvimento .....</b>	<b>34</b>
<b>4.1. Introdução.....</b>	<b>35</b>
<b>4.2. Requisitos.....</b>	<b>35</b>
<b>4.2.1. Requisitos de Negócio.....</b>	<b>35</b>
<u>4.2.1.1. Universo de Informação.....</u>	<u>35</u>
<u>4.2.1.2. Universo de Espacial .....</u>	<u>36</u>
<u>4.2.1.3. Regras .....</u>	<u>36</u>
<u>4.2.1.4. Tratamento de Registos.....</u>	<u>41</u>
<b>4.2.2. Requisitos Tecnológicos .....</b>	<b>41</b>
<u>4.2.2.1. Fontes de Informação .....</u>	<u>41</u>
<u>4.2.2.2. Processos .....</u>	<u>42</u>
<u>4.2.2.3. Refreshamento/Atualização.....</u>	<u>43</u>
<u>4.2.2.4. Geração de mapas temáticos .....</u>	<u>44</u>
<u>4.2.2.5. Funcionalidades WebGIS.....</u>	<u>44</u>
<b>4.3. Implementação.....</b>	<b>45</b>
<b>4.3.1. Configurações .....</b>	<b>45</b>
<b>4.3.2. Itens.....</b>	<b>45</b>
<b>4.3.3. Precedências.....</b>	<b>47</b>
<b>4.3.4. Processo de instalação .....</b>	<b>48</b>
<b>4.3.5. Sequência de processamento.....</b>	<b>54</b>

<b>5. Resultados e Validações</b> .....	<b>56</b>
5.1. Objectivo .....	59
5.2. Ciclos de Testes.....	59
5.3. Tipo de testes .....	59
5.3.1. Testes de Carga.....	59
5.3.2. Testes de Utilização .....	60
5.3.3. Testes de Desempenho.....	60
5.3.4. Testes de Unidade.....	60
5.3.5. Testes de Integração .....	60
5.3.6. Testes Funcionais.....	60
5.4. Validações Efectuadas .....	61
5.4.1. Componente de ETL .....	61
5.4.2. Componente de GIS.....	62
<b>6. Conclusão</b> .....	<b>64</b>
6.1. Conclusão Final.....	65
6.2. Avaliação Crítica.....	65
6.3. Trabalho Futuro .....	66
<b>7. Bibliografia</b> .....	<b>67</b>
<b>8. Anexos</b> .....	<b>69</b>

## Índice de figuras

- Figura 1 - Arquitetura de BI (página 7)
- Figura 2 - Arquitetura Tradicional de um DW (página 9)
- Figura 3 - Arquitetura em duas camadas de um DW (página 10)
- Figura 4 - Arquitetura em três camadas de um DW (página 10)
- Figura 5 - Arquitetura com uma BD integrada de um DW (página 11)
- Figura 6 - Modelo em Estrela (página 12)
- Figura 7 - Modelo em Floco de Neve (página 12)
- Figura 8 - Componentes principais de um SIG (página 16)
- Figura 9 - Componentes principais de um LI (página 17)
- Figura 10 - DataStage Designer (página 25)
- Figura 11 - DataStage Director (página 26)
- Figura 12 - Toad (página 27)
- Figura 13 - ZOC (página 28)
- Figura 14 - Filezilla (página 29)
- Figura 15 - ArcGis (página 31)
- Figura 16 - Geo-referenciação da operadora A (página 40)
- Figura 17 - Geo-referenciação da concorrência (página 41)
- Figura 18 – Esquema de dependência das tabelas (página 47)
- Figura 19 – Configuração ArcGIS Desktop 1 (página 50)
- Figura 20 – Configuração ArcGIS Desktop 2 (página 51)
- Figura 21 – Configuração ArcGIS Desktop 3 (página 51)
- Figura 22 – Configuração ArcGIS Server 1 (página 52)
- Figura 23 – Configuração ArcGIS Server 2 (página 52)
- Figura 24 – Configuração ArcGIS Server 3 (página 53)

- Figura 25 – Configuração ArcGIS Server 4 (página 53)
- Figura 26 – Configuração ArcGIS Server 5 (página 53)
- Figura 27 – Configuração ArcGIS Server 6 (página 54)
- Figura 28 – Diagrama Principal do Processo (página 55)
- Figura 29 – Mapa por quota da operadora A (página 58)
- Figura 30 – Mapa de variação líquida da operadora A (página 58)
- Figura 31 – Mapa de variação de quota da operadora A (página 59)
- Figura 32 - Extracção de tráfego (página 69)
- Figura 33 - Agregação de 12 semanas da operadora A (página 70)
- Figura 34 – Geo-referenciação da operadora A (página 71)
- Figura 35 – Geo-referenciação da concorrência (página 72)
- Figura 36 - Apuramento de quotas por distrito (página 90)
- Figura 37 - Apuramento de quotas por concelho (página 91)
- Figura 38 - Apuramento de quota por antena (página 91)
- Figura 39 - Validação da criação do serviço passo 1 (página 92)
- Figura 40 - Validação da criação do serviço passo 2 (página 92)
- Figura 41 - Validação da criação do serviço passo 3 (página 93)
- Figura 42 - Validação da criação do serviço passo 4 (página 93)
- Figura 43 – Validação da mxd no desktop (página 94)
- Figura 44 – Validação da mxd no ArcGIS Server (página 94)

## Índice de tabelas

Tabela 1 – Níveis das sub-regiões estatísticas (página 19)

Tabela 2 – Descrição da informação das tabelas fontes (página 42)

Tabela 3 – Tabela de agregados com apuramento de quotas por operadora (página 57)

## Lista de abreviaturas

<b>Abreviatura</b>	<b>Descrição</b>
BD	<i>Base de Dados</i>
BI	<i>Business Intelligence</i>
DBA	<i>Database Administrator</i>
DDL	<i>Data Definition Language</i>
DM	<i>Data Mart</i>
DMGEO	<i>Data Mart Geográfico</i>
DML	<i>Data Manipulation Language</i>
DSS	<i>Decision Support System</i>
DW	<i>Data Warehouse</i>
EIS	<i>Executive Information System</i>
ETL	<i>Extract Transform Load</i>
FS	<i>File System</i>
GIS	<i>Geographic Information System</i>
GPRS	<i>General Packet Radio Service</i>
IDX	<i>Index</i>
LI	<i>Location Intelligence</i>
MSISDN	<i>Número de Telefone</i>
MXD	<i>Mapa de ArcGIS</i>
ODS	<i>Operational Data Store</i>
OLAP	<i>On-line Analytical Processing</i>
OPRA	<i>Operadora A</i>
SI	<i>Sistema de Informação</i>
SIG	<i>Sistema de Informação Geográfica</i>
SMS	<i>Short Message Service</i>
WFS	<i>Web Feature Service</i>
WGS	<i>World Geographic System</i>
WMS	<i>Web Mapping Service</i>

## 1. Apresentação

1.1. Contexto Académico .....	2
1.2. Contexto Científico .....	2
1.3. Contexto Institucional.....	3
1.4. Objectivos da Dissertação.....	3

## 1.1. Contexto Académico

Este trabalho corresponde à dissertação final de curso do Mestrado em Engenharia Informática da Universidade de Évora e equivale a uma unidade de 48 ECTS (European Credit Transfer and Accumulation System).

A dissertação foi baseada num dos projectos efectuados durante o estágio curricular que teve como duração 6 meses (de Novembro de 2008 a Maio de 2009) e com uma carga horária de pelo menos 40 horas semanais.

O estágio curricular decorreu para a instituição IDW (*Integration, Development and Warehousing*) em *outsourcing* numa empresa de telecomunicações.

A coordenação externa deste projecto esteve a cargo do Dr. Jorge Basílio enquanto que o orientador pedagógico ficou a cargo do Prof. Luis Arriaga da Cunha do Departamento de Informática da Universidade de Évora.

## 1.2. Contexto Científico

Hoje em dia, mais do que nunca as grandes empresas precisam de tomar decisões com o máximo de rigor e segurança possível. As empresas de telecomunicações não fogem à regra e para tal necessitam de processar grandes quantidades de informação que estão distribuídas por diversos sistemas.

É aqui que entram os sistemas de Data Warehouse que permitem uma organização e melhoramento da grande volumetria de dados produzidos por essas empresas. A sua utilização permite que as análises dos dados sejam mais completas e de qualidade superior.

Aliado a isto, é também muito importante para uma empresa de telecomunicações saber onde devem actuar, onde o seu negócio é mais ou menos dominante; aí torna-se particularmente importante a referenciação geográfica que fornece dados essenciais para o melhoramento do seu negócio, como por exemplo, saber onde deve ter campanhas publicitárias mais agressivas de modo a não se deixar vencer pela concorrência.

### 1.3. Contexto Institucional

A IDW, Integration Development and Warehousing, é uma empresa portuguesa fundada no ano de 2003 com sede em Lisboa.

Oferece uma gama de serviços, prestações e produtos privilegiando a sua integração, capacidades técnicas, de suporte e manutenção que lhes podem ser oferecidas.

As áreas de principal aposta da empresa são Business Intelligence, Segurança, Consultoria, Business Process Automation, Data Management & Business Continuity.

### 1.4. Objectivos da Dissertação

Um dos objectivos desta dissertação passa por explicar o que são e quais as funcionalidades dos sistemas de Data Warehouse, detalhando as suas características, modelos de arquitectura de dados e conceito. É também igualmente importante frisar a importância destes sistemas numa empresa de telecomunicações móvel e o que leva a que este tipo de sistema seja usado e mantido.

Outro dos objectivos passa pela utilização de um modelo de referência geográfica com base na informação conseguida através do volume de dados trabalhado por um Data Warehouse e a importância em termos de negócio que essa informação pode vir a ter.

## 2. Enquadramento do Projecto

<b>2.1. Business Intelligence</b> .....	<b>6</b>
<b>2.1.1. Conceito e enquadramento histórico</b> .....	<b>6</b>
<b>2.1.2. Business Intelligence nas empresas de Telecomunicações</b> .....	<b>6</b>
<b>2.1.3. Arquitectura de um sistema de Business Intelligence</b> .....	<b>7</b>
<b>2.2. Data Warehouse</b> .....	<b>7</b>
<b>2.2.1. Enquadramento e motivação</b> .....	<b>7</b>
<b>2.2.2. Definição de Data Warehouse</b> .....	<b>8</b>
<b>2.2.3. Arquitectura de um Data Warehouse</b> .....	<b>9</b>
<u>2.2.3.1. Arquitectura tradicional</u> .....	<b>9</b>
<u>2.2.3.2. Arquitectura em duas camadas</u> .....	<b>10</b>
<u>2.2.3.3. Arquitectura em três camadas</u> .....	<b>10</b>
<u>2.2.3.4. Arquitectura com uma base dados integrada</u> .....	<b>10</b>
<b>2.2.4. Modelo de dados de um Data Warehouse</b> .....	<b>11</b>
<u>2.2.4.1. Modelo Estrela</u> .....	<b>11</b>
<u>2.2.4.2. Modelo Floco de Neve</u> .....	<b>12</b>
<b>2.2.5. Processos Básicos</b> .....	<b>12</b>
<u>2.2.5.1. Extracção</u> .....	<b>13</b>
<u>2.2.5.2. Transformação</u> .....	<b>13</b>
<u>2.2.5.3. Carregamento e Indexação</u> .....	<b>13</b>
<u>2.2.5.4. Garantia de Qualidade</u> .....	<b>13</b>
<u>2.2.5.5. Publicação</u> .....	<b>13</b>
<u>2.2.5.6. Consulta/Actualização</u> .....	<b>14</b>
<u>2.2.5.7. Segurança</u> .....	<b>14</b>
<u>2.2.5.8. Recuperação e Backup</u> .....	<b>14</b>
<b>2.3. Geo-Referenciação</b> .....	<b>14</b>
<b>2.3.1. Conceito</b> .....	<b>14</b>
<b>2.3.2. Sistema de Informação nas Empresas de Telecomunicações</b> .....	<b>15</b>
<b>2.3.3. Sistema de Informação Geográfica</b> .....	<b>15</b>
<b>2.3.4. Location Intelligence</b> .....	<b>16</b>

---

<b>2.3.5. Geomarketing .....</b>	<b>17</b>
<b>2.4 Outros Conceitos .....</b>	<b>18</b>
<b>2.4.1. Data Mart.....</b>	<b>18</b>
<u>2.4.1.1. Data Mart versus Data Warehouse.....</u>	18
<u>2.4.1.2. Definições de Data Mart Geográfico.....</u>	18
<b>2.4.2. ETL.....</b>	<b>21</b>
<b>2.4.3. OLAP .....</b>	<b>21</b>
<u>2.4.3.1. OLAP vs OLTP .....</u>	22
<b>2.4.4. ODS.....</b>	<b>22</b>

## 2.1. Business Intelligence

### 2.1.1. Conceito e enquadramento histórico

O conceito de BI pode ser entendido como um conjunto de ideias e métodos que são usados para tomar decisões em negócios baseadas em factos sobre o mesmo.

Este termo ficou conhecido em 1989 por meio de Howard Dresner mas já antes, há milhares de anos atrás, outros povos faziam BI à sua maneira “jogando” com as informações disponibilizadas pela natureza (como as condições climáticas, comportamentos animais, períodos de seca e chuvas, etc.) para tomarem decisões que melhorassem o seu bem-estar e o da sua comunidade.

Apesar dessa distância temporal o conceito é exactamente o mesmo. O BI como o conhecemos hoje baptizado por Dresner surgiu por volta da década de 70 com o aparecimento dos primeiros sistemas de gestão de bases de dados, aliados a uma maior complexidade do negócio fizeram aparecer os primeiros produtos direccionados para analistas de negócios. Daí para cá o interesse tem aumentado de forma exponencial devido à disponibilização de análises, testes e simulações baseadas em hipóteses e fazendo uso da experiência adquirida em situações anteriores semelhantes que as empresas necessitam para as decisões estratégicas.

Com esta evolução foram também aparecendo novas ferramentas de trabalho entre as principais: Data Marts, OLAP, Data Mining, EIS e DSS e programas que geram relatórios. É também importante referir que BI não é uma ferramenta de análise nem um sistema ou serviço de informações.

### 2.1.2. Business Intelligence nas empresas de Telecomunicações

Em mercados muito competitivos como são actualmente os das empresas de telecomunicações o cliente é o objectivo principal pois é ele que escolhe a empresa, entenda-se prestador de serviços, e é determinante nas ofertas de serviços e nos impactos causados na empresa ao escolher/centrar-se num tipo específico de clientes.

A competitividade tem como consequência o aumento da concorrência que se pode manifestar de diversas formas, como por exemplo, as empresas estão sempre em busca de novos mercados procurando maneiras de entrar neles normalmente através de fusões ou aquisições de outras empresas, outra das formas, é tentar vender o melhor produto possível, isto é, ter a melhor relação possível qualidade/preço integrando vários serviços (vídeo, voz, dados, etc...) de modo a formar o melhor pacote ou produto.

É aqui que entram as soluções de BI que ajuda as empresas de telecomunicações a tomar as melhores opções para atingir vantagem competitiva. Quando analisamos a evolução da indústria das

telecomunicações verificamos que a estratégia do modelo de negócio tem-se vindo alterar com o passar do tempo deixando de se focar no produto e passando para o cliente. Já não faz sentido vender produtos e serviços para as massas mas sim produtos que possam ser adaptáveis individualmente a chave passa por saber antecipar as necessidades dos clientes e com recurso ao BI é possível satisfazê-las.

### 2.1.3. Arquitectura de um sistema de Business Intelligence

A figura 1 representa o desenho de uma arquitectura típica de Business Intelligence:



Figura 1 – Arquitectura de BI

Os dados dos sistemas operacionais podem vir de diversas fontes sendo que o ETL é a componente que se dedica à extracção, transformação e carregamento desses mesmos dados que mais tarde serão guardados em bases de dados o que se torna uma vantagem para análises futuras. O output final do BI, ou seja, aquele que chega às mãos do cliente/utilizador são normalmente relatórios, portal de internet ou documentos com análises de dados (normalmente em Excel).

## 2.2. Data Warehouse

### 2.2.1. Enquadramento e motivação

Na actualidade as bases de dados assumem um papel cada vez mais importante nas empresas e quanto maior forem maior será a volumetria de dados associados ao seu respectivo negócio fazendo com que seja cada vez mais complicado a análise dos dados existentes pois os sistemas de informação estão dispersos pelas diferentes áreas da empresa tornando praticamente impossível a tarefa de aceder a todas as

informações necessárias de maneira a que se pudessem tomar decisões fiáveis que assegurassem o futuro da empresa.

É aqui que aparece o Data Warehouse um novo conceito que consiste em organizar os dados de maneira integrada, com histórico, uniformizada e gerando uma única fonte de dados para abastecer os Data Marts. Isso permite à chefia das empresas tomar decisões baseadas em factos concretos e não em intuições, cruzando informações de diversas fontes, agilizando a tomada de decisão e diminuindo os erros.

*“A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management’s decision-making process.” — Bill Inmon*

### 2.2.2. Definição de Data Warehouse

Existem diversas maneiras de definir um DW mas nenhuma de forma rigorosa. Em português significa, literalmente, armazém de dados e é um sistema com o objectivo de responder à gestão de uma base de dados. Pode ser definido como um conjunto de hardwares e softwares que são usados para organizar um grande volume de dados gerados por diversas fontes. Durante as suas análises um DW nunca acrescenta ou retira informação apenas extrai a existente que é organizada de modo a que os seus dados sejam facilmente perceptíveis e credíveis de modo a apoiar na decisão e gestão do negócio. Sendo assim, um DW é uma “base de dados” de apoio à decisão que é mantida separadamente dos sistemas da operacionais da empresa.

De acordo com Bill Inmon, autor de *Building the Data Warehouse*, que é considerado o autor do conceito de DW, existem basicamente quatro características que o descrevem:

- **Orientado ao assunto:** os dados são organizados de acordo com o assunto em vez da aplicação. São centralizados na análise e modelação dos dados para quem tem de tomar decisões sobre assuntos que podem ser, por exemplo, o cliente, o produto ou as vendas.
- **Integrado:** quando os dados estão em várias aplicações diferentes no ambiente operacional a codificação dos dados normalmente é inconsistente. Daí ser importante uniformizar os dados porque quando são inseridos num DW devem ser consistentes quer nos seus formatos quer em nomenclaturas para uma fácil pesquisa e organização.
- **Variação em relação ao tempo:** o DW pode guardar dados durante muitos anos para serem utilizados em comparações, tendências e previsões. Estes dados não são actualizados, esta característica é muito importante do ponto de vista do negócio porque mantém um histórico do seu comportamento ao longo do tempo e permite que os resultados actuais tenham algum padrão com que sejam comparados.

- **Não volatilidade:** Os dados não são actualizados ou modificados a partir do momento em que estes entram no DW, são apenas consultados e acedidos. Os dados que são introduzidos no DW são estáticos e reflectem situações ocorridas no passado ficando assim registadas para que no futuro possam ser consultadas.

### 2.2.3. Arquitectura de um Data Warehouse

Para além da arquitectura tradicional de um DW existem também outras arquitecturas relevantes, como são o caso da arquitectura em três camadas, em duas camadas e com uma base de dados integrada que alimenta um DW.

*“Existem quatro níveis de dados numa arquitectura, que são: operacional, atómico (ou de Data Warehouse), departamental (ou nível de Data Mart) e individual.” – Bill Inmon*

É importante reconhecer que não existe uma arquitectura correcta para um DW. Para algumas empresas pode ser mais conveniente usar a arquitectura em duas camadas porque ela minimiza o custo e a complexidade de construção do DW para outras que requerem grande performance e escalabilidade a arquitectura em três camadas pode ser a mais apropriada.

As empresas quando estão a planear a construção de um DW devem examinar as alternativas disponíveis de arquitecturas e seleccionar aquela que satisfaça os seus requisitos estratégicos e organizacionais.

#### 2.2.3.1. Arquitectura tradicional

Nesta arquitectura os dados são integrados directamente para o DW sendo posteriormente trabalhados segundo as necessidades do cliente/utilizador final.

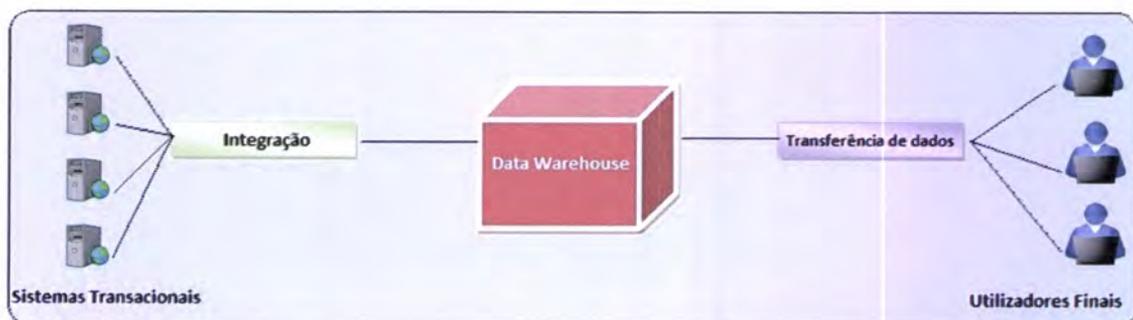


Figura 2 – Arquitectura Tradicional de um DW

### 2.2.3.2. Arquitectura em duas camadas

Esta arquitectura é a que se implementa mais rapidamente uma vez que não contempla a passagem dos dados por um DW, ou seja, os dados são actualizados directamente dos sistemas transaccionais para os DM.

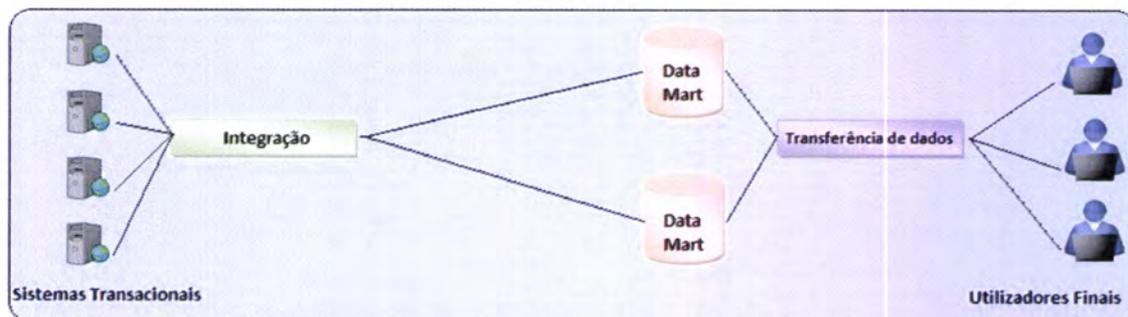


Figura 3 – Arquitectura em duas camadas de um DW

### 2.2.3.3. Arquitectura em três camadas

É a arquitectura mais utilizada pelos analistas uma vez que apresenta um bom desempenho de consultas, a agregação de dados é rápida e suporta um grande número de serviços integrados. O seu funcionamento é o seguinte: os DM tem como principal fonte um DW específico de uma determinada área de negócio da empresa a que pertence e as suas estruturas são construídas com base nesse pressuposto com o objectivo de dar uma resposta o mais rapidamente possível.

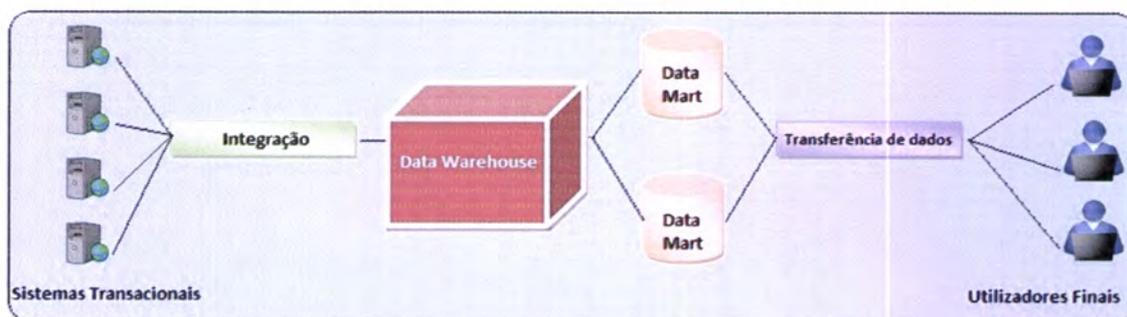


Figura 4 – Arquitectura em três camadas de um DW

### 2.2.3.4. Arquitectura com uma BD integrada

Nesta arquitectura a bd integrada é utilizada para alimentar o DW e é nela que os dados são extraídos, integrados e carregados. Ao contrário do DW na bd integrada os dados podem ser actualizados. É uma arquitectura especialmente vocacionada para ambientes instáveis em que o factor tempo é decisivo nas decisões tomadas pelo negócio.

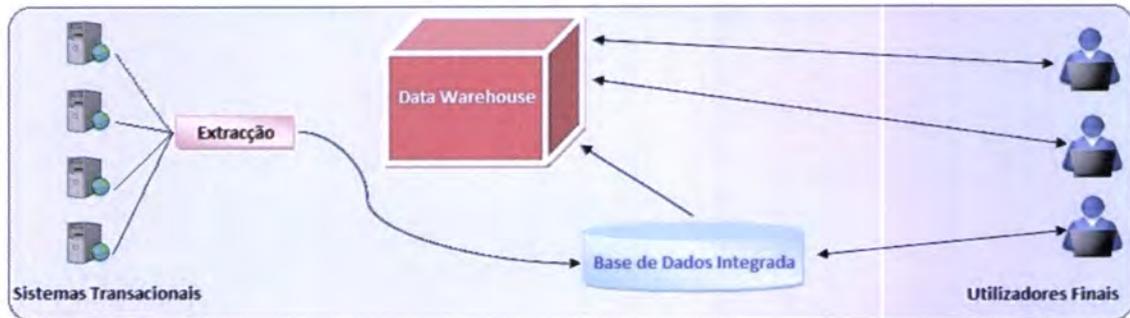


Figura 5 – Arquitectura com bd integrada de um DW

#### 2.2.4. Modelo de dados de um Data Warehouse

O papel que é desempenhado por um modelo dados é extremamente importante na perspectiva da iteração que pode haver num DW, isto é, sempre que existem desenvolvimentos que tenham de ser “fundidos” a fim de se tornarem um só, se ambos tiverem o mesmo modelo de dados os custos para essa união serão muito baixos pois todos os seus componentes estarão a usar as mesmas estruturas. Os dois principais modelos de dados de um DW são o Snow Flake (em português Floco de Neve) e o Star Schema (em português Estrela).

##### 2.2.4.1. Modelo Estrela

Este modelo é chamado de estrela porque apresenta um esquema que faz lembrar as pontas de uma estrela, em que a tabela factual se encontra ao centro e as tabelas de dimensões estão nas extremidades. A tabela factual liga-se às tabelas de dimensões através de várias ligações enquanto as tabelas de dimensão apenas se ligam à tabela factual. Este atributo diminui o número de uniões e facilita o cruzamentos entre tabelas no entanto as tabelas de dimensão são muito grandes ocupando por isso muito espaço em disco é também um modelo de simples navegação para os softwares possibilitando um acesso mais rápido aos dados.

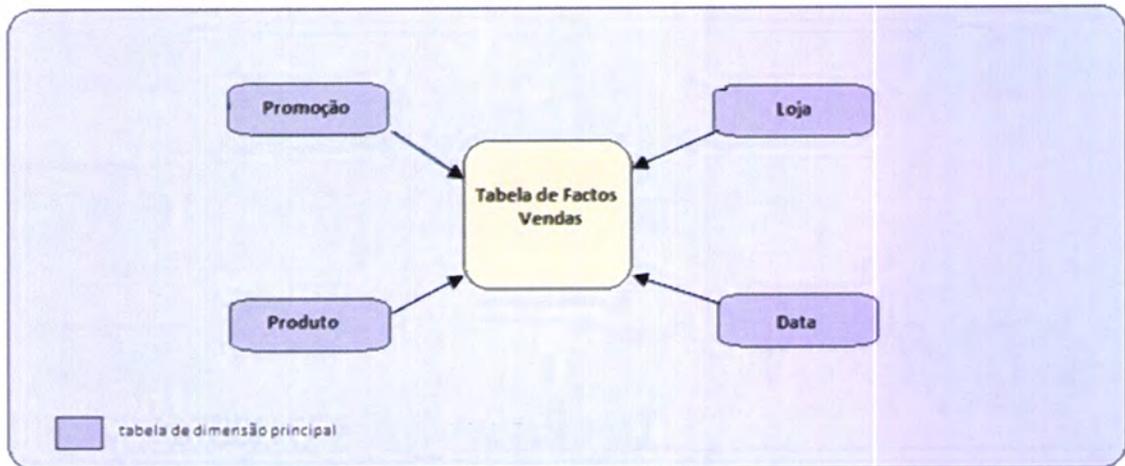


Figura 6 – Modelo em Estrela

2.2.4.2. Modelo Floco de Neve

Pode ser considerado como uma extensão do modelo em estrela uma vez que cada uma das pontas da estrela do modelo anterior passa a ser ela própria o centro de uma estrela passando assim a existirem vários níveis de agregação. As tabelas de dimensão principais passam a ser normalizadas por outras tabelas de dimensão auxiliares. Assim, utilizam-se mais tabelas para representar as mesmas dimensões tornando o modelo mais complexo, no entanto, ocupam menos espaço em disco do que o modelo estrela.

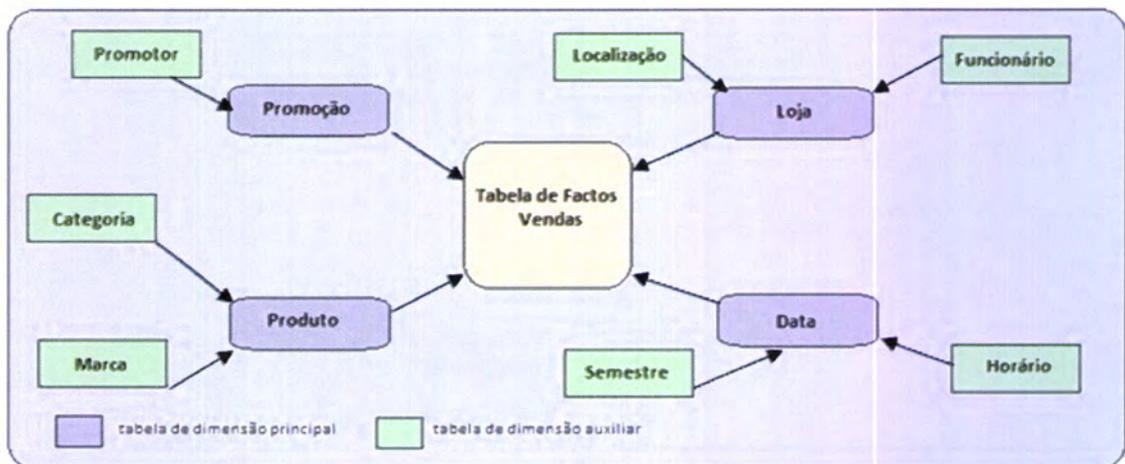


Figura 7 – Modelo em Floco de Neve

**2.2.5. Processos Básicos**

Qualquer Data Warehouse pode ser composto por sub-processos básicos sendo eles: extracção, transformação, carregamento e indexação, garantia de qualidade, publicação, consulta/actualização, segurança, recuperação e backup.

#### 2.2.5.1. Extracção

É através do processo de extracção que se pode ler e entender os dados fonte e copiar partes que são necessárias para a área de Data Staging (estrutura intermédia entre os sistemas operacionais e o DW) no futuro. É importante também referir que os dados podem-se encontrar em vários formatos (como por exemplo fontes Informix ou arquivos ASCII) sendo necessária a sua conversão para a estrutura standard.

#### 2.2.5.2. Transformação

O objectivo da transformação de dados é o de obter os dados necessários para o negócio uniformizando-os, para isso, deverá seguir um conjunto de regras: correcção de erros de digitação, resolução de conflitos de dados, tratamento de dados perdidos, conversão para formatos padrão, remoção de campos seleccionados que não são úteis para o DW e construção de agregados para melhoria de performance em consultas comuns.

#### 2.2.5.3. Carregamento e Indexação

Após a transformação, a volumetria de dados a ser carregada é muito grande e estes serão agregados e indexados conforme o seu tipo de informação. Irão ser feitos particionamentos (normalmente por dia ou mês mas também poderão ser por outros campos) e será assegurado o controlo da sua qualidade através da verificação da sua integridade.

#### 2.2.5.4. Garantia de Qualidade

A garantia da qualidade de dados que pode ser assegurada através de diferentes métodos, entre eles: verificação de inconsistências através de testes de contagens (duplicados), somas, tempos, utilização de ferramentas que devolvem relatórios de excepções sobre um conjunto de dados carregados ou validação de carregamentos e mapeamentos das tabelas, comparando tabelas finais com as fontes do processo.

#### 2.2.5.5. Publicação

Garantida a qualidade de dados os utilizadores serão notificados de que os novos dados estão prontos. Serão também comunicadas quaisquer mudanças que tenham sido efectuadas nas dimensões e novos pressupostos.

#### 2.2.5.6. Consulta/Actualização

A actualização envolve adequação a mudanças em status, hierarquias e organização de dados e também a sua correcção. Quanto às consultas, podem ser efectuadas consultas *ad hoc*, escrita de relatórios, aplicações de tomadas de decisão e mineração de dados.

#### 2.2.5.7. Segurança

A segurança num DW passa por reconhecer os utilizadores e fornecer-lhes direitos específicos de acesso aos dados. Nem todos os utilizadores têm acesso aos mesmos dados. Quem atribui essas permissões a determinados utilizadores será um DBA respeitando um determinado conjunto de regras/normas/leis pois cada utilizador deverá ter a informação necessária para cumprir a sua função.

#### 2.2.5.8. Recuperação e Backup

O ponto fulcral desta etapa reside em como, onde e quando tirar os “snapshots” de dados legados e dos metadados para propósitos de arquivamento e recuperação de problemas que possam vir a surgir.

## 2.3. Geo-Referenciação

### 2.3.1. Conceito

Para se explicar o conceito de Sistema de Geo-Referenciação é importante ter em conta uma série de outros conceitos para que se possa entender na totalidade o que está por trás desta ideia. Para começar é necessário saber que um sistema é um conjunto de procedimentos metodológicos que permitem produzir novas formas de conhecimento a partir de outros conhecimentos já adquiridos e que informação é a transmissão de conhecimento, algo que se transmite e obtém, por forma a enriquecer o nosso sistema de conhecimentos.

Enquadrando estes conceitos na Geo-Referenciação podemos assim dizer que é uma associação de sistemas de referência terrestre por meio de sistemas de coordenadas aplicando para esse efeito modelos de transformações matemáticas. Quando se geo-referência um mapa é estar a obter as coordenadas do ponto do mapa que se quer geo-referenciar num determinado sistema de referência. Para isso o mapa já deverá ter geo-referenciados os seus pontos de controlo que são os locais que tem formatos perfeitamente identificáveis como estradas, linhas de comboio, montanhas, rios, lagos entre outros. O modo mais comum e actual de se poder levantar esses pontos de controlo é feito através de GPS que significa *Global Positioning System*, ou em português sistema de posicionamento global.

A informação que se recolhe numa determinada área que é analisada é a informação espacial que confere à informação uma determinada localização (posição relativa a um referencial).

### 2.3.2. Sistema de Informação nas Empresas de Telecomunicações

A informação numa grande empresa de telecomunicações é um recurso imprescindível. Daí haver necessidade de garantir a sua fiabilidade e coerência. A sua importância reside no facto de ser através dela que a empresa consegue competir com a concorrência no mercado e ir acompanhando as suas investidas, respondendo aos seus requisitos internos e à interactividade com o exterior.

Os SI nas empresas de telecomunicações são por isso sistemas ou subsistemas interligados que utilizam equipamentos na aquisição, armazenamento, manipulação e controlo de dados incluindo os softwares e hardwares utilizados. O acesso à informação disponibilizada pelo SI e a capacidade de aplicar os conhecimentos são muito importantes para o aumento do desenvolvimento de actividades comerciais que possibilitem ganhar vantagem no mercado competitivo que é o das telecomunicações.

### 2.3.3. Sistema de Informação Geográfica

É uma forma particular, dos sistemas de informação, aplicada a dados espaciais geo-referenciados com o fim de guardar, manipular, visualizar e analisar dados de natureza espacial (referenciados à superfície da terra), representando-a sob a forma de mapas. Trata-se de programas de computador que tem por objectivo gerir e analisar informações espaciais ou não espaciais (que estejam ligadas a entidades espaciais), visando a tomada de decisão. É um sistema para capturar, integrar, manipular, analisar e visualizar dados geo-referenciados.

Envolve uma base de dados espacialmente referenciada e software apropriado (figura 8). Contém subsistemas para entrada, representação, armazenamento, pesquisa e saída de dados e também para a sua gestão, transformação e análise. Contempla ainda geração de relatórios, gráficos e estatísticas.

*"O DW pode aproveitar as interessantes ferramentas de um sistema de informação geográfica (SIG) para apresentar a informação contida numa determinada morada ou rota de dados. O que na verdade pode incentivar a melhorias de projectos para incluir atributos que permitem a análise mais completa dos nossos dados através das capacidades do GIS" - Ralf Kimball & Margo Deen*

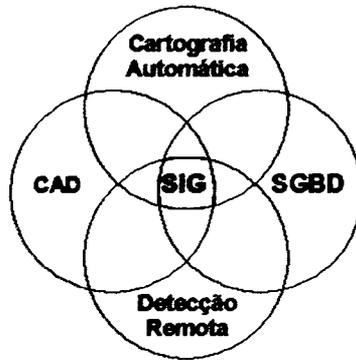


Figura 8 – Componentes principais de um SIG

#### 2.3.4. Location Intelligence

Location Intelligence é a capacidade de organizar e interpretar factos complexos através da utilização de relações geográficas inerentes a todas as informações. Ao combinar dados de localização geográfica com dados de negócio as empresas obtêm informação crítica que as ajuda a tomar as melhores decisões para otimizar processos e aplicações importantes que podem melhorar o seu desempenho e resultados. As suas ferramentas (figura 9) fomentam diversas fontes de dados como mapas aéreos ou sistemas de informação geográfica. O termo Location Intelligence também é usado para descrever a integração da componente geográfica em processos de BI e ferramentas OLAP.

Um LI surge sempre com um problema de negócio, ou seja, sempre que é solicitado é com intuito de solucionar, corrigir ou melhorar algo que o negócio tenha sugerido. Numa empresa de telecomunicações, pode ser necessário para ajudar a resolver os seguintes problemas: descobrir quais são as zonas com mais baixa cobertura de rede móvel para se saber onde colocar um reforço de sinal, saber qual a região que tem mais ou menos quota de mercado para que o marketing actue em conformidade ou saber quais os locais mais propícios a expandir a rede.

O que mais difere um LI de um SIG são os seguintes aspectos:

- O SIG pertence normalmente a um departamento enquanto que o LI é transversal a toda a empresa
- O LI surge com problemas de negócio enquanto que o SIG não
- O LI tem uma oferta de soluções maior que o SIG



Figura 9 – Componentes principais de um LI

### 2.3.5. Geomarketing

É uma forma de dividir geograficamente o território de acordo com nichos e perfis específicos de consumidores de determinadas regiões que tem em conta um série de variáveis, entre as quais, comportamentais, socioeconómicas ou culturais.

O seu objectivo é o de conhecer o território de actuação da empresa. A correcta delimitação das respectivas áreas possibilita por em prática actividades como a divulgação de produtos e serviços ou atendimento diferenciado ao consumidor. Para tal o geomarketing tem de se basear em diferentes matérias como a geografia, marketing, estatística e economia.

A principal vantagem do geomarketing é o aumento substancial de eficácia na tomada de decisões permitindo uma análise mais rápida e prática de informações espacialmente distribuídas que possibilitam às empresas lançar estratégias de marketing que lhes possam trazer resultados positivos. Para além do apoio à tomada de decisões e/ou identificação do seu público-alvo serve para aprofundar o conhecimento sobre seus clientes e encontrar futuras oportunidades e consumidores. Para aplicar o conceito é preciso alinhar as informações por divisão geográfica uma vez que é importante ter qualidade e volume das mesmas através de:

- Mapas e informações de pesquisas comportamentais;
- Censo de domicílios;
- Orçamento familiar

A partir desta informação, a empresa consegue avaliar e descobrir qual é a região mais adequada para aplicar um teste de determinado produto, serviço ou estabelecer um “ponto-de-venda”. Com tudo isto, o geomarketing tem um papel fundamental para as empresas porque aliado às novas tecnologias possibilita o acesso a informações actualizadas, distribuídas geograficamente e analisadas em tempo real.

## 2.4. Outros Conceitos

### 2.4.1. Data Mart

Apesar do DM, tal como o DW, ainda não ter uma definição consensual, pode-se definir como um subconjunto de dados dentro de um DW que se aplica numa determinada área tornando-se num local específico de acesso à informação do DW.

Um DM é alimentado por uma única fonte de dados, o DW, e é construído com a finalidade de responder às necessidades de um utilizador específico por isso é muito importante a sua usabilidade para que os seus dados sejam facilmente acedidos e processados.

*“os Data Marts são muito interessantes para resolver certos problemas, mas não são necessariamente substitutos de um projecto de DW. Um DM não deve ser um pequeno DW, com a finalidade de ser rápido ou possuir dados ainda não suportados para o DW” – Ralph Kimball*

#### 2.4.1.1. Data Mart Versus Data Warehouse

Relativamente às definições entre DM e DW existem opiniões diferentes não sendo consensual a sua descrição.

Segundo Bill Inmon o DW e os Data Marts têm estruturas essencialmente diferentes. Na sua opinião, é difícil integrar um conjunto de Data Marts e mesmo que se conseguisse não resultaria num DW. Para Bill Inmon, o DM resulta do DW.

Já Ralph Kimball tem uma opinião contrária, afirmando que o DW é constituído pela união de todos os seus Data Marts.

#### 2.4.1.2. Definições de Data Mart Geográfico

Existem algumas definições que são básicas para se entender o conceito de um DM Geográfico e estas são:

- **Código Postal** - O Código Postal é composto por 7 algarismos, seguido da designação postal (nome de uma localidade ou freguesia). O 1º algarismo do código corresponde a uma das 9 regiões de Código Postal do país. Assim, o primeiro conjunto de 4 algarismos, identifica as zonas de Código Postal em que cada uma destas regiões se divide, correspondendo na maioria dos casos à área de actuação de um Centro de Distribuição Postal. Cada zona de Código Postal é composta por uma ou mais áreas de Código Postal, cada uma com uma designação postal, a qual fora dos centros urbanos, corresponde em geral, a freguesias.

O segundo conjunto de 3 algarismos do Código Postal identifica pequenas parcelas do território dentro de cada área de Código Postal: ruas, partes de ruas, bairros, lugares ou agrupamentos de pequenos lugares.

- **Divisão Administrativa** - Estrutura hierárquica dinâmica estabelecida e alterada por lei que divide o território nacional em 3 tipos de entidades: Distrito, Concelho, Freguesia.  
*Distrito* - Grande divisão administrativa, que se subdivide em Concelhos.  
*Concelho* (município) - Circunscrição administrativa, que se subdivide em freguesias.  
*Freguesia* - Circunscrição administrativa em que se subdivide o Concelho.
  
- **NUTS** - Nomenclatura Comum das Unidades Territoriais Estatísticas - nomenclatura que define sub-regiões estatísticas em que se divide o território dos Estados-Membros União Europeia.  
 As sub-regiões estatísticas estão subdivididas em 3 níveis:
  - NUTS 1
  - NUTS 2
  - NUTS 3

Nível	População máxima	População mínima
NUTS 1	7.000.000	3.000.000
NUTS 2	3.000.000	800.000
NUTS 3	800.000	150.000

Tabela 1 – Níveis das sub-regiões estatísticas

Estas sub-regiões devem ser obrigatoriamente utilizadas, pelos Estados-Membros e pela Comissão Europeia, para a recolha, compilação e divulgação a nível comunitário de dados estatísticos.

- **Célula** – é uma área geográfica coberta por uma antena transmissora (BTS) associada a um operador. Uma célula pode ter uma dimensão variável de 2 a 50 Km de diâmetro, dependendo do tipo de terreno (mais ou menos montanhoso) e da potência de transmissão. Ao conjunto de antenas de um operador chama-se um sistema de células. Ao utilizar o serviço de um operador tem-se acesso ao seu sistema de células. Ao sair do sistema de células que, em Portugal tipicamente acontece quando se sai do País, esse sistema pode permitir que o seu telefone passe para outro sistema de células, ou seja, outro operador vizinho sem se notar. A esta situação (pertencer a um operador mas estar a utilizar a infra-estrutura de outro) chama-se roaming.

- **Célula de activação** - Célula onde é detectado o primeiro registo de tráfego associado ao cartão da operadora A.
  
- **Célula Preferencial do Mês** - Célula atribuída a cada cartão da operadora A onde é efectuado mais tráfego no mês.
  1. São considerados os cartões com registo de tráfego no universo de células e os cartões activos sem tráfego do parque da operadora A.
  2. É considerado todo o tráfego de Voz, SMS e GPRS.
  3. No tráfego de voz é considerado o tráfego originado e o tráfego terminado (ambas tem o mesmo peso no processo de cálculo).
  4. Na ausência de tráfego no período, será considerada a célula referente ao último período com tráfego desde que se verificou a última activação do cartão da operadora A ou na ausência deste, a informação da célula de activação do cartão.
  
- **Célula Preferencial do Trimestre** - Célula atribuída a cada cartão da operadora A onde se registou mais actividade durante um período de três meses. A actividade da célula será determinada com base em:
  1. Quantidade total de tráfego durante o período 3 meses em análise;
  2. Duração total do tráfego durante o período de 3 meses em análise;
  3. A célula preferencial será aquela que contém maior quantidade de tráfego e dentro desta a que tiver maior duração das comunicações.

Apenas são considerados para a determinação da célula preferencial a 3 meses, a informação de célula resultante de tráfego. Caso durante o trimestre não exista registo de tráfego, a célula preferencial a 3 meses será a mesma referente ao registo do mês anterior;
  
- **Célula Preferencial por Banda Horária** - Célula atribuída a cada cartão da operadora A onde foi efectuado maior quantidade de tráfego para cada período/banda horária. São aplicadas as seguintes regras:
  1. São considerados os cartões com registo de tráfego, bem como os cartões activos sem tráfego e que constam do parque da operadora A.
  2. É considerado todo o tráfego de voz, SMS e GPRS.
  3. A célula preferencial para cada um dos períodos é determinada utilizando os seguintes critérios por ordem apresentada:
    - Maior quantidade de tráfego no período;
    - Maior período de tráfego;
    - Célula de menor ordem

4. Na ausência de tráfego em cada um dos três períodos considerados é atribuída a célula correspondente ao último período com tráfego registado, ou na ausência deste a informação da célula de activação do cartão.

#### 2.4.2. ETL

ETL é a sigla em inglês para Extract Transform Load.

É uma metodologia composta por ferramentas de extracção, transformação e carregamento de dados.

A extracção é normalmente feita periodicamente, os métodos utilizados servem para isolar os dados que vão ser usados pelos sistemas de apoio à decisão. É necessário filtrar apenas os dados que não terão utilidade para reduzir ao máximo o desperdício de espaço.

A fase de transformação começa logo após a altura em que os dados são extraídos dos sistemas operacionais para a Staging Area. É nesta fase que é adicionada informação temporal aos dados para que possam ficar guardados com um histórico. É iniciado um conjunto de processos como a limpeza e integração dos dados que têm como objectivo garantir a sua consistência no DW.

O processo de carregamento é a última fase e só pode ser realizado após a extracção e transformação dos dados. Nesta etapa carregam-se os dados tratados e armazenados na Staging Area para o DW.

#### 2.4.3. OLAP

OLAP é a sigla em inglês para Online Analytical Processing e é um mecanismo de manipulação de dados de elevada capacidade, destinado a suportar e operar sobre estruturas de dados multidimensionais.

*“As técnicas OLAP servem fundamentalmente para comparar grupos de dados relacionando-os com diversos critérios, de que se podem salientar a escala temporal, a distribuição geográfica ou a perspectiva económico-social” – Carlos Pampulim Caldeira*

O principal objectivo da filosofia OLAP é o de maximizar a sua função analítica para isso tem necessidade de encontrar o melhor modelo para orientar e organizar a informação disponível.

Na tecnologia OLAP o passado é guardado pois os dados são utilizados para análise e previsões futuras, sendo necessário que fiquem registadas informações sobre o passado.

### 2.4.3.1. OLAP vs OLTP

OLTP (Online Transaction Processing) são as aplicações que se preocupam com processamento individualizado de determinados eventos. Um sistema OLTP é uma aplicação que gere um conjunto de transacções num período, normalmente em tempo real, que explicam como são desenvolvidas as actividades de uma empresa.

*“Os sistemas OLAP contrapõem-se aos OLTP por serem ferramentas de análise de dados. Ou seja, o objecto das aplicações OLAP não é a forma como a informação é guardada num repositório de dados, mas o modo como essa mesma informação pode ser eficientemente analisada.” – Carlos Pampulim Caldeira*

Os sistemas OLTP e OLAP não podem existir em conjunto na mesma aplicação, o que faz com que sejam incompatíveis entre si. As suas arquitecturas são diferentes, os dados podem ser os mesmos mas nos sistemas OLAP são organizados de modo a beneficiar os utilizadores que pretendem analisar a informação do que aqueles que pretendam “actos isolados” de processamento, como por exemplo, registos de dados ou alteração de informação existente nos sistemas.

### **2.4.4. ODS**

O ODS é a sigla em inglês para *Operational Data Store* que significa base de dados operacional e é constituído por dados que foram sujeitos a diferentes acções de transformação e conversão provenientes de diferentes sistemas operacionais.

A sua informação é organizada por áreas de análise e é regularmente actualizada com um período curto de latência.

O ODS pode ser considerado como uma cópia da base de dados, pois possibilita obter uma perspectiva geral e actual dos dados de produção, onde se podem fazer análises e extrair informação sem afectar em nada os dados de produção.

Para Ralf Kimball, o ODS é uma fonte de dados pormenorizada em tempo real enquanto que Bill Immon considera que é um sistema operacional, aparte do DW, que suporta constantes acessos e actualizações operacionais. Ambos concordam que é uma estrutura de dados integrada, orientada ao assunto,

constantemente actualizada com informação pormenorizada de modo a suportar os sistemas transaccionais.

---

## 3. Metodologias utilizadas

<b>3.1. Ferramentas</b> .....	<b>25</b>
<b>3.1.1. DataStage</b> .....	<b>25</b>
<b>3.1.2. Toad</b> .....	<b>26</b>
<b>3.1.3. ZOC</b> .....	<b>27</b>
<b>3.1.4. FileZilla</b> .....	<b>28</b>
<b>3.1.5. ArcGIS</b> .....	<b>29</b>
<u>3.1.5.1. ArcView</u> .....	<b>29</b>
<u>3.1.5.2. ArcEditor</u> .....	<b>30</b>
<u>3.1.5.3. ArcInfo</u> .....	<b>30</b>
<b>3.2. Linguagens</b> .....	<b>31</b>
<b>3.2.1. PL-SQL</b> .....	<b>31</b>
<b>3.2.2. Shell Script</b> .....	<b>32</b>

## 3.1. Ferramentas

### 3.1.1. DataStage

É uma ferramenta de ETL que foi concebida para gerir fluxos de dados. Suporta integração de dados em tempo real, permite fazer a recolha, integração e transformação (ETL) de grandes volumes de dados com alto desempenho de estruturas simples até às mais complexas. A recolha de dados pode ser proveniente de diversos tipos de fontes garantindo que os dados integrados são utilizados pela maioria das marcas mais populares de software empresarial como é o caso da SAP, Siebel e Oracle. Possui uma interface intuitiva para um programador que lhe permite maximizar a eficácia e implementação na construção dos seus projectos. A empresa que o desenvolveu inicialmente chama-se Ascential Software sendo mais tarde adquirido pela IBM.

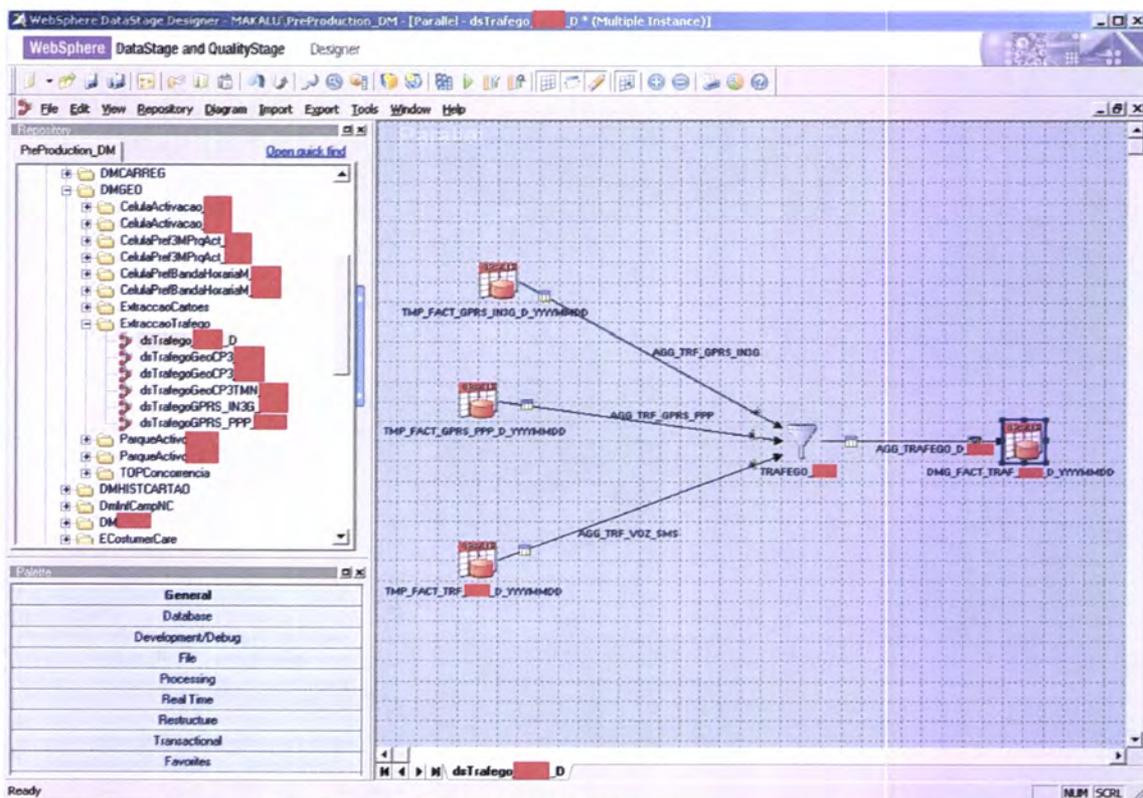


Figura 10 – DataStage Designer



Job name	Status	Started	On date
dsTratago..._D_20081229	Finished (see log)	15:26	19-02-2009
dsTratago..._D_20081230	Finished (see log)	15:34	19-02-2009
dsTratago..._D_20081231	Finished (see log)	15:42	19-02-2009
dsTratagoGeoCP3...20081229	Finished (see log)	15:26	19-02-2009
dsTratagoGeoCP3...20081230	Finished (see log)	15:34	19-02-2009
dsTratagoGeoCP3...20081231	Finished (see log)	15:42	19-02-2009
dsTratagoGeoCP3...20081229	Finished (see log)	15:29	19-02-2009
dsTratagoGeoCP3...20081230	Finished (see log)	15:37	19-02-2009
dsTratagoGeoCP3...20081231	Finished (see log)	15:45	19-02-2009
dsTratagoGeoCP3TMN...20081229	Finished (see log)	15:25	19-02-2009
dsTratagoGeoCP3TMN...20081230	Finished (see log)	15:32	19-02-2009
dsTratagoGeoCP3TMN...20081231	Finished (see log)	15:41	19-02-2009
dsTratagoGPRS_IN3G...20081229	Finished (see log)	15:25	19-02-2009
dsTratagoGPRS_IN3G...20081230	Finished (see log)	15:33	19-02-2009
dsTratagoGPRS_IN3G...20081231	Finished (see log)	15:41	19-02-2009
dsTratagoGPRS_FPPP...20081230	Finished (see log)	15:33	19-02-2009
dsTratagoGPRS_FPPP...20081231	Finished (see log)	15:41	19-02-2009
dsTratagoGPRS_FPPP...20081229	Finished (see log)	15:25	19-02-2009
dsTratago..._D	Completed	11:42	03-02-2009
dsTratagoGeoCP3...	Completed	11:42	03-02-2009
dsTratagoGeoCP3...	Completed	16:59	18-02-2009
dsTratagoGeoCP3TMN...	Completed	15:24	19-02-2009
dsTratagoGPRS_IN3G...	Completed	11:44	03-02-2009
dsTratagoGPRS_FPPP...	Completed	11:44	03-02-2009

Status of jobs: 24 entries

Figura 11 – DataStage Director

### 3.1.2. Toad

O Toad é um software que é usada para desenvolvimento e administração de bases de dados relacionais usando o SQL consegue interpretar o dicionário Oracle: tabelas, índices, procedimentos e através de browsers multi-tabelas consegue desempenhar muitas outras funções. Suporta as seguintes base de dados: Oracle Database, Microsoft SQL Server, IBM DB2, e MySQL. Corre em todas as plataformas Windows 32-bit, incluindo Windows 95, 98, NT, 2000, XP e Vista. Não suporta nenhum ambiente que não seja Microsoft embora tenha uma plataforma independente do Toad que se chama Tora que oferece algumas das funcionalidades básicas da ferramenta. A empresa que desenvolveu o Toad chama-se Quest.

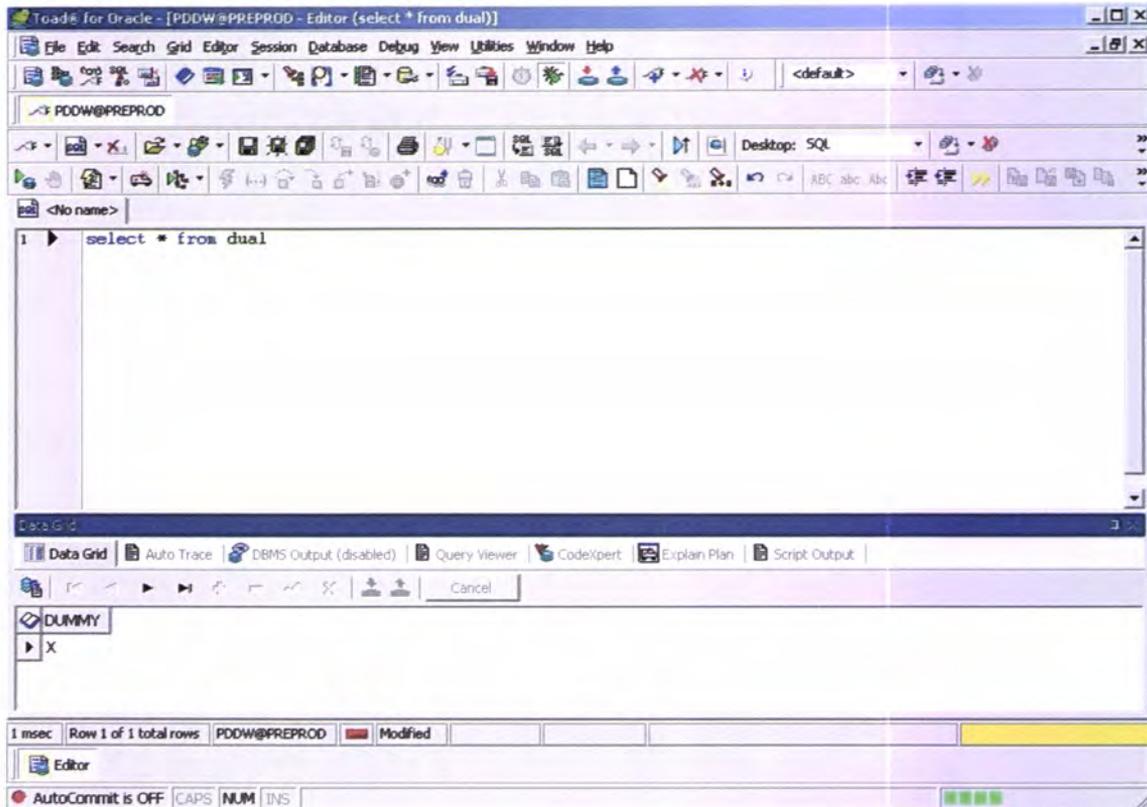


Figura 12 – Toad

### 3.1.3. ZOC

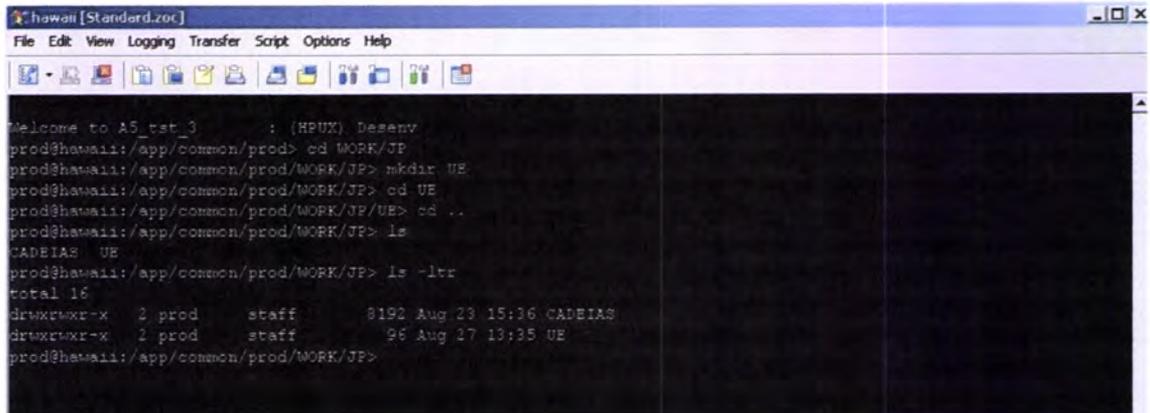
O ZOC é um software que permite trabalhar em Unix utilizando todos os seus comandos e permitindo navegar através das directorias existentes na máquina. É um terminal e cliente Telnet para Windows e Mac OS que suporta Telnet, Modem, SSH 1 e 2, ISDN de série TAPI, rlogin e outros meios de comunicação. O seu emulador tem pleno de cores, meta-chaves e locais impressão, VT102, VT220 e vários tipos de ANSI, bem como WYSE, TVI, TN3270, e da Sun CDE, teclado completo, remapeamento, REXX script e suporte named pipes.

Protocolos disponíveis:

- Secure Shell (SSH v1/v2) com a chave pública / privada e autenticação porto encaminhamento (tunneling)
- Telnet (RFC-Telnet, pura sockets TCP, SSL-Telnet)
- Modem via porta serial e TAPI (Windows modem)
- RDIS através CAPI V2.0 (incluindo X.25 e X.31 apoio)
- rlogin

- File Transfer protocolos: ASCII, X-Modem, Y-Modem, Z-modem, Kermit, SCP

É um software comercial desenvolvido por Markus Schmidt Emtec da Innovative Software.



```

hawaii [Standard.zoc]
File Edit View Logging Transfer Script Options Help

Welcome to A5_tst_3      : (HPUX) Desenv
prod@hawaii:/app/common/prod> cd WORK/JP
prod@hawaii:/app/common/prod/WORK/JP> mkdir UE
prod@hawaii:/app/common/prod/WORK/JP> cd UE
prod@hawaii:/app/common/prod/WORK/JP/UE> cd ..
prod@hawaii:/app/common/prod/WORK/JP> ls
CADEIAS UE
prod@hawaii:/app/common/prod/WORK/JP> ls -ltr
total 16
drwxrwxr-x  2 prod  staff    8192 Aug 23 15:36 CADEIAS
drwxrwxr-x  2 prod  staff    96 Aug 17 13:35 UE
prod@hawaii:/app/common/prod/WORK/JP>

```

Figura 13 – ZOC

### 3.1.4. FileZilla

O FileZilla é um software utilizado para manipulação de ficheiros e directorias através de diferentes máquinas/servidores através do seu interface gráfico bastante intuitivo em que as pastas estão dispostas “em árvore” e o utilizador arrasta os ficheiros de um local para outro acedendo de forma local ou remota. Qualquer acção que o utilizador tenha é exibida na mensagem de log que regista os comandos enviados pelo FileZilla e as respostas enviadas pelo servidor. É um cliente FTP, SFTP e FTPS que permite também criar uma lista de FTP com todas as suas configurações (conexão, porta, protocolo, login, etc.). É um software de código livre tanto para Windows como para Linux.

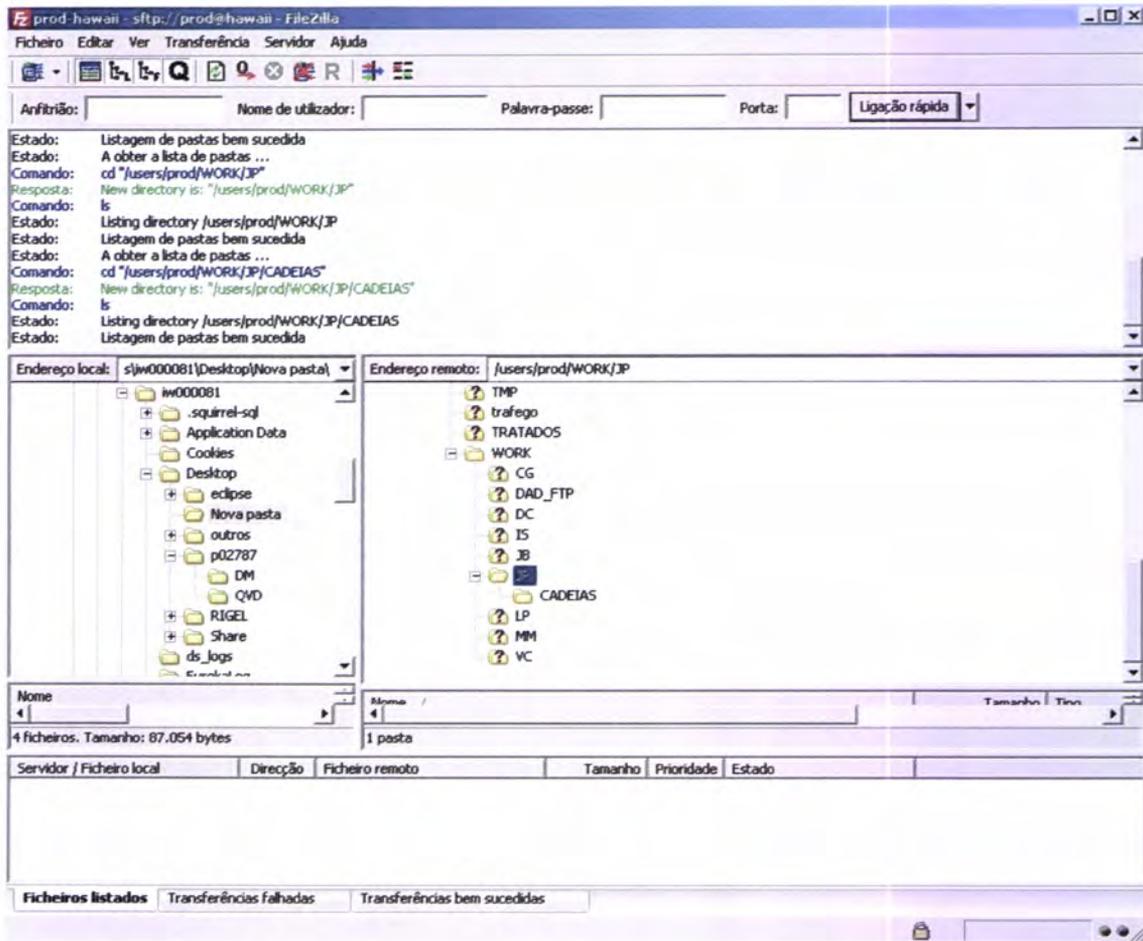


Figura 14 – Filezilla

### 3.1.5. ArcGis

O ArcGis é um sistema escalável, integrado e de fácil compreensão que vai ao encontro dos utilizadores de SIG. Ao nível das aplicações de desktop existem três “pacotes” de ArcGIS: ArcView, ArcEditor e ArcInfo.

#### 3.1.5.1. ArcView

É uma solução SIG com grande poder de visualização, query, análise e criação de mapas desenvolvido para trabalhar em ambiente multi-utilizador (Microsoft Windows e Linux). Disponibiliza ferramentas interactivas para explorar, seleccionar, visualizar, editar, analisar, classificar e aplicar simbologia de dados bem como criar automaticamente, gerir e actualizar os metadados. É “Web-enabled.” isto significa que pode navegar dinamicamente (browse) e utilizar dados geográficos disponíveis na World Wide Web. ArcView inclui ArcCatalog, ArcMap, e o ArcToolbox.

As suas funcionalidades são:

- Explorar, visualizar, e executar queries espaciais
- Aceder a um conjunto de ferramentas de edição para as shapefiles, e executar um conjunto de operações de edição simples sobre as geodatabases
- Executar operações de geo-processamento
- Criar mapas de apresentação (saídas gráficas) de forma simples através de um conjunto de ferramentas cartográficas base e de wizards
- Criar relatórios e gráficos em duas e três dimensões avançados
- Executar projecção de informação espacial (vector e raster)
- Visualizar mapas da internet
- Importar e exportar dados de variados formatos (interoperabilidade)

#### 3.1.5.2. ArcEditor

É uma aplicação intermédia do ArcGIS tem mais funcionalidades que o ArcView mas menos que o ArcINFO e está desenhado para organizações que aplicam o sistema integrado de SIG tal como o seu nome indica disponibilizando mais funções avançadas de edição. O exemplo da sua utilidade encontra-se nas organizações que tem uma base de dados de dados geográficos estruturada sobre um SGBD e tem sobre ela o ArcSDE. A edição e alteração de dados neste tipo de estruturas de geodatabase avançada apenas são possíveis dispondo do ArcEditor. Assim com varias licenças de ArcView é possível vários a visualizar os dados e com ArcEditor a editar os dados.

As suas funcionalidades são:

- É possível editar e actualizar as bases de dados geográficas e criar estilos personalizados de acesso aos dados
- Pode-se ainda executar funcionalidades avançadas de edição, como por exemplo adicionar e modificar o comportamento das unidades (features), regras e relações entre as mesmas. Pode-se ainda ter um conjunto completo de análises espaciais sobre os dados

#### 3.1.5.3. ArcInfo

É a aplicação mais completa e rica nas operações de SIG, e dentro dos produtos de ArcGIS. Tem todas as funcionalidades do ArcEditor incluindo ainda uma panóplia de ferramentas no ArcToolbox e uma versão completa do ArcInfo WorkStation. (ARC, ArcEdit, ArcPlot, AML e todas as extensões). ArcInfo é um SIG completo para a criação, actualização, análise, query, processamento, gestão de DB e visualização de dados.

A sua principal funcionalidade é poder ter acesso as funções integradas de análise e integra-las noutras aplicações, isto é possível pois é desenvolvido com tecnologia COM (Component Object Model), pelo que se pode criar modelos de grande complexidade do mundo real e atribuir-lhe regras e comportamentos aos objectos (com).

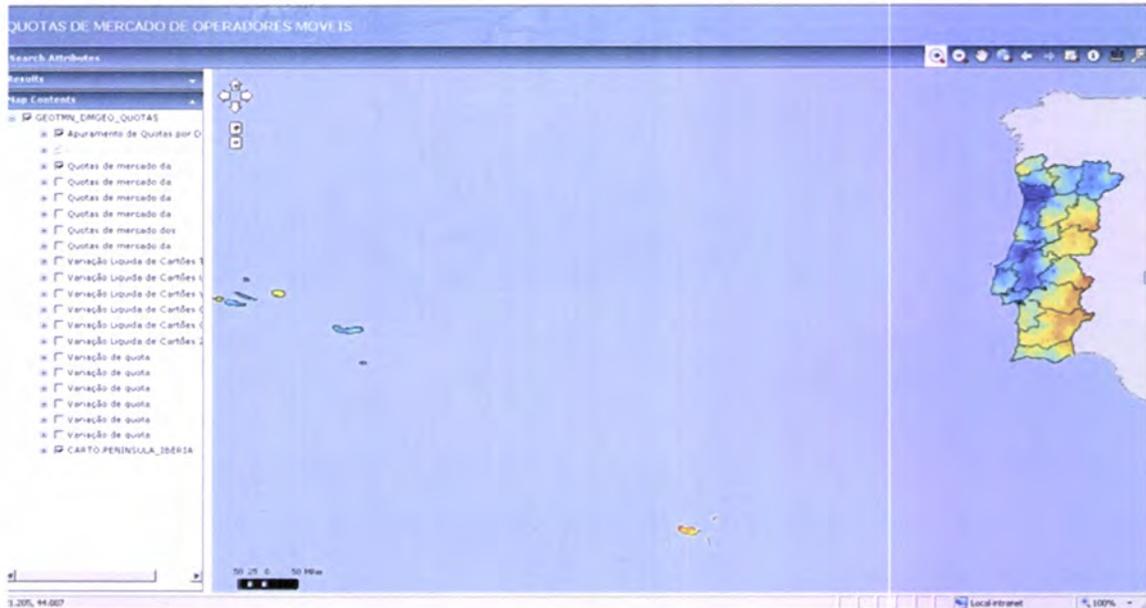


Figura 15 - ArcGis

## 3.2. Linguagens

### 3.2.1. PL-SQL

O PL-SQL é uma linguagem extensiva ao à linguagem SQL, isto é, é como se fosse um complemento do SQL especializada nas bases de dados da Oracle. A sigla significa *Procedural Language- Structured Query Language*. Por ser uma linguagem processual retira muitas das definições da linguagem SQL. A unidade básica da linguagem é o bloco (de código). Todos os programas em PL-SQL são compostos por blocos que estão uns dentro de outros. Cada bloco, regra geral, desempenha uma acção lógica no programa.

Algumas definições importantes dos principais componentes desta linguagem:

**Stored Procedure e Função** - Um stored procedure (ou em português Procedimento Armazenado) é um programa PL-SQL que pode ser accionado por uma aplicação, um trigger ou uma ferramenta ORACLE. A principal diferença entre procedure e uma função está no facto de que a função executa comandos e retornar um resultado enquanto que o procedure apenas executa os seus comandos.

**Package** - Um package (ou em português Pacote [de código]) é um arquivo que agrupa funções, stored procedures, cursores e variáveis tudo no mesmo local.

**Trigger** - É um programa PL-SQL armazenado na base de dados e que é executado imediatamente antes ou após os comandos *Insert*, *Update* e *Delete*. Após estes últimos comandos convém sempre fazer um *Commit*.

**DBLink**- Através dos *Data Base Link* podemos criar ligações entre schemas dentro da mesma base de dados ou de bases de dados diferentes podendo assim cruzar tabelas e usar os seus dados mesmo sendo de locais diferentes.

**Blocos PL-SQL** - Esses blocos são compostos por procedures e funções. Um bloco com a estrutura básica segue as seguintes regras:

[Declare

-- Secção opcional, em que todos os objectos são declarados]

Begin

-- Secção em que os comandos PL-SQL são colocados

[Exception

-- Secção em que os erros são tratados]

End;

### 3.2.2. Shell Script

O Shell Script é uma linguagem de programação interpretada e imperativa. Para se definir bem o que é esta linguagem de programação podemos dizer que são as linhas de comando Unix guardadas num arquivo de texto, normalmente com a extensão *.sh*, estas são interpretadas na linha de comandos e posteriormente executadas sequencialmente realizando as funções pedidas. Para além de executar os programas no terminal o Shell Script também possui os seus próprios comandos, variáveis e funções. Existem também vários tipos de shell como por exemplo: *bash*, *csh*, *ksh* e *ash*. O mais utilizado actualmente é o *bash*. Para a primeira linha de comando terá de ser *#!/bin/bash* para o definir como tal. Uma das suas principais vantagens é que não precisa de ser compilado basta definir o seu tipo na primeira linha de código como foi dito anteriormente e para ser executado utilizar o comando *chmod* para definir os seus privilégios de utilização.

Exemplos de alguns dos comandos utilizados:

**echo** Imprime texto na tela

**read** Captura dados do usuário e coloca numa variável

**exit** Finaliza o script

**sleep** Dá uma pausa de segundos no script

**clear** Limpa a tela

**if** Controla o fluxo que testa uma ou mais expressões

**case** Controla o fluxo que testa várias expressões ao mesmo tempo

**for** Controla o fluxo que testa uma ou mais expressões

**while** Controla o fluxo que testa uma ou mais expressões

Para as variáveis:

**-eq** se é igual

**!=** se é diferente

**-gt** se é maior

**-lt** se é menor

---

## 4. Descrição do Desenvolvimento

<b>4.1. Introdução</b> .....	<b>35</b>
<b>4.2. Requisitos</b> .....	<b>35</b>
<b>4.2.1. Requisitos de Negócio</b> .....	<b>35</b>
<u>4.2.1.1. Universo de Informação</u> .....	35
<u>4.2.1.2. Universo de Espacial</u> .....	36
<u>4.2.1.3. Regras</u> .....	36
<u>4.2.1.4. Tratamento de Registos</u> .....	41
<b>4.2.2. Requisitos Tecnológicos</b> .....	<b>41</b>
<u>4.2.2.1. Fontes de Informação</u> .....	41
<u>4.2.2.2. Processos</u> .....	42
<u>4.2.2.3. Refrescamento/Atualização</u> .....	43
<u>4.2.2.4. Geração de mapas temáticos</u> .....	44
<u>4.2.2.5. Funcionalidades WebGIS</u> .....	44
<b>4.3. Implementação</b> .....	<b>45</b>
<b>4.3.1. Configurações</b> .....	<b>45</b>
<b>4.3.2. Itens</b> .....	<b>45</b>
<b>4.3.3. Precedências</b> .....	<b>47</b>
<b>4.3.4. Processo de instalação</b> .....	<b>48</b>
<b>4.3.5. Sequência de processamento</b> .....	<b>54</b>

## 4.1. Introdução

O desenvolvimento divide-se em duas componentes, a de ETL e a de GIS.

A componente de ETL, têm como âmbito a automatização do processo de cálculo de quotas trimestral (agregações de 12 semanas). O processo das quotas divide-se em dois processos, um processo de extracção de informação relativa ao universo de cartões da operadora A e concorrência e um segundo processo relativo à geo-referenciação do universo e determinação das respectivas quotas de mercado. É utilizado o ODS como fonte de dados.

Para a componente de GIS é implementado um conjunto de processos espaciais na tecnologia ArcGIS em que é feito o processo de constituição de elementos geográfico do documento de ArcGIS Desktop (mxd) que suportará o processo de publicação de dados na componente WEB (ArcGIS Server).

## 4.2. Requisitos

Existem dois tipos de requisitos, os de negócio, que são aqueles onde se encontram descritas todas as funcionalidades a implementar pelo processo de carregamento de dados e especificações funcionais a disponibilizar pelo processo, e os tecnológicos, são os que descrevem as necessidades tecnológicas do desenvolvimento para dar resposta ao pedido.

### 4.2.1. Requisitos de Negócio

O processo desenvolvido é estruturado de forma a existir apenas uma única extracção de informação do tráfego on-line. O processo do cálculo de quotas divide-se nas seguintes etapas:

- Extracção diária de tráfego
- Agregação semanal de tráfego
- Agregação das últimas 12 semanas
- Cálculo da quota nacional (processo semanal)
- Geo-referenciação de parque da Operadora A e concorrência
- Cálculo da quota por unidade territorial (processo mensal)

[anexo processos]

#### 4.2.1.1. Universo de Informação

Um processo de extracção de informação tem associado um conjunto de regras que delimitam o universo de informação a carregar pelo processo.

Para cada um dos operadores móveis serão identificados:

***Cartões com tráfego on-net*** - cartões que apenas realizam tráfego dentro da sua rede de fidelização. Isto é, para a Operadora A, são os que realizam tráfego apenas entre cartões da Operadora A enquanto que para a concorrência será extrapolado com base no volume de cartões da Operadora A que apenas realiza tráfego on-net. A identificação deste universo será feita ao nível do Distrito.

***Cartões com tráfego off-net*** – cartões que realizam tráfego entre diversas redes de operadoras móveis. A sua determinação tem como objectivo a disponibilização de informação para análises de geomarketing. Para a Operadora A, são os cartões da Operadora A que realizam comunicações com outros cartões pertencentes a outras operadoras de comunicações enquanto que para a concorrência são os cartões da concorrência que efectuem tráfego para a rede da Operadora A.

***Cartões associados ao serviço casa*** – cartões associados à rede fixa (serviço casa fixo) disponibilizados pelos operadores móveis. No tráfego da concorrência serão considerados os cartões cujo MSISDN começa por 2. São excluídos os cartões da Operadora A, cujo MSISDN esteja na gama dos cartões Serviço Casa da Operadora A. Na concorrência são excluídos os MSISDN da concorrência que pertencem à rede fixa e que efectuem tráfego para a rede da Operadora A.

Os cartões são identificados pelo seu código único (cod\_cartao), isto é, cada cartão só têm um único código associado. É um número inteiro que pode ir até nove dígitos. Enquanto que o MSISDN é o número de telefone. Um MSISDN pode ter estado associado a vários cartões (cod\_cartao) mas um cartão só tem ou teve um único MSISDN associado.

#### 4.2.1.2. Universo Espacial

O universo do estudo das quotas corresponde a todo o território nacional. O detalhe irá variar entre a unidade das células e as de Distrito. Serão ainda feitos apuramentos de quotas ao nível nacional com detalhe ao nível de concelho.

#### 4.2.1.3. Regras

As regras que o desenvolvimento teve que respeitar são as seguintes: sazonalidade, quarentena, extracção do universo, cálculo da quota da concorrência, agregação semanal, agregação do “trimestre” móvel, geo-referenciação e redistribuição de cartões indefinidos.

##### ❖ **Regra de Sazonalidade**

Para a definição das semanas de sazonalidade são definidas as seguintes regras:

- Todas as semanas de sazonalidade são excluídas do processo de agregação.
- Nessas semanas (ex: mês de Agosto) a geo-referenciação dos cartões deste mês é feita com base na informação do mês anterior.

Este critério é adoptado pois verifica-se uma transferência de células preferencialmente para o sul do país e zonas costeiras, devido ao período de férias (migração da população).

#### ❖ Regra de Quarentena

A primeira e última semanas do ano são consideradas como semanas de quarentena. Para cada uma delas será aplicado o seguinte critério:

- As semanas de quarentena (52 e 01) são excluídas dos processos de agregação para cálculo de quotas.
- À primeira semana do ano e às suas 11 semanas seguintes é aplicado o factor de correcção (Factor correctivo do período de quarentena) da quota da concorrência de forma a garantir o acréscimo da concorrência na mesma proporção que a Operadora A.

Os novos cartões da concorrência que aparecem após o natal e que, inicialmente apenas realizam tráfego dentro da sua rede, tem também de ser contemplados no estudo.

#### ❖ Regra de Extração do Universo

Para o processo de extração de tráfego são consideradas as seguintes regras para determinação do universo de cartões.

##### 1. Cartões de tráfego

- São elegíveis para o universo da Operadora A e da concorrência todos os cartões que efectuem comunicações do tipo voz e vídeo chamada sendo guardada a informação sobre a chamada
- A distinção dos cartões de tráfego para cada um dos operadores da concorrência é feita com base no código tipo origem-destino.

O código tipo origem-destino (cod\_tipo\_orig\_dest) é um número inteiro que caracteriza o operador destino do registo da chamada, cdr, que se encontra no ODS e serve de fonte ao processo de extração.

##### 2. Cartões “casa”

Para o universo da Operadora A e da concorrência são excluídos os cartões associados ao serviço casa, uma vez que não é possível distinguir o serviço fixo da concorrência. A sua identificação é feita pelo código tipo origem-destino e no MSISDN da concorrência (números começados por 2).

#### ❖ Regra de cálculo da quota da concorrência

Regras para extrapolação de quota da concorrência:

## 1. Cartões on-net da concorrência

- Ao nível da quota nacional a determinação do número de cartões da concorrência é feito através do processo de extrapolação dos dados apurados para os cartões da Operadora A, factor de extrapolação nacional, que está definido em função do operador.

## 2. Cartões on-net da concorrência – quota por distrito (factor de exclusividade):

- Ao nível de distrito a determinação é feita pelo número de cartões que efectuam tráfego do tipo on-net através de um processo de extrapolação dos dados apurados para os cartões da Operadora A, factor de exclusividade, que está definido em função do operador.

A determinação do factor de exclusividade segue as seguintes regras:

- A exclusividade (cartões on-net / cartões totais) para a Operadora A é determinada num dado momento, com detalhe ao nível de distrito.
- Para o momento considerado é determinada a regressão linear entre quota de mercado da Operadora A e a exclusividade.
- A extrapolação do valor de tráfego on-net para a concorrência e a respectiva quota de mercado determinada num dado momento é com base nos parâmetros da regressão linear.
- O valor de exclusividade da concorrência, por cada distrito, é extrapolado tendo por base a regressão linear. Este parâmetro é guardado em tabela de referência.

### ❖ Regra de agregação semanal

Para a determinação do universo da quota semanal é feita a agregação dos cartões distintos que realizam tráfego. Na agregação semanal dos cartões da Operadora A é considerado o tráfego off-net e on-net enquanto que na concorrência apenas é considerado tráfego off-net.

#### 1. Regras de agregação semanais da Operadora A

As regras para a agregação semanal são as seguintes:

- Os cartões que realizam tráfego da Operadora A serão extraídos do ODS, diariamente, e serão agregados à semana.
- A determinação do universo de cartões da Operadora A numa semana é feito verificando os cartões que:
  - Realizam tráfego na semana em iteração;
  - Ainda não foram registados na tabela agregada da semana (tabela com valores distintos de Cartões);

#### 2. Regras de agregação semanais da concorrência

A determinação dos cartões concorrência numa semana obedece aos seguintes critérios:

- Todos os cartões da concorrência, registados no ODS, que realizam tráfego com cartões da Operadora A na semana em iteração;
- Cartões registados no ODS com o código tipo origem-destino na gama dos valores da concorrência e tendo em consideração o número chamado e o número chamador.

❖ **Regras de agregação do “trimestre” móvel**

É feita a agregação de 12 semanas dos cartões da Operadora A e concorrência. No processo de agregação trimestral são excluídos os cartões que estão registados nas semanas definidas como sazonalidade e de período de quarentena.

❖ **Regras de geo-referenciação**

A geo-referenciação dos cartões da Operadora A e concorrência que efectua tráfego é feita com base nos seguintes pressupostos:

- O processo de geo-referenciação do universo de cartões da concorrência é feito tendo por base o número da Operadora A mais contactado por cada cartão.
- No período onde existem semanas sazonais é feita com base na informação relativa à célula preferencial a três meses do mês anterior à sazonalidade, e os novos cartões são geo-referenciados na célula preferencial no mês posterior.

➤ **Geo-referenciação Operadora A**

A geo-referenciação dos cartões da Operadora A é feita segundo os seguintes critérios:

1. Para o universo de cartões é identificada a célula preferencial a três meses, disponibilizada pelo processo de geo-referenciação do parque activo. É ainda disponibilizada a informação de concelho e distrito associado à célula.
2. Para os cartões cuja célula não foi identificada, é determinado o código postal da conta a que está associada o cartão.
3. Para os cartões cujo código postal de conta não foi identificado é determinado o código postal do utilizador (contribuinte).
4. Para os cartões, cujo código postal do utilizador não foi identificado, são distribuídos pelo território.
5. Os cartões identificados no ponto 2, 3 e 4, bem como os que ficam com o estado indefinido são redistribuídos ao nível das antenas (células agregadas).
6. São feitos os apuramentos ao nível de Distrito e de Concelho.

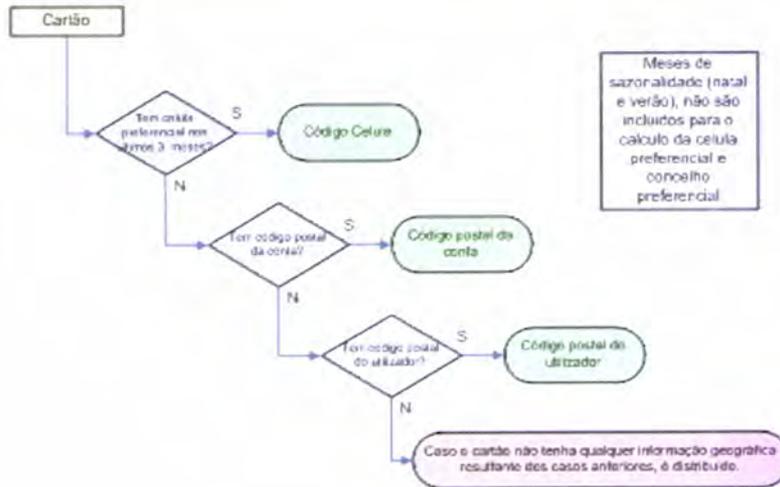


Figura 16 – Geo-referenciação da operadora A

### ➤ Geo-referenciação Concorrência

A geo-referenciação dos cartões da concorrência é feita segundo os seguintes critérios:

1. Para o universo de cartões da concorrência apurados no tráfego do ODS são identificados os respectivos cartões TOP (número de telefone da Operadora A com o qual um cartão da concorrência realiza mais tráfego).
2. A geo-referenciação dos cartões da concorrência é feita atribuindo a célula preferencial a 3 meses do cartão TOP que lhe está associado. É ainda disponibilizada a informação de Concelho e Distrito associado à célula.
3. Caso não exista célula preferencial para o cartão TOP associado, é atribuído o código postal da conta do respectivo cartão TOP.
4. Caso não exista informação de código postal de conta do cartão TOP associado será determinado o código postal do utilizador (contribuinte) do respectivo cartão TOP.
5. Caso não exista informação de código postal do utilizador do cartão TOP serão distribuídos pelo território nacional.
6. Os cartões identificados nas alíneas 3, 4 e 5, bem como os que ficam com o estado indefinido serão redistribuídos ao nível das antenas (células agregadas).
7. São feitos os apuramentos ao nível de distrito e de concelho.

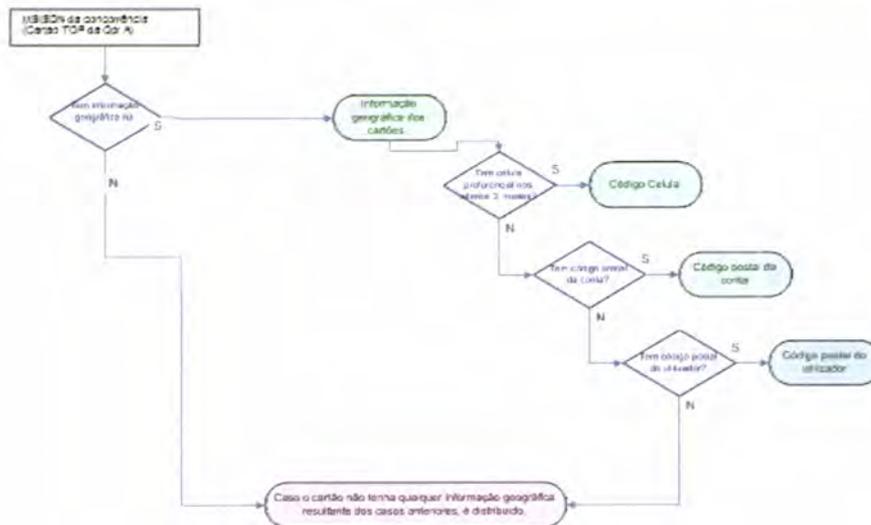


Figura 17 – Geo-referenciação da concorrência

#### ❖ Regras de redistribuição de cartões indefinidos

Para os cartões que não seja possível associar uma dimensão espacial serão distribuídos no território. Os critérios para a distribuição são os seguintes:

- Em função da contribuição de cada antena para a quota nacional da Operadora A, é feita a distribuição dos cartões indefinidos de forma proporcional.
- É feita a agregação da distribuição dos cartões indefinidos para as unidades de concelho e distrito.

#### 4.2.1.4. Tratamento de Registos

As regras, na componente gráfica, utilizadas para o tratamento de registos são as seguintes:

- **Sistema de coordenadas** - a informação geográfica utilizada está projectada segundo o sistema de coordenadas geográfico WGS84.
- **Mapa de quota de mercado** – para os níveis de concelho, distrito e código postal é apresentado, ao nível alfanumérico, a distribuição da quota por operador e as respectivas variações.

### 4.2.2. Requisitos Tecnológicos

#### 4.2.2.1. Fontes de Informação

São consideradas fontes de informação todas as estruturas que fornecem input ao sistema ou a uma parte dele e podem ter diferentes formatos ou pertencer a diferentes sistemas.

**ODS** - fonte de dados para a determinação das quotas de mercado da Operadora A e da concorrência.

**DM** – fonte de dados para as tabelas de referência relacionadas com a célula, dia, semana, mês, trimestre, ano.

**DMGEO** - fonte de dados para as tabelas de referência relacionadas as dimensões espaciais freguesia, concelho, distrito.

Tabela	Descrição do conteúdo da informação das tabelas
HIST_CONTA	Contas dos clientes da Operadora A
REF_CELULA	Referências da Célula
HIST_CONTRIBUINTE	Clientes da Operadora A
REF_DIA	Referência para informação sobre o dia
FACT_CDR_HIST	Guarda a informação de histórico proveniente do (ODS)
DMG_REF_FREGUESIA	Referência com informação relativa às freguesias
DMG_AGG_GE_HOR_GBL_M1	Informação da célula preferencial a 3 meses e do mês
DMG_AGG_GEOMKT_BASE_M1	Informação base relativa às hierarquias espaciais da célula preferencial de cada cartão
DMG_REF_ANTENA	Referência da antena
DMG_REF_CELULA_ANTENA	Referência que liga a antena com a célula
DMG_REF_COD_POSTAL_BASE	Referência que identifica o código postal base
DMG_REF_CODIGO_POSTAL	Informação de código postal
DMG_REF_CONCELHO	Referência com informação relativa a concelho e a NUTSIII
DMG_REF_DISTRITO	Referência do Data Mart Geográfico de Distritos

Tabela 2 – Descrição da informação das tabelas fontes

#### 4.2.2.2. Processos

São processos todos os conjuntos de acções sequenciais que contribuem para que se atinja um objectivo. Tipicamente possuem restrições ao nível de *timing* de processamento, performance, reprocessamento e tecnológicas.

#### ❖ Cálculo das quotas

Os cálculos das quotas são efectuadas tendo por base a informação de cartões de 12 semanas de tráfego (trimestre móvel) segundo as seguintes características:

- **Líder por concelho** – Com base na quota de mercado de cada uma das operadoras é determinado para cada concelho a operadora que o lidera.
- **Quotas nacionais** – Para cada uma das operadoras (Operadora A e concorrência) é feita a determinação da quota nacional tendo por base o total de cartões aferidos.
- **Quotas por unidades territoriais** – Para cada uma das operadoras (Operadora A e concorrência) é feita a determinação da quota por unidade territorial (distrito, concelho, código postal e antena), tendo por base o total de cartões aferidos. A distribuição por unidade territorial contempla o processo de redistribuição de cartões classificados como “geo-referenciação indefinida”, pelas unidades territoriais de acordo com regras definidas.
- **Adições líquidas por operadora** – O cálculo das adições líquidas representam o acréscimo ou decréscimo absoluto de cartões no parque de cada um dos operadores entre dois trimestres consecutivos (ex: 1T -2T). Para o cálculo é utilizado o universo total de cartões de cada uma das operadoras.
- **Variação de quota por operadora** - O cálculo da variação de quota representará o acréscimo ou decréscimo da quota das operadoras entre dois trimestres consecutivos (ex: 2T -1T). Para o cálculo será utilizado o universo total de cartões de cada uma das operadoras face ao universo total de cartões.

#### 4.2.2.3. Refrescamento/Actualização

Requisitos que especificam a origem e o destino dos dados a serem refrescados bem como a periodicidade dos refrescamentos ou que acções os desencadeiam.

##### ❖ **Tabelas de Referência**

Serão considerados os seguintes processos de refrescamento:

- Para as tabelas de referência relativas às antenas foi desenvolvido um processo automático que garante a actualização da referência das antenas com base na informação das células (REF\_CELULA) e que garante a consistência temporal necessária para os processos de análise das quotas.

- A actualização das seguintes tabelas de referência é feita através da inserção manual de novos valores, sendo elas:
  - Factor exclusividade da concorrência
  - Factor correctivo do período de quarentena
  - Semanas de sazonalidade
  - Extrapolação nacional

#### 4.2.2.4. Geração de mapas temáticos

Os mapas temáticos a implementar são os seguintes:

**Mapas realizados ao nível da agregação de células** (corresponde à antena):

- **Adições Líquidas (Operadora A + concorrência)** – Representação do valor das adições líquidas do trimestre, por antena.
- **Quota de mercado Operadora A** - Representação do valor da quota da operadora A do trimestre, por antena.
- **Quota de mercado da concorrência** (todas as outras operadoras) – Representação do valor da quota da concorrência do trimestre, por antena.
- **Variação da quota de mercado da concorrência** (todas as outras operadoras) – Representação do valor de variação de quota de toda a concorrência do trimestre, por antena
- **Variação da quota de mercado para cada operadora** – Representação do valor da variação de quota do trimestre, por antena

#### 4.2.2.5. Funcionalidades WebGIS

Para a disponibilização da análise de quotas é utilizada a plataforma ArcGIS Server. As funcionalidades são disponibilizadas directamente pela ferramenta. O acesso à interface será realizado através da funcionalidade de autenticação, cujo desenvolvimento não é do âmbito deste projecto.

Das funcionalidades BUILT-IN disponibilizadas pela componente servidor é utilizada a Web Services incluindo mapa e imagem (WMS e WFS).

Das funcionalidades BUILT-IN disponibilizadas pela componente cliente são utilizadas as seguintes:

- Zoom
- Consulta e pesquisa de atributos.
- Os mapas a disponibilizar são sobre a forma de mapas temáticos e/ou mapas de classes, sendo a sua legenda igualmente disponibilizada na tabela de conteúdos.

## 4.3. Implementação

### 4.3.1. Configurações

No ficheiro de configurações que existe para cada user de uma máquina devem ser configuradas as parametrizações, utilizadas pelos scripts, para acederem às bases de dados ou directorias pretendidas. Por norma deve-se identificar o que se pretende aceder, através de comentários, no próprio código.

Um exemplo de como se deve fazer uma parametrização no ficheiro de parametrizações:

```
#####
# Acesso ao ODS
#####
par_user_ods=userods
par_pass_ods=abc123!
par_bd_ODS=ODS
```

### Componente Gráfica (GIS)

#### ❖ Configurações em Windows

A instalação da componente ArcGIS Desktop foi instalada no path “Drive:\Program Files \ ArcGIS”, sendo a componente de Python em c:\Pthon25. É importante salientar que esta configuração também é feita no ArcGIS Server.

#### ❖ Livrarias adicionais

- Python

Foi instalada a livreria Cx\_Oracle para conexão às bases de dados.

- Cliente Oracle 10g

Está disponível na máquina servidor de ArcGIS Desktop as livrarias e aplicações do cliente Oracle 10g.

### 4.3.2. Itens

Todos os itens que foram criados no âmbito deste projecto encontram-se abaixo organizados segundo o tipo a que pertencem.

Procedimentos:

- INSERT\_DMG\_MTD\_FACTOR\_ACTIVADO.sql
- INSERT\_DMG\_MTD\_FACTOR\_CORRECTIVO.sql
- INSERT\_DMG\_MTD\_FACTOR\_EXCLUSIVIDADE.sql
- INSERT\_DMG\_MTD\_FACTOR\_EXTRAP\_NACION.sql
- INSERT\_DMG\_MTD\_GAMA\_CONCORRENCIA.sql
- INSERT\_DMG\_MTD\_QUARENTENA.sql
- INSERT\_DMG\_MTD\_SAZONALIDADE.sql
- INSERT\_DMG\_REF\_COD\_POSTAL\_BASE.sql
- PRC\_DMGEIO\_AGG12SEMAN\_CONC.sql
- PRC\_DMGEIO\_AGG12SEMAN\_OPRA.sql
- PRC\_DMGEIO\_GREF\_ANTEN\_CONC.sql
- PRC\_DMGEIO\_GREF\_ANTEN\_OPRA.sql
- PRC\_DMGEIO\_APURAM\_CRT\_NAC.sql
- PRC\_DMGEIO\_OPER\_QUOTA\_NAC.sql
- INSERT\_DMG\_MTD\_EXCLUSAO.sql
- PRC\_DMGEIO\_APURAM\_CRT\_ANT.sql
- PRC\_UPDATE\_DMG\_REF\_ANTENA.sql
- IS\_NUMBER.sql
- PRC\_DMGEIO\_DROP\_GEOMKT.sql
- DMG\_REF\_RECLASS\_RASTER\_INSERT.sql
- PRC\_DMGEIO\_APURAM\_GIS.sql
- PRC\_DMGEIO\_CACHE.py
- PRC\_DMGEIO\_IDW\_QUOTAS.py
- PRC\_DMGEIO\_REF\_ANTENAS.py

Estruturas das tabelas (\*.ddl) e dos índices(\*.idx):

- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li>• DMG_MTD_FACTOR_EXCLUSIVIDADE.ddl</li> <li>• DMG_MTD_FACTOR_EXCLUSIVIDADE.idx</li> <li>• DMG_MTD_FACTOR_EXTRAP_NACION.ddl</li> <li>• DMG_MTD_FACTOR_EXTRAP_NACION.idx</li> <li>• DMG_MTD_GAMA_CONCORRENCIA.ddl</li> <li>• DMG_MTD_GAMA_CONCORRENCIA.idx</li> <li>• DMG_MTD_QUARENTENA.ddl</li> <li>• DMG_MTD_QUARENTENA.idx</li> <li>• DMG_MTD_SAZONALIDADE.ddl</li> <li>• DMG_MTD_SAZONALIDADE.idx</li> <li>• DMG_REF_ANTENA.ddl</li> <li>• DMG_REF_CELULA_ANTENA.ddl</li> <li>• DMG_REF_CELULA_ANTENA.idx</li> <li>• DMG_REF_COD_POSTAL_BASE.ddl</li> <li>• SEQ_COD_ANTENA.ddl</li> <li>• SEQ_COD_CELULA_ANTENA.ddl</li> <li>• DMG_AGG_APURA_QUOTA_M1.ddl</li> <li>• DMG_AGG_APURA_QUOTA_M1.idx</li> <li>• DMG_AGG_CRT_CONC_12S_S1.ddl</li> <li>• DMG_AGG_CRT_CONC_S1.ddl</li> <li>• DMG_AGG_CRT_OPRA_12S_S1.ddl</li> <li>• DMG_AGG_CRT_OPRA_S1.ddl</li> <li>• DMG_AGG_GEOREF_CONC_M1.ddl</li> <li>• DMG_FACT_CRT_OPERADOR.ddl</li> <li>• DMG_FACT_CRT_OPERADOR.idx</li> </ul> | <ul style="list-style-type: none"> <li>• DMG_FACT_QUOTA_OPERADOR.ddl</li> <li>• DMG_MTD_EXCLUSAO_PARAM_AGENTE.ddl</li> <li>• DMG_MTD_EXCLUSAO_PARAM_AGENTE.idx</li> <li>• DMG_MTD_EXCLUSAO_PARAM_CATEG.ddl</li> <li>• DMG_MTD_EXCLUSAO_PARAM_CATEG.idx</li> <li>• DMG_MTD_EXCLUSAO_PARAM_PPP_PPS.ddl</li> <li>• DMG_MTD_EXCLUSAO_PARAM_PPP_PPS.idx</li> <li>• DMG_MTD_EXCLUSAO_PARAM_TARIF.ddl</li> <li>• DMG_MTD_EXCLUSAO_PARAM_TARIF.idx</li> <li>• DMG_MTD_EXCLUSAO_PARAMETRO.ddl</li> <li>• DMG_MTD_EXCLUSAO_PARAMETRO.idx</li> <li>• DMG_MTD_FACTOR_ACTIVADO.ddl</li> <li>• DMG_MTD_FACTOR_ACTIVADO.idx</li> <li>• DMG_MTD_FACTOR_CORRECTIVO.ddl</li> <li>• DMG_MTD_FACTOR_CORRECTIVO.idx</li> <li>• DMG_AGG_CRT_OPER_M1.ddl</li> <li>• DMG_AGG_CRT_OPER_M1.idx</li> <li>• DMG_AGG_CRT_ORIDEST_M1.ddl</li> <li>• DMG_AGG_CRT_ORIDEST_M1.idx</li> <li>• DMG_FACT_CRT_NGEO.ddl</li> <li>• DMG_AGG_AP_QUOTA_CC_M1.ddl</li> <li>• DMG_AGG_AP_QUOTA_DD_M1.ddl</li> <li>• DMG_AGG_APURA_QUOTA_M1.ddl</li> <li>• DMG_REF_RECLASS_RASTER.ddl</li> <li>• DMG_AGG_APURA_QUOTA_M1.idx</li> </ul> |
|--|--|

Scripts:

- dsTrafegoGeoCP3OPRAh.sh
- dsTrafegoGeoCP3OPRAreph.sh
- dsDMG\_AGG\_CRT\_OPRA\_S1h.sh
- dsDMG\_AGG\_CRT\_CONC\_S1h.sh
- dsDMG\_AGG\_CRT\_OPRA\_12S\_S1h.sh
- dsDMG\_AGG\_CRT\_CONC\_12S\_S1h.sh
- dsDMG\_REF\_ANTENAh.sh.sh
- dsExtracaoCDRs\_BDLh.sh
- dsExtracaoCDRs\_OPRAh.sh
- dsExtracaoCDRsh.sh
- dsGEOR\_CONC\_TOP\_Kh.sh
- dsDMG\_AGG\_GEOREF\_CONC\_M1h.sh
- dsQUOTA\_OPRA\_CELPRFh.sh
- dsDMG\_FACT\_CRT\_OPERADORh.sh
- dsDMG\_FACT\_QUOTA\_OPERADORh.sh
- dsDMG\_AGG\_CRT\_OPER\_M1h.sh
- DSgeomkt\_CONCh.sh
- DSgeomkt\_FREGh.sh
- dsDMG\_AGG\_APURA\_QUOTAh
- dsDMG\_EXEC\_ArcGIS\_ANTENAh
- dsDMG\_EXEC\_ArcGIS\_CACHEh
- dsDMG\_EXEC\_ArcGIS\_QUOTAh

#### Projectos de DataStage:

- ExtracaoTrafego.dsx
- ExtracaoCartoes.dsx
- TOPConcorrenca.dsx

#### Ficheiros ArcGIS:

- Arcgisservices.exe (aplicação)
- DMGEO\_QUOTAS.mxd(mapa)

[anexo itens]

#### 4.3.3. Precedências

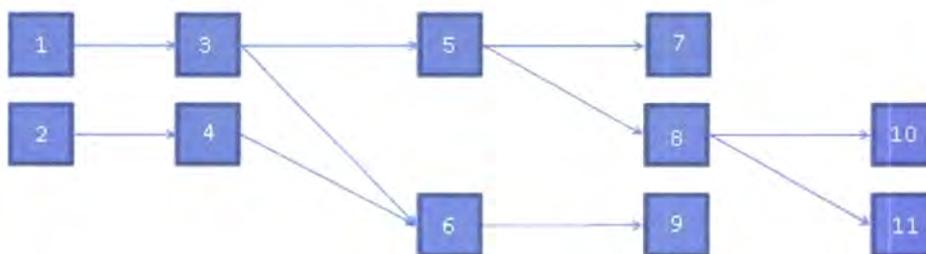


Figura 18 – Esquema de dependência das tabelas

1 - DMG\_AGG\_CRT\_CONC\_S1

❖ Contém a informação agregada de cartões da concorrência por semana

2 - DMG\_AGG\_CRT\_OPRA\_S1

- ❖ Contém a informação agregada de cartões da Operadora A por semana  
3 - DMG\_AGG\_CRT\_CONC\_12S\_S1
- ❖ Contém a informação agregada de cartões da concorrência de 12 semanas  
4 - DMG\_AGG\_CRT\_OPRA\_12S\_S1
- ❖ Contém a informação agregada de cartões da operadora A de 12 semanas  
5 - DMG\_AGG\_GEOREF\_CONC\_M1
- ❖ Contém os cartões que foram geo-referenciados da concorrência (mensal)  
6 - DMG\_FACT\_CRT\_OPERADOR
- ❖ Contém o valor da quantidade de cartões por tipo origem-destino (semanal)  
7 - DMG\_FACT\_CRT\_NGEO
- ❖ Contém o total de cartões não geo-referenciados/distribuídos por operador (mensal)  
8 - DMG\_AGG\_CRT\_ORIDEST\_M1
- ❖ Contém o valor da quantidade de cartões por tipo origem destino (mensal)  
9 - DMG\_FACT\_QUOTA\_OPERADOR
- ❖ Contém o valor da quota semanal por tipo origem destino (semanal)  
10 - DMG\_AGG\_APURA\_QUOTA\_M1
- ❖ Contém o valor da quota, da variação líquida e absoluta da quota por operadora  
11 - DMG\_AGG\_CRT\_OPER\_M1
- ❖ Contém o valor da quantidade de cartões por operadora (mensal)

#### 4.3.4. Processo de Instalação

##### ❖ Criação de Directorias

Para uma boa organização do DW é necessário distribuir correctamente os itens pelo seu tipo para que seja mais intuitivo o acesso a um em específico. Sendo assim são criadas novas directorias no sistema de ficheiros.

Máquina Unix:

- ~/DDL/DMGEO
- ~/SQL/DMGEO
- ~/SH/DMGEO
- ~/TMP/DMGEO
- ~/LOG/DMGEO

Máquina Windows onde está instalado a componente ArcGIS Desktop:

- %PATH\_DSK %\MXD\DMGEO
- %PATH\_DSK %\PY\DMGEO
- %PATH\_DSK %\TMP\DMGEO
- %PATH\_DSK%\LOG\DMGEO

Máquina WINDOWS (máquina remota) onde está instalado a componente ArcGIS Server:

- %PATH\_SRV%\MXD\DMGEO
- %PATH\_SRV%\EXE

%PATH\_DSK% - Path contendo a raiz onde será criada a estrutura de ficheiros da componente ArcGIS Desktop

%PATH\_SRV% - Path contendo a raiz onde será criada a estrutura de ficheiros da componente ArcGIS Server.

#### ❖ **Compilação**

Os procedimentos são criados num user e bd específicos. É através deles que são chamados a maioria dos itens criados e onde são efectuados todos os cálculos relacionados com o processo de modo a responder às regras impostas.

- PRC\_DMGEO\_AGG12SEMAN\_CONC.sql
- PRC\_DMGEO\_AGG12SEMAN\_OPRA.sql
- PRC\_DMGEO\_GREF\_ANTEN\_CONC.sql
- PRC\_DMGEO\_GREF\_ANTEN\_OPRA.sql
- PRC\_DMGEO\_APURAM\_CRT\_NAC.sql
- PRC\_DMGEO\_OPER\_QUOTA\_NAC.sql
- PRC\_DMGEO\_APURAM\_CRT\_ANT.sql
- PRC\_UPDATE\_DMG\_REF\_ANTENA.sql
- PRC\_DMGEO\_APURAM\_GIS.sql

#### ❖ **Criação de tabelas e respectivos índices**

Existem duas formas de o fazer: no unix através do sqlplus ou dentro da base de dados com um ide (Integrated Development Environment) como o toad usando o código que vem nos itens de \*.ddl e \*.idx. Para qualquer um dos casos é necessário ter acesso schema (user), base de dados e password.

Se as tabelas forem particionadas deve-se passar o parâmetro consoante a sua partição, por exemplo, se forem particionadas por dia deve ser passado o dia da partição como parâmetro.

#### ❖ **Operações de DML (Linguagem de Manipulação de Dados)**

Efectuam-se da mesma forma que as operações de ddl. As tabelas são carregadas com a informação que vem sob a forma de queries sql para poder carregar as respectivas tabelas.

- INSERT\_DMGMTD\_FACTOR\_ACTIVICO.sql
- INSERT\_DMGMTD\_FACTOR\_CORRECTIVO.sql
- INSERT\_DMGMTD\_FACTOR\_EXCLUSIVIDADE.sql
- INSERT\_DMGMTD\_FACTOR\_EXTRAP\_NACION.sql
- INSERT\_DMGMTD\_GAMA\_CONCORRENCIA.sql
- INSERT\_DMGMTD\_QUARENTENA.sql
- INSERT\_DMGMTD\_SAZONALIDADE.sql
- INSERT\_DMGMREF\_COD\_POSTAL\_BASE.sql
- INSERT\_DMGMTD\_EXCLUSAO.sql
- IS\_NUMBER.sql
- DMGMREF\_RECLASS\_RASTER\_INSERT.sql

#### ❖ Importação de ETL

Devem ser importados os seguintes mapeamentos de DataStage para posteriormente serem compilados e executados:

- ExtracaoTrafego.dsx
- ExtracaoCartoes.dsx
- TOPConcorrencia.dsx

#### ❖ Configurações ArcGIS

##### *ArcGIS Desktop*

Criar conexões às bases de dados. Para tal, criar uma ligação com o nome “Carto” seguindo os passos:

1. Iniciar a aplicação ArcCatalog (Start > programs > ArcGIS > ArcCatalog)
2. Na tabela de conteúdos (TOC) aceder a “Database Connections”

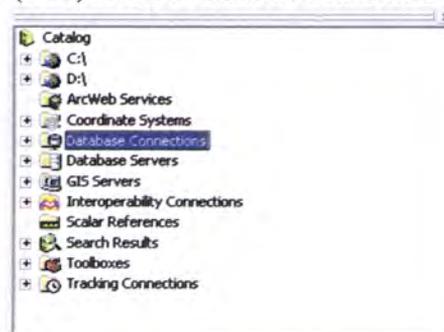


Figura 19 – Configuração ArcGIS Desktop 1

3. Definir uma nova conexão através da opção “add Spatial Database Connection”

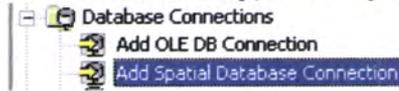


Figura 20 – Configuração ArcGIS Desktop 2

4. Criar a conexão à CARTO@BD

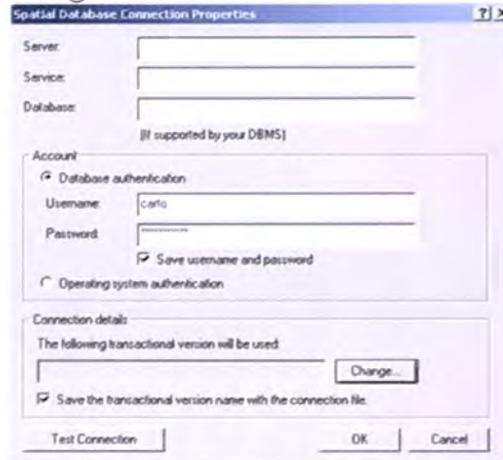


Figura 21 – Configuração ArcGIS Desktop 3

### *ArcGIS Server*

Criar o serviço de ArcGIS Server para o processo de cálculo de quotas de mercado e respectiva publicação de informação. Para isso são executados os seguintes procedimentos na máquina que contém o ArcGIS Server.

1. Acesso a componente ArcGIS Manager
2. Criar o serviço através dos seguintes passos:
  - a. Para criar um novo serviço de mapas, seleccionar a componente de “SERVICES”
  - b. Seleccionar a opção “Add New Service”
  - c. Utilizar os parâmetros relativos aos ficheiros MXD
    - i. Map Template : ~/MXD /DMGEO/DMGEO\_QUOTAS.mxd
    - ii. Data Frame : “Active Data Frame”
  - d. Terminar o processo de criação do serviço aceitando os parâmetros apresentados por defeito
3. Criação de Cache no serviço
  - a. Colocar o serviço como “started”
  - b. Editar o serviço de arcgisserver
  - c. No separador cache definir de acordo com a imagem seguinte

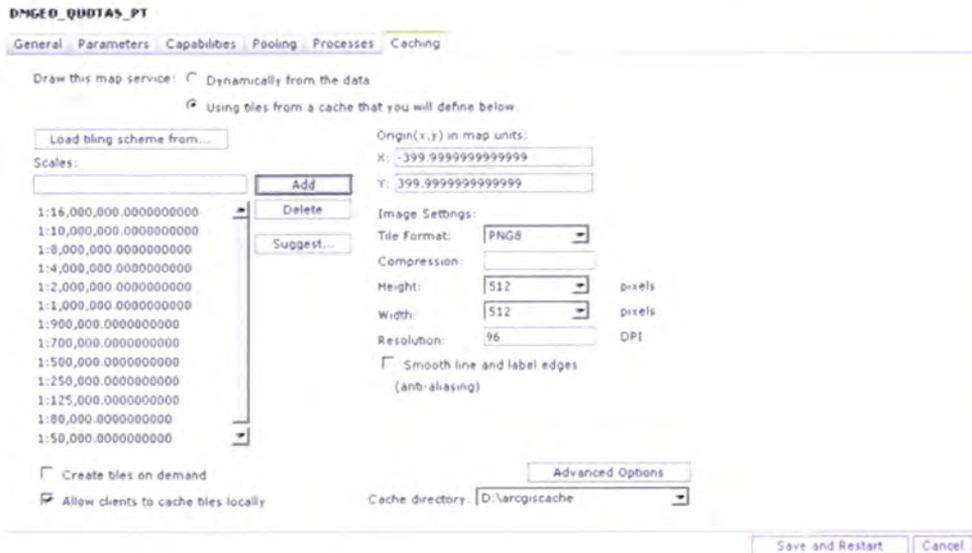


Figura 22 – Configuração ArcGIS Server 1

### Criação de aplicação Web (ArcGIS Server)

1. Criar uma nova aplicação Web que utiliza o serviço anteriormente criado e seleccionar a componente de “Applications”
2. Seleccionar a opção “Web Application”

#### Create Web Application - Application

Specify the name of the application you want to create.

Name:

Description:

[Hide Advanced Options](#)

Host:  Port:

URL of Host:

Figura 23 – Configuração ArcGIS Server 2

3. Definir os layers que a constituirá. Para tal executar os seguintes passos:
  - a) A partir do catálogo escolher o WEB Service definidos localmente



**Create Web Application - Page Properties**

Select the properties to apply to the web page.

Title text:

Theme:  Click to change theme

Web page links:

Name	URL
Add New Link	

Figura 27 – Configuração ArcGIS Server 6

#### 4.3.5. Sequência de Processamento

- ❖ Para a componente de ETL:

##### *Carregamento Inicial*

Processo de extracção do tráfego:

- dsTrafegoGeoCP3OPRAreph.sh

##### *Processo Diário*

Processo de actualização da tabela DMG\_REF\_ANTENA e da tabela DMG\_REF\_CELULA\_ANTENA:

- dsDMG\_REF\_ANTENAh.sh

Processo de extracção de cartões:

- dsExtracaoCDRsh.sh
- dsExtracaoCDRs\_BDLh.sh
- dsExtracaoCDRs\_OPRAh.sh
- dsAGG\_CRT\_OPRA\_S1h.sh
- dsAGG\_CRT\_CONC\_S1h.sh

##### *Processo Semanal*

Estes processos apenas deverão ser executados uma vez por semana. Os processos têm como dependência os processos diários, o que implica que sejam executados com a periodicidade de YYYYMMDD+7 relativamente ao início do processo diário.

Processo de agregação de 12 semanas:

- dsDMG\_AGG\_CRT\_OPRA\_12S\_S1h.sh
- dsDMG\_AGG\_CRT\_CONC\_12S\_S1h.sh

Processo de cálculo de quotas:

- dsDMG\_FACT\_CRT\_OPERADORh.sh
- dsDMG\_FACT\_QUOTA\_OPERADORh.sh

*Processo Mensal*

Processo de cálculo da geo-referenciação:

- dsQUOTA\_OPRA\_CELPRFh.sh
- dsGEOR\_CONC\_TOP\_Kh.sh
- dsDMG\_AGG\_GEOREF\_CONC\_M1h.sh
- dsDMG\_AGG\_CRT\_OPER\_M1h.sh

❖ Para a componente de GIS:

*Processo Mensal*

- dsDMG\_AGG\_APURA\_QUOTAh.sh
- dsDMG\_EXEC\_ArcGIS\_QUOTAh.sh
- dsDMG\_EXEC\_ArcGIS\_CACHEh.sh
- dsDMG\_EXEC\_ArcGIS\_ANTENAh.sh

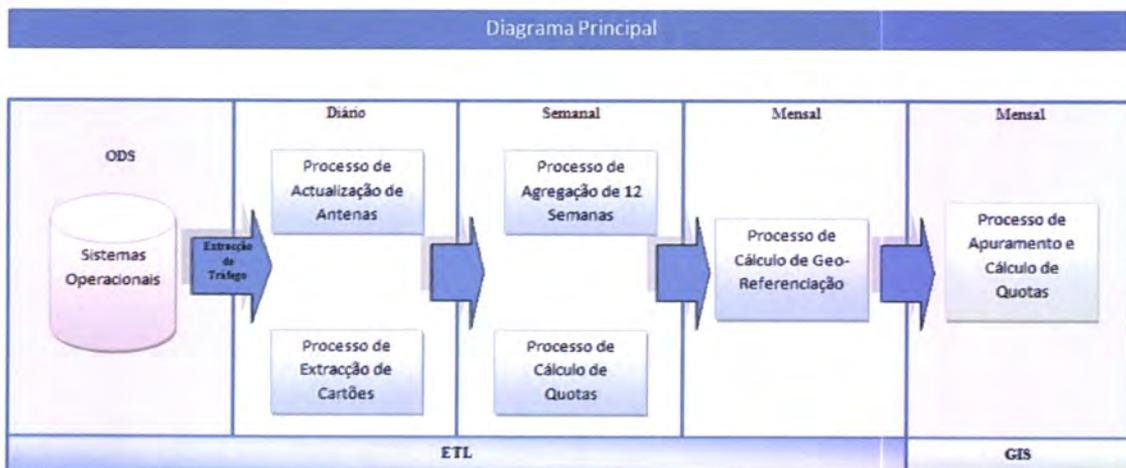


Figura 28 – Diagrama Principal do Processo

## 5. Resultados e Validações

<b>5.1. Objectivo .....</b>	<b>57</b>
<b>5.2. Ciclos de Testes.....</b>	<b>59</b>
<b>5.3. Tipo de testes .....</b>	<b>59</b>
<b>5.3.1. Testes de Carga.....</b>	<b>59</b>
<b>5.3.2. Testes de Utilização .....</b>	<b>60</b>
<b>5.3.3. Testes de Desempenho.....</b>	<b>60</b>
<b>5.3.4. Testes de Unidade .....</b>	<b>60</b>
<b>5.3.5. Testes de Integração .....</b>	<b>60</b>
<b>5.3.6. Testes Funcionais.....</b>	<b>60</b>
<b>5.4. Validações Efectuadas .....</b>	<b>61</b>
<b>5.4.1. Componente de ETL .....</b>	<b>61</b>
<b>5.4.2. Componente de GIS.....</b>	<b>62</b>

## 5. Validação dos Resultados

### 5.1. Objectivo

Assim que o desenvolvimento é concluído este passa pela fase de testes em que são validadas as suas duas componentes (ETL e GIS). A validação é feita através de diversos tipos de testes de modo a provar que os dados resultantes são coerentes, íntegros, fiáveis e correspondem às expectativas pretendidas, descritas nos requisitos impostos pelo cliente.

É então elaborado um relatório que descreve pormenorizadamente o processo de validação e as metodologias utilizadas nos testes de modo a que fique claro para o cliente que os resultados estão correctos.

As validações dos resultados terão por base os dados carregados nas tabelas finais e geração de mapas por quota, variação líquida e variação de quota.

<b>DMG_AGG_APUARA_QUOTA_MI</b>	
<b>Cod_mes</b>	Number(6)
<b>Cod_antena</b>	Number(8)
<b>Quota_opra_ant</b>	Number(10,2)
<b>Quota_oprb_ant</b>	Number(10,2)
<b>Quota_oprc_ant</b>	Number(10,2)
<b>Quota_oprd_ant</b>	Number(10,2)
<b>Quota_opre_ant</b>	Number(10,2)
<b>Quota_oprf_ant</b>	Number(10,2)
<b>Var_liq_opra</b>	Number(15)
<b>Var_liq_oprb</b>	Number(15)
<b>Var_liq_oprc</b>	Number(15)
<b>Var_liq_oprd</b>	Number(15)
<b>Var_liq_opre</b>	Number(15)
<b>Var_liq_oprf</b>	Number(15)
<b>Var_quota_opra</b>	Number(10,2)
<b>Var_quota_oprab</b>	Number(10,2)
<b>Var_quota_oprc</b>	Number(10,2)
<b>Var_quota_oprd</b>	Number(10,2)
<b>Var_quota_opre</b>	Number(10,2)
<b>Var_quota_oprf</b>	Number(10,2)

Tabela 3 – Tabela de agregados com apuramento de quotas por operadora

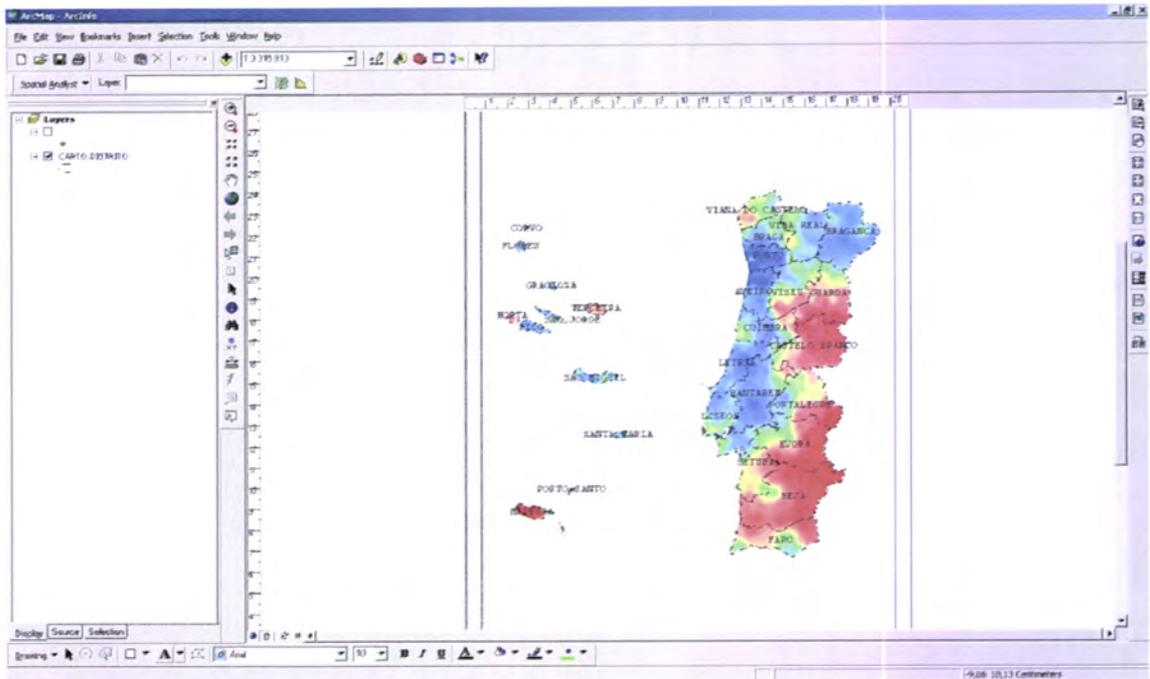


Figura 29 – Mapa por quota da operadora A

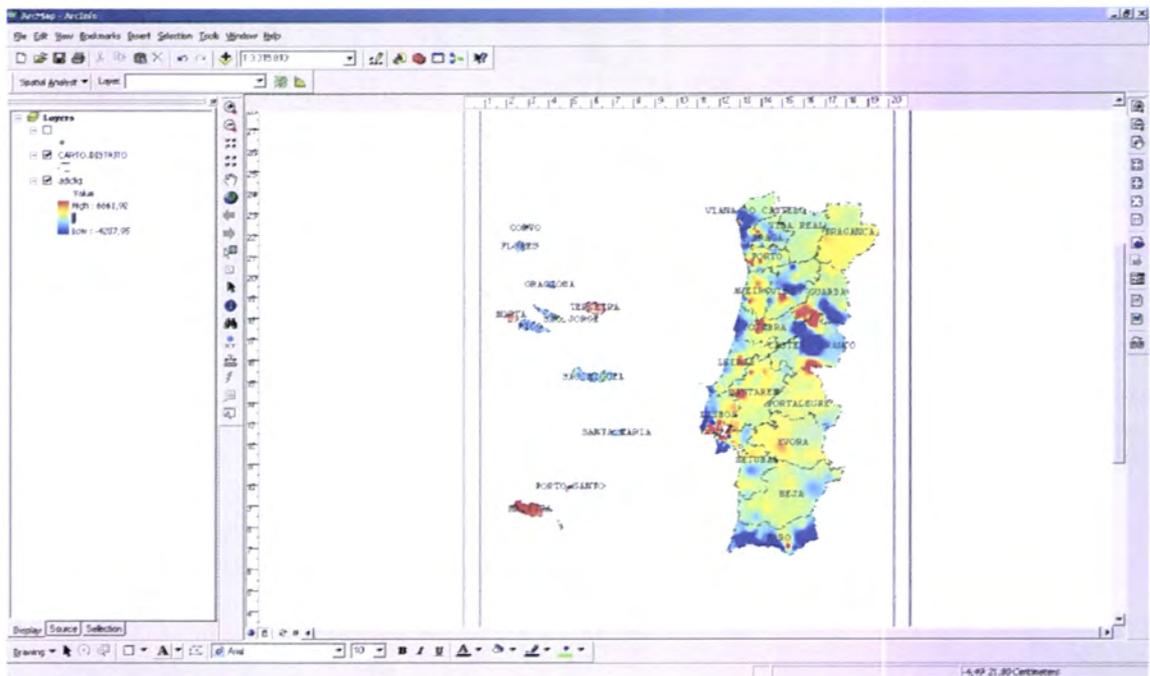


Figura 30 – Mapa de variação líquida da operadora A

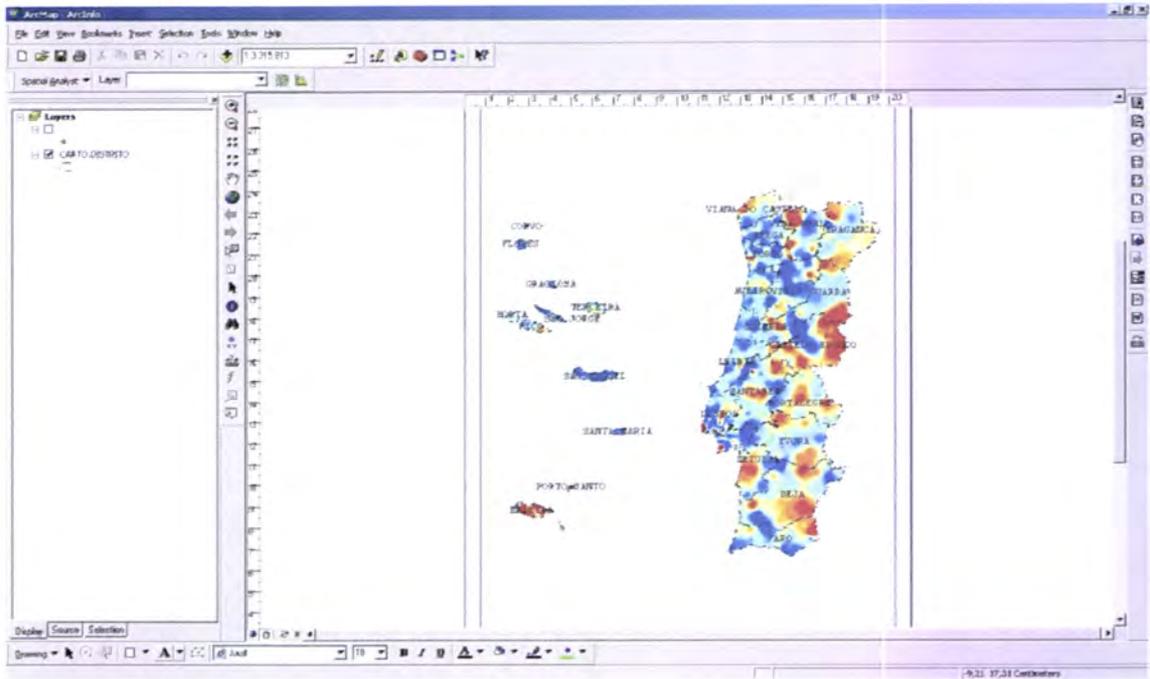


Figura 31 – Mapa de variação de quota da operadora A

## 5.2. Ciclo de testes

Um teste consiste em avaliar os dados ou uma amostra dos mesmos num ambiente igual ou semelhante aquele em que vão ser utilizados, executando de forma controlada e validando se obtém o resultado esperado.

Um ciclo de testes é normalmente caracterizado por uma sequência de casos de testes que aprovam ou rejeitam um determinado requisito. Um relatório apresenta pelo menos um ciclo de testes para cada requisito que é testado. Apesar de existirem alguns tipos de teste os casos de teste devem basear-se em particular nos três principais tipos que são: desempenho, funcionais e carga.

## 5.3. Tipo de Testes

Existem vários tipos de testes mas os principais são os testes de: carga, utilização, desempenho, unidade, integração e funcionais. É através deles, não necessariamente todos, que se consegue garantir a qualidade, eficácia e fiabilidade dos desenvolvimentos.

### 5.3.1. Testes de Carga

Avaliam o comportamento dos sistemas quando estes estão expostos a enormes cargas de processamento a ponto de não o conseguirem executar. É utilizado para avaliar e validar os limites que o processo atinge na realização das suas operações. Com base nessa avaliação consegue-se calcular os recursos e quantidade de dados a processar de modo a que o sistema se mantenha estável.

### **5.3.2. Testes de Utilização**

Servem para validar uma interface, é um método para detectar problemas entre o utilizador e o sistema. O utilizador ou grupo de utilizadores ao explorar a interface pode deparar-se com problemas/dificuldades que devem ser reportados para que esta seja mais intuitiva possível.

### **5.3.3. Testes de Desempenho**

Consta em avaliar o tempo de resposta do sistema, isto é, validar se o tempo de execução dos processos necessários a dar a resposta pretendida pelo sistema estão dentro dos padrões aceitáveis definidos previamente. Esta validação tem de ser feita num ambiente o mais próximo possível do ambiente de produção uma vez que é quase impossível recriá-lo em testes.

### **5.3.4. Testes de Unidade**

Quando se utiliza este tipo de testes tem-se como objectivo validar componentes isolados do desenvolvimento, isto é, para validar uma função em específico é necessário que o processo seja executado e também ter acesso ao código fonte na íntegra do mesmo o que faz com que este tipo de testes sejam dinâmicos.

### **5.3.5. Testes Integração**

Quanto a este tipo de testes, são usados para validar uma sequência de funções ou serviços com as dependências reais de modo a garantir que funcionam em conjunto. São normalmente utilizados para avaliar processos em que interagem modelos diferentes para evitar executar cada um deles, o que faria perder muito tempo desnecessariamente, o processo é dividido em pequenas fracções e vai-se testando cada uma delas sendo assim mais fácil de identificar o erro.

### **5.3.6. Testes Funcionais**

Todos os projectos vêm com documentação anexada que inclui os requisitos propostos pelo cliente e que devem ser implementados pela equipa de desenvolvimento. Este tipo de testes tem como objectivo

garantir que os requisitos estão correctamente implementados, com base na informação da documentação são efectuados os testes necessários para garantir a sua validação.

## 5.4. Validações Efectuadas

### 5.4.1. Componente de ETL

#### Caso de teste 1 – Execução do processo e verificação dos logs

- Garante a execução do processo implementado, o correcto output para os ficheiros de log e que não permanecem objectos no FS ou na base de dados que sejam desnecessários.

#### Caso de teste 2 – Verificação do carregamento das tabelas da operadora A, concorrência, Quota/Operador

- Garante que a estrutura das tabelas se encontra de acordo com o descrito na documentação do pedido.

#### Caso de teste 3 – Contagem de duplicados nas tabelas agregadas

- Verifica a não existência de registos duplicados nas tabelas.

#### Caso de teste 4 – Análise da Integridade Referencial

- Verifica que todas as ocorrências distintas de alguns dos campos presentes nas tabelas do modelo, existem nas respectivas tabelas de referência.

#### Caso de teste 5 – Validação da geo-referenciação sobre os números da concorrência

- Valida o processo de geo-referenciação da concorrência identificando os números distintos da concorrência.

#### Caso de teste 6 – Verificação do valor da soma das quotas

- Valida que a soma das percentagens das quotas é igual a 100, isto para garantir que não existem quotas mal calculadas ou em falta.

#### Caso de teste 7 – Verificação das quotas nacionais sem sazonalidade das três principais operadoras

- Analisa a contagem das quotas relativas a um mês em que não existe sazonalidade, ou seja, qualquer um excepto Natal e Verão.

#### Caso de teste 8 – Verificação das quotas nacionais com sazonalidade das três principais operadoras

- Analisa a contagem das quotas relativas ao mês de sazonalidade.

---

Caso de teste 9 – Verificação das quotas nacionais com quarentena das três principais operadoras

- Analisa a contagem das quotas relativas a um período de quarentena.

Caso de teste 10 – Validação do critério de exclusão de cartões do tipo Banda Larga

- Valida a distribuição das quotas efectuadas pelo tráfego de banda larga contando todos os cartões únicos e verifica a sua exclusão.

Caso de teste 11 – Validação do critério de exclusão de cartões do tipo “Serviço em casa”

- Valida a distribuição das quotas efectuadas pelo tráfego de “Serviço em casa” contando todos os cartões únicos e verifica a sua exclusão.

Caso de teste 12 – Análise de quota por dimensão geográfica

- Verifica a quota para uma determinada dimensão geográfica analisando a sua distribuição e comparando com as fontes.

Os resultados dos casos de teste são apresentados nos anexos. Alguns serão encurtados porque se tornam repetitivos, como por exemplo, validar a estrutura ou o mapeamento de uma tabela será sempre pelo mesmo processo para todas as tabelas e como são muitas não existe necessidade de estar a repetir o mesmo teste inúmeras vezes.

[anexo testes\_etl]

#### 5.4.2. Componente de GIS

Caso de teste 1 – Execução do processo e verificação dos logs

- Garante a execução do processo implementado, o correcto output para os ficheiros de log e que não permanecem objectos no FS ou na base de dados que sejam desnecessários.

Caso de teste 2 – Verificação da estrutura das tabelas com e sem dimensão geográfica

- Garante que a estrutura das tabelas se encontra de acordo com o descrito na documentação do pedido.

Caso de teste 3 – Verificação dos mapeamentos das tabelas de dimensão geográfica

- Verifica que as tabelas são correctamente carregadas de acordo com as fontes especificadas na documentação.

Caso de teste 4 – Validação da correcta criação do serviço e da aplicação do serviço ArcGIS Server

- Verifica que a criação e aplicação do serviço ArcGIS Server respeitam todas as normas que estão estipuladas na documentação.

Caso de teste 5 – Confrontação da visualização do MXD no Desktop com o MXD no ArcGIS Server

- Compara e valida os resultados (mxd) do ArcGIS Desktop com os do ArcGIS Server.

[anexo testes\_gis]

## 6. Conclusão

6.1. Conclusão Final.....	65
6.2. Avaliação Crítica.....	65
6.3. Trabalho Futuro .....	66

### 6.1. Conclusão Final

Um projecto de DW tem sempre de se ir adaptando às necessidades apresentadas pela empresa com o fim de as conseguir colmatar o melhor possível, o que faz com que existam sempre novos requisitos e/ou pressupostos apresentados pelo negócio, tais como, novas funcionalidades prontas para serem implementadas. É por isto que jamais se pode afirmar que um DW está terminado.

Embora desenvolver um projecto de DW seja algo financeiramente dispendioso, num futuro não muito longínquo é de prever que todas as empresas tenham o seu próprio DW e que passem a aceder aos seus dados através de sistemas e diferentes aplicações, lucrando com a informação adquirida e posteriormente guardada em base de dados.

A aplicação de tecnologias como a de geo-referenciação aliada à evolução e optimização dos processos de DW são claramente apostas que devem ser feitas por parte de empresas com um grande número de clientes, nomeadamente as de telecomunicações.

Na actualidade o acesso à informação por parte das mesmas é cada vez mais rápido e eficaz e por isso não se justifica que fiquem para trás, não investindo na informação que este tipo de inovações pode fornecer.

A vantagem competitiva que se pode tirar a partir da análise dos dados devolvidos pelo processo é evidente. Permite á empresa que os utiliza saber onde e como actuar no mercado já que disponibiliza informação válida para vários departamentos como o de venda, publicidade ou marketing.

Não é um exagero dizer que apostar na geo-referenciação tem um risco baixo, uma vez que o auxílio prestado aos vários departamentos da companhia que aí investe faz com que esta receba os seus dividendos mais tarde.

### 6.2. Avaliação Crítica

A integração neste projecto “fechado” de uma grande empresa de telecomunicações permite avaliar o modo como se processa o desenvolvimento de um projecto com impacto significativo para empresa da seguinte forma:

- Devido ao facto de o desenvolvimento, testes e produção serem efectuados em ambientes diferentes, isto é, as bases de dados e os schemas não são os mesmos, há inúmeras situações de erro por parte de quem está a desenvolver fazendo atrasar o planeamento que foi feito.
- Para se conseguir desenvolver e testar a parte gráfica, ou seja, para se poder criar o projecto no ArcGIS é necessária obter a licença do produto, visto que esta é uma licença flutuante, que

significa que só  $x$  utilizadores a podem estar a usar ao mesmo tempo. Muitas vezes era necessário esperar ou verificar quem a estava a utilizar para fazer log off.

- Algumas vezes existem dificuldades em conseguir implementar os requisitos estabelecidos pelo negócio (nomeadamente as regras do desenvolvimento) o que requer a necessidade de reunir diversas vezes com o gestor de projecto e este por sua vez com o interlocutor do negócio.
- Como consequência de serem necessários objectos de diferentes ambientes houve a necessidade de solicitar às equipas de produção a replicação dos mesmos para os ambientes de desenvolvimento. Estas situações levam tempo e muitas vezes existe a necessidade de prestar esclarecimentos para justificar os pedidos.
- Devido à complexidade do uso da ferramenta ArcGIS, especialmente para conseguir a sua correcta configuração, foi necessário o apoio de um especialista.

### 6.3. Trabalho Futuro

Existem alguns aspectos que podem fazer evoluir este projecto, nomeadamente quando se trata da informação respeitante à concorrência.

Em vez de a informação ser obtida através do cálculo de extrapolação com base na informação do tráfego da operadora A, a informação dos números da concorrência poderia ser extraída através do log das antenas porque todas as antenas conseguem distinguir todos os números activos que estão na sua área de abrangência. Desta forma tanto o processo para determinar o universo da concorrência como a geo-referenciação dos números da concorrência seria mais fidedigno e os cálculos mais simplificados.

A geo-referenciação que é feita pelas diferentes áreas também poderá ser melhorada para que do ponto de vista do utilizador seja mais fácil identificar as zonas que se pretendem geo-referenciar, ou seja, usando processos gráficos que permitissem definir os limites das áreas de modo mais preciso e fiável especialmente quando se aumenta o zoom nessa zona.

Para além disso seria importante conseguir reduzir a margem de erro da geo-referenciação das antenas de modo a que se consiga realmente a posição exacta das mesmas. Embora nas bases de dados se encontrem as coordenadas, a geo-referenciação feita através do ArcGIS faz com que haja uma divergência na sua localização de alguns metros.

---

## 7. Bibliografia

- Caldeira, C. P. (2008). *Data Warehousing Conceitos e modelos com exemplos práticos*. Lisboa: Sílabo
- Câmara, G.(acedido em Agosto de 2010) *Arquitetura de Sistemas de Informação Geográfica* [on-line]  
<http://www.dpi.inpe.br/gilberto/livro/introd/cap3-arquitetura.pdf>
- ESRI (acedido em Agosto de 2010) [on-line]<http://www.esriportugal.pt/>
- Gama, J. (acedido em Agosto de 2010) Capítulo: *Olap Bases de Dados Multidimensionais* [on-line]  
[http://www.liaad.up.pt/~jgama/Mestrado/Bases\\_Dados/olap.pdf](http://www.liaad.up.pt/~jgama/Mestrado/Bases_Dados/olap.pdf)
- IDW (acedido em Agosto de 2010). *IDW Integration, Development and Warehousing* [on-line]  
<http://www.idw.pt>
- Inmon, W. H. (2002). *Building the Data Warehouse 3<sup>rd</sup> Edition*. New York: John Wiley & Sons, Inc
- Kimball & Ross, R. & M. (2002). *The Data Warehouse Lifecycle Toolkit 2<sup>nd</sup> Edition*. New York: John Wiley & Sons, Inc.
- Mattison, R. (1997). *Data Warehousing and Data Mining for Telecommunication*. Norwood, Massachusetts: Artech House, Inc
- Panegassi, L. (2006). *Data Warehouse: Monografia apresentada à disciplina Trabalho de Conclusão de Curso, do Curso de Ciência da Computação*. São Paulo: formato pdf
- Silvers, F. (2008). *Building and Maintaining a Data Warehouse*. Boca Raton, Florida: Taylor & Francis Group, LLC
- Tanaka, A.(1997), *Sistemas de Apoio à Inteligência do Negócio* [online]  
<http://www.uniriotec.br/~tanaka/SAIN/>
- Santos & Silva (acedido em Agosto de 2010) *Linguagem Shell Script* [on-line]  
<http://fit.faccat.br/~jonis/shellsript.pdf>
- Vercellis, C.(2009) *Business Intelligence: Data Mining and Optimization for Decision Making*. Reino Unido: John Wiley & Sons, Ltd

Wikipedia (acedido em Julho de 2010), *Data Warehouse* [online]  
[http://pt.wikipedia.org/wiki/Data\\_Warehouse](http://pt.wikipedia.org/wiki/Data_Warehouse)

Wikipedia (acedido em Julho de 2010), *Sistema de Informação Geográfica* [online]  
[http://pt.wikipedia.org/wiki/Sistema\\_de\\_informa%C3%A7%C3%A3o\\_geogr%C3%A1fica](http://pt.wikipedia.org/wiki/Sistema_de_informa%C3%A7%C3%A3o_geogr%C3%A1fica)

## 8. Anexos

[anexo processos]

### ❖ Processo diário de extracção de tráfego

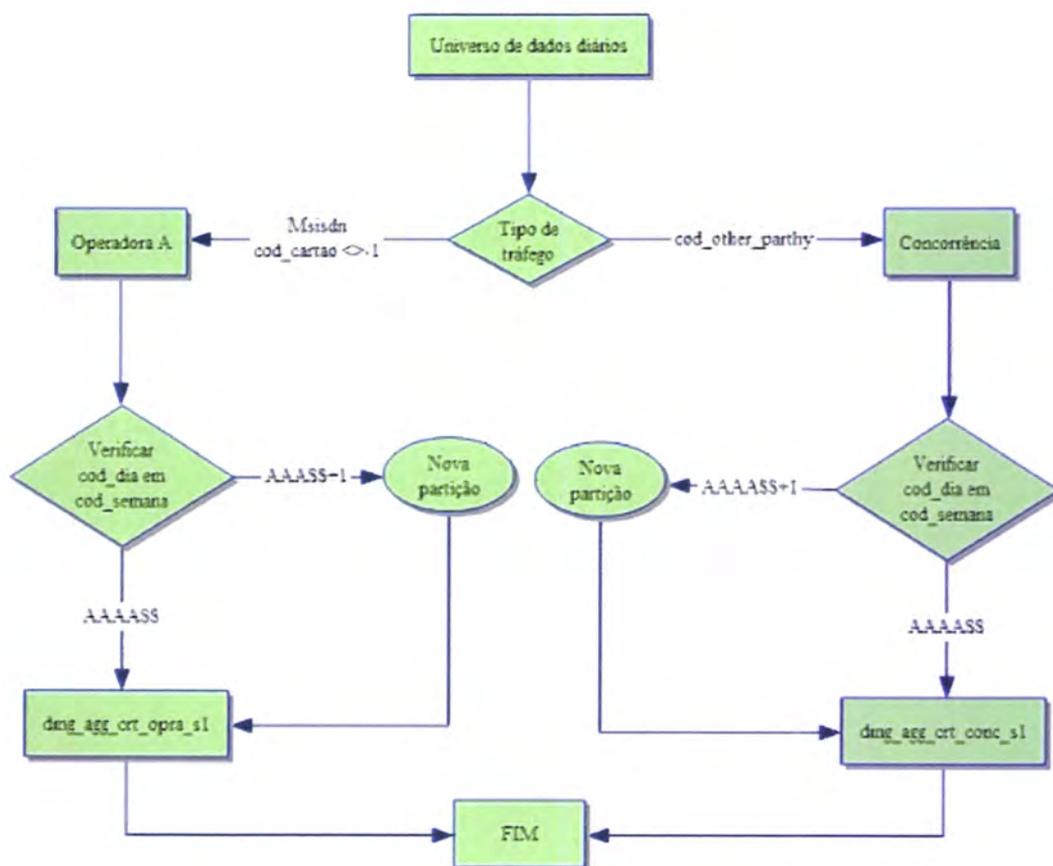


Figura 32 – Extracção de tráfego

❖ Processo semanal de agregação de 12 semanas da Operadora A para cálculo de quotas

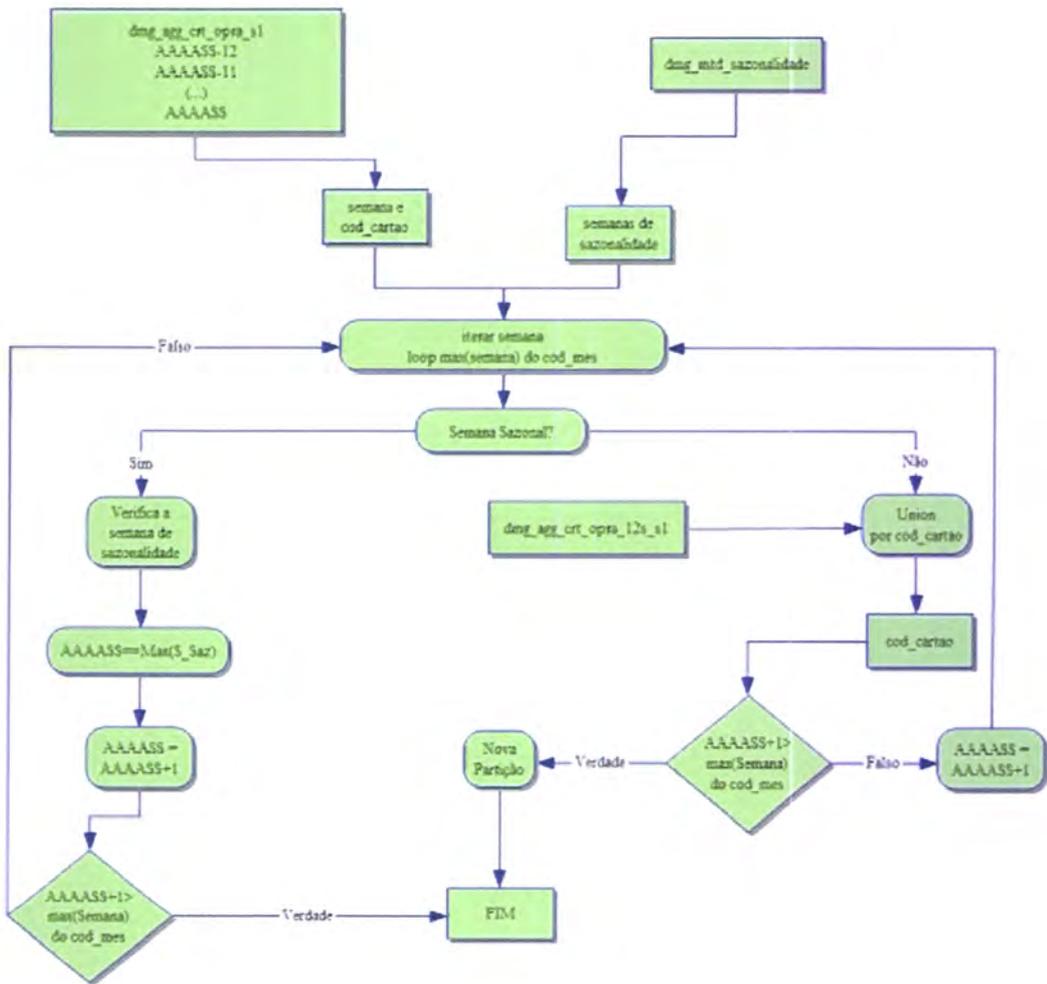


Figura 33 - Agregação de 12 semanas da Operadora A

❖ Processo semanal de agregação de 12 semanas da Operadora A para cálculo de quotas

Em termos de esquematização é um processo idêntico ao anterior com a diferença de se usar o `cod_tipo_orig_dest` para identificar que operador da concorrência se está a tratar e em vez de se utilizar o `cod_cartao` tem de se recorrer ao `msisdn` da concorrência.

❖ Processo mensal de geo-referenciação da Operadora A para cálculo de quota

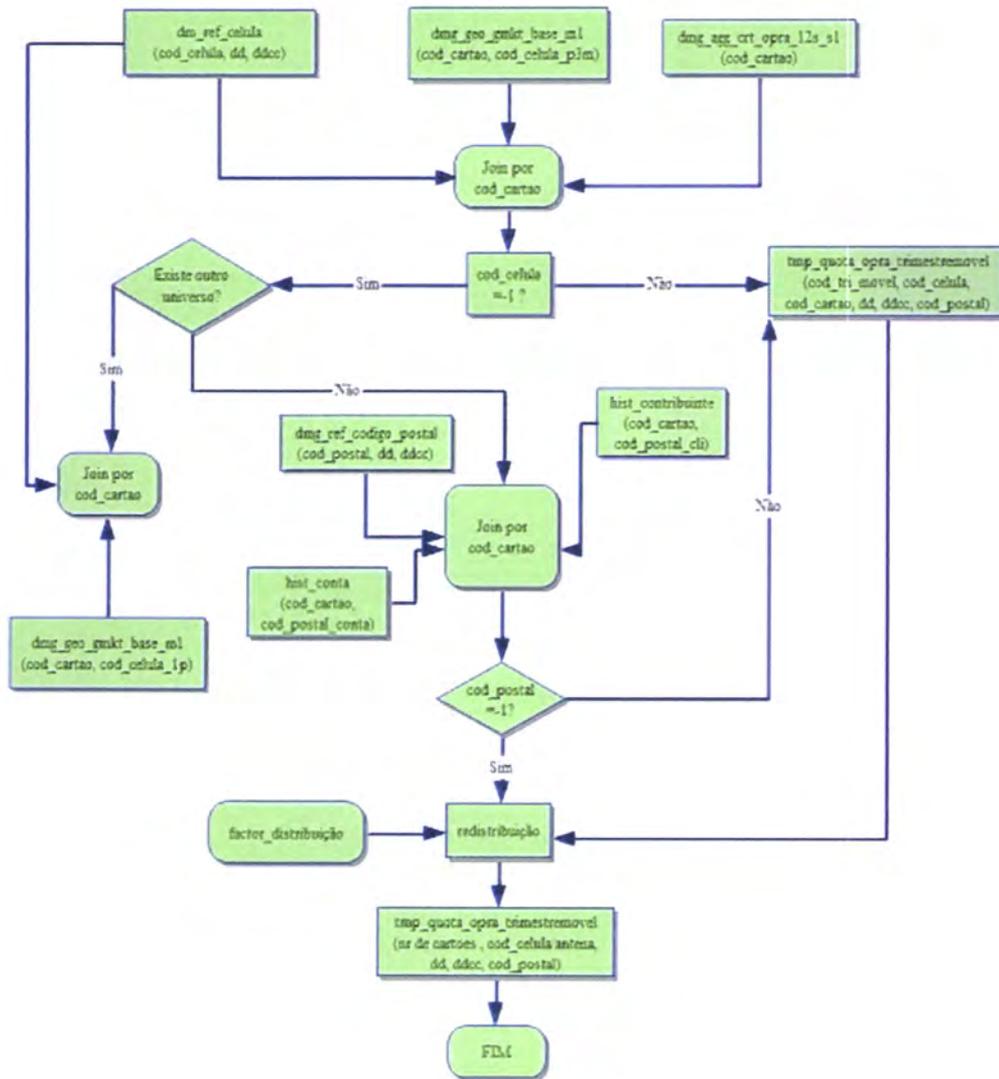


Figura 34 – Geo-referenciação da Operadora A

❖ Processo mensal de geo-referenciação da concorrência para cálculo de quota

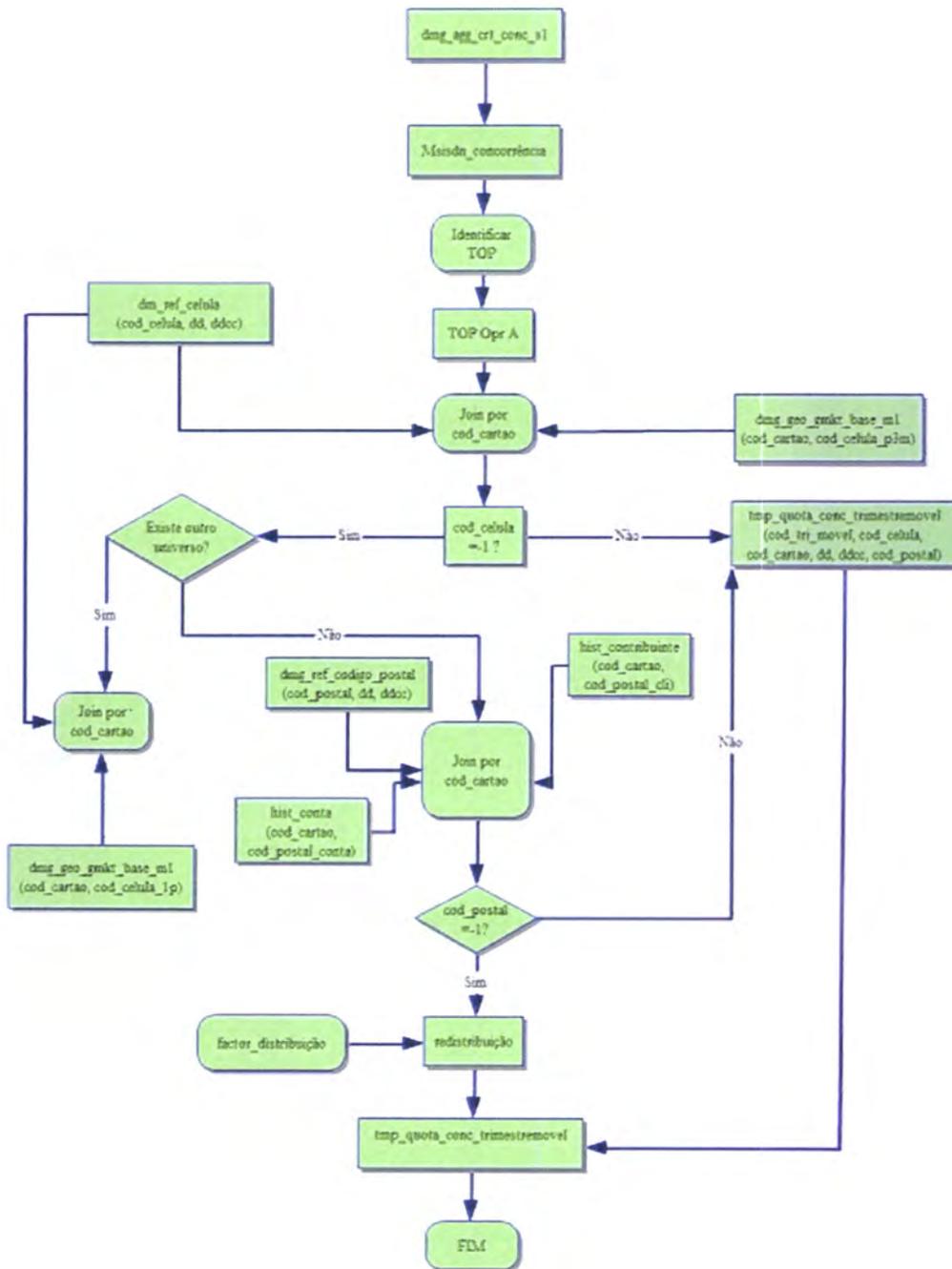


Figura 35 – Geo-referenciação da concorrência

[anexo itens]

Devido ao facto de os procedimentos serem muito extensos, anexo um exemplo de cada item. Seria impossível anexar todo o código do trabalho uma vez que a dissertação tem um limite máximo de páginas que não pode ser excedido.

#### ❖ dsDMG\_REF\_ANTENAh.sh

```
#!/bin/ksh

#####
# Processo: dsDMG_REF_ANTENAh.sh
# Descricao: ShellScript para a execução dos procedimentos de actualização
#            das tabelas DMG_REF_ANTENA e DMG_REF_CELULA_ANTENA
#
# Parametro: [AAAAMDD] -> Dia de Carregamento (Obrigatório) {DIA_A_TRATAR}
#            [XXXXXXXX] -> Database link
#####

# Preparacao das Variaveis de Ambiente
#####
. $HOME/.dm_env

# Definicao de variaveis do processo
#####
PROC=`echo $0 | awk -F "/" '{ print \$NF}'`

if [ $# -ne 2 ]
then
  echo "A quantidade de parametros não foi a esperada"
  echo "dsDMG_REF_ANTENAh.sh <DIA_A_TRATAR> <DBLINK_DW>"
  exit -1
else
  DIA_A_TRATAR=$1
  PROCEDIMENTO="PRC_UPDATE_DMG_REF_ANTENA"
  DBLINK_DW=$2
fi

TBS_TEMP="TBS_DATA_TMP"

#####
# Definicao de variaveis
#####
#cd $HOME

parORADSN=`.obtainDSpar parORADSNMGE0`
parORAPWD=`.obtainDSpar parORAPWDDMGE0`
OWNER_GEO=`.obtainDSpar parORAUIDDMGE0`
OWNER_DW=`.obtainDSpar parORAUIDDW`

#coneccao BD
ORADMGE0=${OWNER_GEO}/${parORAPWD}@${parORADSN}

LOG_FIC=$LOG_DIR/DMGEO/$PROC.$PROCEDIMENTO.$DIA_A_TRATAR.log
FIC_SQL_LOG=${LOG_DIR}/DMGEO/${PROC}.${PROCEDIMENTO}.${DIA_A_TRATAR}.sql.log
FIC_SQL=$TMP_DIR/DMGEO/tmp_$PROC.$PROCEDIMENTO.$DIA_A_TRATAR.sql

#####
# Passo 1: Executar procedure
#####

echo "Processo $PROC $PROCEDIMENTO $DIA_A_TRATAR iniciado `date`." > $LOG_FIC
echo "Processo $PROC $PROCEDIMENTO $DIA_A_TRATAR iniciado `date`."

echo "Passo 1. $PROC $PROCEDIMENTO $DIA_A_TRATAR `date`." >> $LOG_FIC

echo "set serverout on;\n " \
```

```

    "exec $PROCEDIMENTO('$DIA_A_TRATAR', '$OWNER_DW', '$DBLINK_DW', '$OWNER_GEO',
'$TBS_TEMP');\n quit\n" > $FIC_SQL

sqlplus -s > ${FIC_SQL_LOG} << END
${ORADMGEO}
@${FIC_SQL}
END

if [ `egrep -c '(ERRO|SP2-[0-9]+:|ORA-[0-9]+:|O/S Message)' $FIC_SQL_LOG` -ne 0 ]
then
echo "ERRO: Erro a correr procedimento $PROCEDIMENTO $DIA_A_TRATAR." >> $LOG_FIC
echo "ERRO: Erro a correr procedimento $PROCEDIMENTO $DIA_A_TRATAR."
cat $FIC_SQL_LOG >> $LOG_FIC
echo "Processo $PROC $PROCEDIMENTO $DIA_A_TRATAR ABORTADO.`date`." >> $LOG_FIC
cat $LOG_FIC
exit -1
fi

#####
# Passo 2. Limpeza de Ficheiros temporarios
#####

rm -f ${FIC_SQL}

rm -f ${FIC_SQL_LOG}

echo "Processo $PROC $PROCEDIMENTO $DIA_A_TRATAR terminado com sucesso.`date`" >>
$LOG_FIC
echo "Processo $PROC $PROCEDIMENTO $DIA_A_TRATAR terminado com sucesso.`date`"

exit 0

```

#### ❖ PRC\_UPDATE\_DMG\_REF\_ANTENA.sql

```

CREATE OR REPLACE PROCEDURE PRC_UPDATE_DMG_REF_ANTENA
(p_cod_dia NUMBER,
 p_owner_dw VARCHAR2 default 'PDDW',
 p_dblink_dw VARCHAR2 default 'DBLINK_DW',
 p_owner_dmg VARCHAR2 default 'DMGPROD',
 p_tablespace_tmp VARCHAR2 default 'TBS_DATA_TMP')
authid current_user AS

/*****
/* DESCRIÇÃO: procedure de actualização da tabela DMG_REF_ANTENA */
*****/

-- Variáveis --
nome_procedure      VARCHAR2(30) := 'PRC_UPDATE_DMG_REF_ANTENA';
passo                VARCHAR2(20);
sql_code             NUMBER(10);
query                VARCHAR2(5120);

-- TABELAS DE ORIGEM DE REREFERENCIA DE DW
v_tab_ref_celula     VARCHAR2(30) := 'REF_CELULA';

-- TABELAS DE REREFERENCIA DE DMGEO
v_tab_ref_antena     VARCHAR2(30) := 'DMG_REF_ANTENA';
v_tab_ref_celula_antena VARCHAR2(30) := 'DMG_REF_CELULA_ANTENA';

-- TABELAS TEMPORÁRIAS
v_tabela_temporaria1 VARCHAR2(30) := 'TMP_DMG_REF_ANTENA';
v_tabela_temporaria2 VARCHAR2(30) := 'TMP_ALL_CELULA_ANTENA';
v_tabela_temporaria3 VARCHAR2(30) := 'TMP_NEW_CELULA_ANTENA';
--utilizadores e owners da bd
v_owner_dmg VARCHAR2(20) := '';
v_dblink_dw VARCHAR2(30) := '';
v_owner_dw VARCHAR2(20) := '';

--VARIABLES DE PROCESSOS
v_max_cod_antena number(15);
v_max_cod_celula_antena number(15);

```

```

v_existe_ant_novas NUMBER(4);
v_seq_antena VARCHAR2(30) := 'SEQ_COD_ANTENA';
v_seq_celula_antena VARCHAR2(30) := 'SEQ_COD_CELULA_ANTENA';
v_sequence NUMBER(2);
BEGIN

dbms_output.enable(100000);

-----
passo := 'P001'; --Tratamento do owner de datamart geografico
-----
v_owner_dmg := '';

IF p_owner_dmg IS NOT NULL then
  v_owner_dmg:= trim(p_owner_dmg);
  IF substr(v_owner_dmg,-1,1) <> '.' THEN
    v_owner_dmg := v_owner_dmg || '.';
  END IF;
END IF;

-----
passo := 'P002'; -- Tratamento do DATABASELINK
-----
v_dblink_dw := '';

IF p_dblink_dw IS NOT NULL THEN
  v_dblink_dw := TRIM(p_dblink_dw);
  IF SUBSTR(v_dblink_dw, 1, 1) <> '@' THEN
    v_dblink_dw := '@' || v_dblink_dw ;
  END IF;
END IF;

-----
passo := 'P003'; --Tratamento do owner de datawarehouse
-----
v_owner_dw := '';

IF p_owner_dw IS NOT NULL then
  v_owner_dw:= trim(p_owner_dw);
  IF substr(v_owner_dw,-1,1) <> '.' THEN
    v_owner_dw := v_owner_dw || '.';
  END IF;
END IF;

-----
passo := 'P010'; --Verificar max cod_antena
-----
v_max_cod_antena:=0;

query:= 'select NVL(max(cod_antena),0) ' ||
        ' from ' ||v_owner_dmg|| v_tab_ref_antena ;

EXECUTE IMMEDIATE query INTO v_max_cod_antena;

-----
passo := 'P011'; --Verificar max cod_celula_antena
-----
v_max_cod_celula_antena:=0;

query:= 'select NVL(max(cod_celula_antena),0) ' ||
        ' from ' ||v_owner_dmg|| v_tab_ref_celula_antena ;

EXECUTE IMMEDIATE query INTO v_max_cod_celula_antena;

-----
passo := 'P020'; --criar sequencia e inicializar no max(cod_antena)
-----
v_sequence:=0;
query:= 'SELECT COUNT( SEQUENCE_NAME) FROM USER_SEQUENCES ' ||
        'WHERE SEQUENCE_NAME LIKE ''' ||v_seq_antena||
        '''';
dbms_output.put_line(query);
EXECUTE IMMEDIATE query INTO v_sequence;

dbms_output.put_line(v_sequence);

```



```

    || ' WHERE a.VAL_LATITUDE<>-1 AND a.VAL_LONGITUDE<>-1 AND a.COD_CONCELHO<>-1 AND
a.COD_DISTRITO<>-1 AND a.COD_FREGUESIA<>-1 AND a.COD_FREGUESIA<>0 '
    || 'GROUP BY a.VAL_LATITUDE, a.VAL_LONGITUDE, LPAD(a.COD_FREGUESIA,6,'0'),
LPAD(SUBSTR(LPAD(a.COD_FREGUESIA,6,'0'),0,4),4,'0'),
LPAD(SUBSTR(LPAD(a.COD_FREGUESIA,6,'0'),0,2),2,'0'))';

```

```
EXECUTE IMMEDIATE query INTO v_existe_ant_novas;
```

```
-----
passo := 'P040'; --criar tabela temporária com as novas antenas
-----
```

```
IF v_existe_ant_novas > 0 THEN
```

```
pgbdt10q(v_tabela_temporaria1 || '_' || p_cod_dia);
```

```

query := 'CREATE TABLE ' || v_tabela_temporaria1 || '_' || p_cod_dia
        || ' TABLESPACE ' || p_tablespace_tmp || ' STORAGE (INITIAL 50M NEXT 50M)
PCTFREE 1 NOLOGGING AS '
        || ' SELECT a.VAL_LATITUDE, a.VAL_LONGITUDE ,LPAD(a.COD_FREGUESIA,6,'0')
as      COD_FREGUESIA,      LPAD(SUBSTR(LPAD(a.COD_FREGUESIA,6,'0'),0,4),4,'0')      as
COD_CONCELHO, LPAD(SUBSTR(LPAD(a.COD_FREGUESIA,6,'0'),0,2),2,'0') as COD_DISTRITO '
        || ' FROM '
        || v_owner_dw
        || v_tab_ref_celula
        || v_dblink_dw
        || ' a, '
        || v_owner_dmg
        || v_tab_ref_antena
        || ' b '
        || ' WHERE a.VAL_LATITUDE<>-1 AND a.VAL_LONGITUDE<>-1 AND a.COD_CONCELHO<>-
1 AND a.COD_DISTRITO<>-1 AND a.COD_FREGUESIA<>-1 and a.COD_FREGUESIA<>0 '
        || ' and
a.VAL_LONGITUDE=b.VAL_LONGITUDE(+) '
        || ' and b.VAL_LATITUDE IS NULL AND b.VAL LONGITUDE IS NULL'
        || 'GROUP BY a.VAL_LATITUDE, a.VAL_LONGITUDE, LPAD(a.COD_FREGUESIA,6,'0'),
LPAD(SUBSTR(LPAD(a.COD_FREGUESIA,6,'0'),0,4),4,'0'),
LPAD(SUBSTR(LPAD(a.COD_FREGUESIA,6,'0'),0,2),2,'0'))';

```

```

dbms_output.put_line(query);
execute immediate query ;

```

```
-----
passo := 'P040.1'; --INSERIR as antenas novas na DMG_REF_ANTENA
-----
```

```

query := ' INSERT INTO '
        || v_owner_dmg
        || v_tab_ref_antena
        || ' (COD_ANTENA,VAL_LATITUDE, VAL_LONGITUDE,COD_FREGUESIA , COD_CONCELHO,
COD_DISTRITO) '
        || ' (SELECT '
        || v_seq_antena
        || '.nextVal as COD_ANTENA,a.VAL_LATITUDE, a.VAL_LONGITUDE,a.COD_FREGUESIA ,
a.COD_CONCELHO, a.COD_DISTRITO '
        || 'FROM '
        || v_tabela_temporaria1 || '_' || p_cod_dia || ' a)';
dbms_output.put_line(query);

```

```

execute immediate query ;
commit;

```

```
-----
passo := 'P050'; --CRIAR UMA TABELA AUXILIAR COM AS CELULAS que correspondem às antenas
-----
```

```
pgbdt10q(v_tabela_temporaria2 || '_' || p_cod_dia);
```

```

query := 'CREATE TABLE ' || v_tabela_temporaria2 || '_' || p_cod_dia
        || ' TABLESPACE ' || p_tablespace_tmp || ' STORAGE (INITIAL 50M NEXT 50M)
PCTFREE 1 NOLOGGING AS '
        || '(SELECT a.COD_ANTENA, b.COD_CELULA '
        || 'FROM '
        || v_owner_dmg
        || v_tab_ref_antena
        || ' a, '
        || v_owner_dw
        || v_tab_ref_celula

```

```

        ||v_dblink_dw
        ||' b '
        ||' WHERE a.VAL_LATITUDE= b.VAL_LATITUDE '
        ||' AND a.VAL_LONGITUDE= b.VAL_LONGITUDE '
        ||'
        ||'
        a.COD_CONCELHO=LPAD(SUBSTR(LPAD(b.COD_FREGUESIA,6,'0'),0,4),4,'0') '
        ||'
        a.COD_DISTRITO=LPAD(SUBSTR(LPAD(b.COD_FREGUESIA,6,'0'),0,2),2,'0') '
        ||' AND a.COD_FREGUESIA=LPAD(b.COD_FREGUESIA,6,'0')) '
        AND
        AND

    dbms_output.put_line(query);

    execute immediate query ;

-----
passo := 'P050.1'; --CRIAR UMA TABELA AUXILAR COM AS CELULAS que correspondem às
antenas
-----
pgbdt10q(v_tabela_temporaria3 ||'_'|| p_cod_dia);

    query := 'CREATE TABLE ' || v_tabela_temporaria3 ||'_'|| p_cod_dia
        || ' TABLESPACE ' || p_tablespace_tmp || ' STORAGE (INITIAL 50M NEXT 50M)
PCTFREE 1 NOLOGGING AS '
        || '(SELECT b.COD_ANTENA, b.COD_CELULA '
        || 'FROM '
        ||v_owner_dmg
        ||v_tab_ref_celula_antena
        ||' a, '
        || v_tabela_temporaria2 ||'_'|| p_cod_dia
        ||' b '
        ||' WHERE b.COD_ANTENA= a.COD_ANTENA(+) '
        ||' AND b.COD_CELULA= a.COD_CELULA(+) '
        ||' AND a.COD_ANTENA IS NULL AND a.COD_CELULA IS NULL) ' ;

    dbms_output.put_line(query);

execute immediate query ;

-----
passo := 'P050.2'; --ACTUALIZAR A TABELA DMG_REF_CELULA_ANTENA
-----

query := ' INSERT INTO '
        ||v_owner_dmg
        ||v_tab_ref_celula_antena
        || ' (COD_CELULA_ANTENA,COD_ANTENA, COD_CELULA) '
        || ' (SELECT '
        ||v_seq_celula_antena
        || '.nextVal as COD_CELULA_ANTENA, a.COD_ANTENA, a.COD_CELULA '
        || 'FROM '
        || v_tabela_temporaria3 ||'_'|| p_cod_dia
        ||' a )' ;

    execute immediate query ;
    commit;

    ELSE
        dbms_output.put_line('Não existem antenas novas');
    END IF;

-----
passo := 'P60'; --eliminar tabela temporaria
-----
PGBDT10Q(v_tabela_temporaria1 ||'_'|| p_cod_dia);
PGBDT10Q(v_tabela_temporaria2 ||'_'|| p_cod_dia);
PGBDT10Q(v_tabela_temporaria3 ||'_'|| p_cod_dia);

-----
passo := 'P70'; --eliminar sequencia
-----
query:= 'DROP SEQUENCE ' ||v_seq_antena;

EXECUTE IMMEDIATE query;
query:= 'DROP SEQUENCE ' ||v_seq_celula_antena;
EXECUTE IMMEDIATE query;

```

```

-- TRATAMENTO DE EXCEPÇÕES --
EXCEPTION
WHEN OTHERS THEN

    sql_code := sqlcode;
    dbms_output.put_line('PASSO: ' || passo);
    dbms_output.put_line('ERRO: ' || sql_code || ' -> ' || sqlerrm(to_char(sql_code)) ||
    ' . ');
    dbms_output.put_line('Processo "' || nome_procedure || '" abortado! > ' ||
to_char(sysdate, 'DD-MM-YYYY HH24:MI:SS'));
    raise_application_error(-20001,'Erro na execução do procedimento!');

END PRC_UPDATE_DMG_REF_ANTENA ;
/

```

#### ❖ DMG\_REF\_ANTENA.ddl

```

CREATE TABLE DMG_REF_ANTENA
(
    COD_ANTENA          NUMBER(5)          NOT NULL,
    VAL_LATITUDE        NUMBER(14,9)       NOT NULL,
    VAL_LONGITUDE       NUMBER(14,9)       NOT NULL,
    COD_FREGUESIA       VARCHAR2(6 BYTE)   NOT NULL,
    COD_CONCELHO        VARCHAR2(4 BYTE)   NOT NULL,
    COD_DISTRITO        VARCHAR2(2 BYTE)   NOT NULL
)
TABLESPACE TBS_DAT_REF_DMGEO;

```

#### ❖ DMG\_REF\_CELULA\_ANTENA.ddl

```

CREATE TABLE DMG_REF_CELULA_ANTENA
(
    COD_CELULA_ANTENA  NUMBER(5)          NOT NULL,
    COD_ANTENA         NUMBER(5)          NOT NULL,
    COD_CELULA         NUMBER(5)          NOT NULL
)
TABLESPACE TBS_DAT_REF_DMGEO;

```

#### ❖ DMG\_REF\_CELULA\_ANTENA.idx

```

CREATE INDEX I_DMG_REF_CELULA_ANTENA_1
ON DMG_REF_CELULA_ANTENA (COD_CELULA)
TABLESPACE TBS_IDX_REF_DMGEO
PCTFREE 1
NOLOGGING;

CREATE BITMAP INDEX B_DMG_REF_CELULA_ANTENA_2
ON DMG_REF_CELULA_ANTENA (COD_ANTENA)
TABLESPACE TBS_IDX_REF_DMGEO
PCTFREE 1
NOLOGGING;

```

[anexo testes\_etl]

### Caso de teste 1

```

[dmdesenv@pirineus: /users/dmdesenv/LOG/DMGEO ]# cat
dsDMG_REF_ANTENAh.sh.PRC_UPDATE_DMG_REF_ANTENA.20090331.log

```

Processo dsDMG\_REF\_ANTENAh.sh PRC\_UPDATE\_DMG\_REF\_ANTENA 20090331 iniciado Fri May 29 11:41:17 WEST 2009.

Passo 1. dsDMG\_REF\_ANTENAh.sh PRC\_UPDATE\_DMG\_REF\_ANTENA 20090331 Fri May 29 11:41:17 WEST 2009.

Processo dsDMG\_REF\_ANTENAh.sh PRC\_UPDATE\_DMG\_REF\_ANTENA 20090331 terminado com sucesso.Fri May 29 11:41:26 WEST 2009

```

[dmdesenv@pirineus: /users/dmdesenv/LOG/DMGEO ]# cat dsExtracaoCDRsh.sh.20090331.log

```

Processo dsExtracaoCDRsh.sh iniciado para o dia de trafego 20090331. Fri May 29 11:41:26 WEST 2009

Passo 1: JOB Compile Fri May 29 11:41:26 WEST 2009.

Passo 2: Lancando dsExtracaoCDRs. Fri May 29 11:41:28 WEST 2009

Passo 3: Apagar ficheiro temporario Fri May 29 11:42:31 WEST 2009

Processo dsExtracaoCDRsh.sh Terminado : Fri May 29 11:42:31 WEST 2009

```
[dmdesenv@pirineus: /users/dmdesenv/LOG/DMGEO ]# cat dsExtracaoCDRs_BDLh.sh.20090331.log
```

Processo dsExtracaoCDRs\_BDLh.sh iniciado para o dia de trafego 20090331. Fri May 29 11:42:33 WEST 2009

Passo 1: JOB Compile Fri May 29 11:42:33 WEST 2009.

Passo 2: Lancando dsExtracaoCDRs\_BDL. Fri May 29 11:42:35 WEST 2009

Passo 3: Apagar tabelas temporarias Fri May 29 11:44:32 WEST 2009

Passo: Apagar tabelas temporarias Fri May 29 11:44:32 WEST 2009

Processo dsExtracaoCDRs\_BDLh.sh Terminado : Fri May 29 11:45:31 WEST 2009

```
[dmdesenv@pirineus: /users/dmdesenv/LOG/DMGEO ]# cat dsExtracaoCDRs_OPRAh.sh.20090331.log
```

Processo dsExtracaoCDRs\_OPRAh.sh iniciado para o dia de trafego 20090331. Fri May 29 11:44:33 WEST 2009

Passo 1: JOB Compile Fri May 29 11:44:33 WEST 2009.

Passo 2: Lancando dsExtracaoCDRs\_OPRA. Fri May 29 11:44:35 WEST 2009

Passo 3: Apagar tabelas temporarias Fri May 29 11:46:16 WEST 2009

Processo dsExtracaoCDRs\_OPRAh.sh Terminado : Fri May 29 11:46:19 WEST 2009

(...)

## Caso de teste 2

```
--operadora A
```

```
SELECT COUNT (0), cod_semana
      FROM dmq_agg_crt_opra_s1
GROUP BY cod_semana
```

COUNT (0)	COD_SEMANA
3121090	200909
845920	200910
1089293	200914

```
SELECT COUNT (0), cod_semana
      FROM dmq_agg_crt_opra_12s_s1
GROUP BY cod_semana
```

COUNT (0)	COD_SEMANA
3121090	200909
3221845	200910
3221845	200914

--concorrenca

```
SELECT COUNT (0), cod_semana
FROM dmG_agg_crt_conc_s1
GROUP BY cod_semana
```

COUNT(0)	COD_SEMANA
2248498	200909
236557	200910
334771	200914

```
select count(0), cod_semana from DMG_AGG_CRT_conc_12s_S1 group by
cod_semana
```

COUNT(0)	COD_SEMANA
1417961	200909
1504563	200910
1504563	200914

--quota operador

```
SELECT *
FROM dmG_fact_quota_operador
WHERE cod_semana = 200909
```

COD_SEMANA	COD_OPER_ORIG_DEST	QUOTA_NACIONAL	VARIACAO_QUOTA	ADICAO_LIQUIDA
200909		86	,38	18961
200909		6	8,63	428480
200909		89	5,06	251131
200909		5	23,02	1143156
200909		88	,05	2350
200909		4	62,85	3121090

```
SELECT *
FROM dmG_fact_quota_operador
WHERE cod_semana = 200910
```

COD_SEMANA	COD_OPER_ORIG_DEST	QUOTA_NACIONAL	VARIACAO_QUOTA	ADICAO_LIQUIDA
200910		86	,55	0
200910		6	11,82	0
200910		89	4,29	0
200910		5	30,35	0
200910		88	,09	0
200910		4	52,89	0

--com sazonalidade

```
SELECT *
FROM dmG_fact_quota_operador
WHERE cod_semana = 200914
```

COD_SEMANA	COD_OPER_ORIG_DEST	QUOTA_NACIONAL	VARIACAO_QUOTA	ADICAO_LIQUIDA
200914		86	,37	18961
200914		6	8,37	428480
200914		89	4,9	251131
200914		5	22,32	1143156
200914		88	,05	2350
200914		4	63,98	3277262

--com quarentena

```
SELECT *
```

```
FROM dmng_fact_quota_operador
WHERE cod_semana = 200914
```

COD_SEMANA	COD_OPER_ORIG_DEST	QUOTA_NACIONAL	VARIACAO_QUOTA	ADICAO_LIQUIDA
200914	86	,38	,38	19895
200914	6	8,83	8,83	456303
200914	89	4,95	4,95	256023
200914	5	23,44	23,44	1211563
200914	88	,05	,05	2463
200914	4	62,33	62,33	3221845

### Caso de teste 3

```
--operadora A
```

```
--DMG_AGG_CRT_OPRA_S1
SELECT cod_semana, cod_cartao
FROM dmng_agg_crt_opra_s1
GROUP BY cod_semana, cod_cartao
HAVING COUNT (*) > 1
--0
```

```
--DMG_AGG_CRT_OPRA_12S_S1
SELECT cod_semana, cod_cartao
FROM dmng_agg_crt_opra_12s_s1
GROUP BY cod_semana, cod_cartao
HAVING COUNT (*) > 1
--0
```

```
--concorrenca
```

```
--DMG_AGG_CRT_CONC_S1
SELECT cod_semana, cod_cartao, msisdn, msisdn_concorrenca,
cod_tipo_orig_dest, qtd_eventos, dur_eventos
FROM dmng_agg_crt_conc_s1
GROUP BY cod_semana, cod_cartao, msisdn, msisdn_concorrenca,
cod_tipo_orig_dest, qtd_eventos, dur_eventos
HAVING COUNT (*) > 1
--0
```

### Caso de teste 4

```
--DMG_AGG_CRT_OPRA_S1
SELECT cod_cartao
FROM dmng_agg_crt_opra_s1 PARTITION (dmng_agg_crt_opra_s1_200909)
MINUS
SELECT cod_cartao
FROM hist_cartao@dblink_dw
WHERE cod_dia = 20090228
--0
```

```
--DMG_AGG_CRT_OPRA_12S_S1
SELECT cod_cartao
FROM dmng_agg_crt_opra_12s_s1 PARTITION
(dmng_agg_crt_opra_12s_s1_200909)
MINUS
SELECT cod_cartao
FROM hist_cartao@dblink_dw
WHERE cod_dia = 20090228
--0
```

```
--DMG_AGG_CRT_CONC_S1
SELECT cod_cartao
  FROM dmg_agg_crt_conc_s1
MINUS
SELECT cod_cartao
  FROM hist_cartao@dblink_dw
 WHERE cod_dia = 20090228
--0

SELECT cod_tipo_orig_dest
  FROM dmgopp.dmg_agg_crt_conc_s1
MINUS
SELECT cod_tipo_orig_dest
  FROM prodods.fact_cdr_hist@dblink_pddwods
 WHERE cod_dia = 20090228
--0
```

### Caso de teste 5

```
SELECT COUNT (a.msisdn_concorrencia)
  FROM tmp_geor_conc_top_k_200902 a,
       dmg_agg_geomkt_base_m1 b,
       dmg_ref_celula_antena c
 WHERE a.cod_cartao <> -1
       AND a.cod_cartao = b.cod_cartao
       AND a.cod_mes = 200902
       AND b.cod_mes = 200902
       AND b.cod_celula_3p = c.cod_celula
       AND a.cod_tipo_orig_dest IN (3000, 5524) --operador c
       AND b.cod_celula_3p IN (SELECT cod_celula
                               FROM dmg_ref_celula_antena
                              WHERE cod_antena = 1630)
--100
```

```
SELECT COUNT (a.msisdn_concorrencia)
  FROM tmp_geor_conc_top_k_200902 a,
       dmg_agg_gmkt_base_u_m1 b,
       hist_cartao_2@dblink_dw c,
       dmg_ref_celula_antena d
 WHERE a.cod_cartao = -1
       AND a.msisdn = c.msisdn
       AND b.cod_cartao_2 = c.cod_cartao_2
       AND b.cod_celula_3p = d.cod_celula
       AND a.cod_mes = 200902
       AND b.cod_mes = 200902
       AND c.cod_dia = 20090228
       AND cod_tipo_orig_dest IN (3000, 5524) --operador c
       AND b.cod_celula_3p IN (SELECT cod_celula
                               FROM dmg_ref_celula_antena
                              WHERE cod_antena = 1630)
--32
```

--100+32=132

```
SELECT qtd_crt_oprc_ant
  FROM dmg_agg_crt_oper_m1
 WHERE cod_mes = 200902 AND cod_antena = 1630
--132
```

### Caso de teste 6

```
SELECT SUM (quota_nacional)
  FROM dmng_fact_quota_operador
 WHERE cod_semana = 200909
--100
```

#### Caso de teste 7

```
--operador A
select sum(crt_totais), sum(crt_totais)/4965861
from dmng_fact_crt_operador
where cod_semana = 200909
and cod_tipo_orig_dest = 0
--0,628509336044646

--operador C
SELECT SUM (crt_totais), SUM (crt_totais) / 4965861
  FROM dmng_fact_crt_operador
 WHERE cod_semana = 200909
 AND cod_tipo_orig_dest IN (3000, 5524)
--0,230202979906204

--operador D
SELECT SUM (crt_totais), SUM (crt_totais) / 4965861
  FROM dmng_fact_crt_operador
 WHERE cod_semana = 200909
 AND cod_tipo_orig_dest IN (3001, 3004, 5525)
--0,086285137662935
```

#### Caso de teste 8

```
--operador A
SELECT SUM (crt_totais), SUM (crt_totais) / 5122033
  FROM dmng_fact_crt_operador
 WHERE cod_semana = 200914
 AND cod_tipo_orig_dest = 0
--0,639836174425272

--operador C
SELECT SUM (crt_totais), SUM (crt_totais) / 5122033
  FROM dmng_fact_crt_operador
 WHERE cod_semana = 200914
 AND cod_tipo_orig_dest IN (3000, 5524)
--0,223184036494884

--operador D
SELECT SUM (crt_totais), SUM (crt_totais) / 5122033
  FROM dmng_fact_crt_operador
 WHERE cod_semana = 200914
 AND cod_tipo_orig_dest IN (3001, 3004, 5525)
--0,0836542833675613
```

#### Caso de teste 9

```
--operador A
SELECT SUM (crt_totais), SUM (crt_totais) / 5168092
  FROM dmng_fact_crt_operador
 WHERE cod_semana = 200914
 AND cod_tipo_orig_dest = 0
```

```
--0,623410922251384

--operador C
SELECT SUM (crt_totais), SUM (crt_totais) / 5168092
  FROM dm_g_fact_crt_operador
 WHERE cod_semana = 200914
 AND cod_tipo_orig_dest IN (3000, 5524)
--0,234431391701231

--operador D
SELECT SUM (crt_totais), SUM (crt_totais) / 5168092
  FROM dm_g_fact_crt_operador
 WHERE cod_semana = 200914
 AND cod_tipo_orig_dest IN (3001, 3004, 5525)
--0,0882923523807239
```

### Caso de teste 10

```
SELECT DISTINCT a.cod_cartao
  FROM pddwods.fact_cdr_hist@dblink_pddwods a,
       pddw.hist_cartao@dblink_dw b
 WHERE a.cod_dia = b.cod_dia
       AND a.cod_dia = 20090228
       AND a.cod_cartao = b.cod_cartao
       AND ( a.cod_other_party IN ('OPRA', 'OPRA99') -- tarif
            OR b.cod_tarifario IN
              (1620, 1621, 6808, 6809, 58, 85, 86, 87) --
            )
MINUS
SELECT DISTINCT cod_cartao
  FROM dm_g_agg_crt_opra_12s_s1
 WHERE cod_semana = 200909
--613

SELECT DISTINCT cod_cartao
  FROM dm_g_agg_crt_opra_12s_s1
 WHERE cod_semana = 200909
       AND cod_cartao IN
       (-912425867, -866653767, -619371967, -613321407,
        -588231507, -490100687, -439955097, -431021227,
        -121888970, -105036147, 15694376, 15694651,
        15730867, 15730880, 15730882, 15730910, 15731171,
        (...),
        21202318, 21202352, 21202383, 21331586,
        21333324, 21401962, 21402273, 21456895,
        21457406 ) --613 cartoes
--0
```

### Caso de teste 11

```
SELECT DISTINCT cod_cartao
  FROM dm_g_agg_crt_opra_12s_s1
 WHERE cod_semana = 200909
       AND cod_cartao IN (
        SELECT DISTINCT a.cod_cartao
```

```

FROM
pddwods.fact_cdr_hist@dblink_pddwods a,
        pddw.hist_cartao@dblink_dw b
WHERE a.cod_dia = b.cod_dia
      AND a.cod_dia = 20090228
      AND a.cod_cartao = b.cod_cartao
      AND b.msisdn < 300000000)
--0

```

## Caso de teste 12

--concorrenca

```

SELECT COUNT (0), cod_tipo_orig_dest
FROM dmng_agg_georef_conc_ml
WHERE cod_celula_prefer IN (SELECT cod_celula
                            FROM dmng_ref_celula_antena
                            WHERE cod_antena = 1630) AND cod_mes =
200902
GROUP BY cod_tipo_orig_dest

```

COUNT(0)	COD_TIPO_ORIG_DEST
57	3001
130	3000
2	3004
2	5524
2	5525

```

SELECT qtd_crt_ant, cod_tipo_orig_dest
FROM dmng_agg_crt_oridest_ml
WHERE cod_antena = 1630
      AND cod_mes = 200902
      AND cod_tipo_orig_dest IN (3000, 3001, 3004, 5524, 5525)
GROUP BY qtd_crt_ant, cod_tipo_orig_dest

```

QTD_CRT_ANT	COD_TIPO_ORIG_DEST
2	3004
2	5525
2	5524
57	3001
130	3000

--operador A

```

SELECT qtd_crt_opra, cod_antena
FROM tmp_quota_opra_celprf_200902
WHERE cod_antena = 1630 AND cod_mes = 200902
GROUP BY qtd_crt_opra, cod_antena

```

QTD_CRT_OPRA	COD_ANTENA
229	1630

```

SELECT qtd_crt_ant, cod_tipo_orig_dest
FROM dmng_agg_crt_oridest_ml
WHERE cod_antena = 1630 AND cod_mes = 200902 AND cod_tipo_orig_dest
= 0
GROUP BY qtd_crt_ant, cod_tipo_orig_dest

```

QTD\_CRT\_ANT COD\_TIPO\_ORIG\_DEST

229

0

[anexo testes\_gis]

### Caso de teste 1

--teste 1

\*\*\*\*\*

Ficheiro de LOG do processo ArcGIS - Calculo de Quotas

Ficheiro : LOG\_GIS\_QUOTAS\_20090702.log

Inicio de Processo : 2009-07-02 as 10:03:51

\*\*\*\*\*

100 -> LogTime -> 2009-07-02 as 10:03:51

#####

##### INICIO DE PROCESSO GEOGRAFICO #####

#####

1 -> LogTime -> 2009-07-02 as 10:03:51

Criar Objecto Geoprocessor

2 -> LogTime -> 2009-07-02 as 10:03:52

Verificar licenca ArcGIS

4 -> LogTime -> 2009-07-02 as 10:03:55

Existe licenca de ArcINFO disponivel

5 -> LogTime -> 2009-07-02 as 10:03:55

Overwrite output

6 -> LogTime -> 2009-07-02 as 10:03:55

Definicao de campos de join como UNQUALIFIED

7 -> LogTime -> 2009-07-02 as 10:03:55

criar conexao oracle

8 -> LogTime -> 2009-07-02 as 10:03:57

Query de analise -> COD\_MES = 200902 <-

10 -> LogTime -> 2009-07-02 as 10:03:57

Iniciar processos de instanciação de definições de conexões e workspace

11 -> LogTime -> 2009-07-02 as 10:03:57

Testar workspaces

20 -> LogTime -> 2009-07-02 as 10:04:11

Existe connexcao ao workspace do user DSVGEOOPRA

30 -> LogTime -> 2009-07-02 as 10:04:13

Existe connexcao ao workspace do user CARTO\_VIA\_DSVGEOOPRA

32 -> LogTime -> 2009-07-02 as 10:04:15

Existe connexcao ao workspace do user CARTO

35 -> LogTime -> 2009-07-02 as 10:04:15

verificar existencia de dados de input

40 -> LogTime -> 2009-07-02 as 10:04:20

Existe dados de entrada de distritos

50 -> LogTime -> 2009-07-02 as 10:04:21

Existe dados de entrada de concelhos

60 -> LogTime -> 2009-07-02 as 10:04:22

Existe feature dataset destino QUOTAS

60 -> LogTime -> 2009-07-02 as 10:04:22

Definicao workspace

70 -> LogTime -> 2009-07-02 as 10:04:24

Terminado processos de instanciacao de definicoes de conexoes e workspace

100 -> LogTime -> 2009-07-02 as 10:04:24

(...)

## Caso de teste 2

As tabelas *DMG\_AGG\_APURA\_QUOTA\_M1*, *DMG\_AGG\_AP\_QUOTA\_DD\_M1*, *DMG\_AGG\_AP\_QUOTA\_CC\_M1*

Encontram-se de acordo com a estrutura apresentada na Análise Técnica sendo assim cada uma composta pelos seguintes campos:

*DMG\_AGG\_APURA\_QUOTA\_M1*

```
SELECT d.cod_mes, d.cod_antena, d.val_latitude, d.val_longitude,
       d.quota_opra_ant,      d.quota_oprb_ant,      d.quota_oprc_ant,
d.quota_oprd_ant,
       d.quota_opre_ant,      d.quota_oprf_ant,      d.var_liq_opra,
d.var_liq_oprb,
       d.var_liq_oprc, d.var_liq_oprd, d.var_liq_opre, d.var_liq_oprf,
       d.var_quota_opra,      d.var_quota_oprb,      d.var_quota_oprc,
d.var_quota_oprd,
       d.var_quota_opre, d.var_quota_oprf
FROM dsvgeopra.dmg_agg_apura_quota_ml d
```

(...)

## Caso de teste 3

Neste caso como são copias de tabelas para outro ambiente basta validar que não diferenças entre elas fazendo um MINUS de uma pela outra e vice-versa

```

SELECT COUNT (*)
  FROM dm_g_agg_apura_quota_ml@dblink_dm_gopp
MINUS
SELECT COUNT (*)
  FROM dm_g_agg_apura_quota_ml
--0

```

```

SELECT COUNT (*)
  FROM dm_g_agg_apura_quota_ml
MINUS
SELECT COUNT (*)
  FROM dm_g_agg_apura_quota_ml@dblink_dm_gopp
--0

```

```

SELECT COUNT (*)
  FROM dm_g_agg_ap_quota_dd_ml@dblink_dm_gopp
MINUS
SELECT COUNT (*)
  FROM dm_g_agg_ap_quota_dd_ml
--0

```

```

SELECT COUNT (*)
  FROM dm_g_agg_ap_quota_dd_ml
MINUS
SELECT COUNT (*)
  FROM dm_g_agg_ap_quota_dd_ml@dblink_dm_gopp
--0

```

(...)

para validar as tabelas de dimensão geográfica é necessário ir a uma amostra de registos e comparar se o que está na BD é igual ao que está na tabela de dimensão geográfica

```

--BD
--DISTRITO:Lisboa
SELECT *
  FROM dm_g_agg_ap_quota_dd_ml
 WHERE cod_mes = 200902 AND cod_distrito = 11

```

COD_MES,	200902
COD_DISTRITO,	11
QUOTA_OPRA_DD,	64,97
QUOTA_OPRB_DD,	4,48
QUOTA_OPRC_DD,	22,02
QUOTA_OPRD_DD,	8,12
QUOTA_OPRE_DD,	0,35
QUOTA_OPRE_DD,	0,07
VAR_LIQ_OPRA,	954.693
VAR_LIQ_OPRB,	65.778
VAR_LIQ_OPRC,	323.644
VAR_LIQ_OPRD,	119.266
VAR_LIQ_OPRE,	5.117
VAR_LIQ_OPRE,	1.010
VAR_QUOTA_OPRA,	64,97
VAR_QUOTA_OPRB,	4,48
VAR_QUOTA_OPRC,	22,02
VAR_QUOTA_OPRD,	8,12
VAR_QUOTA_OPRE,	0,35
VAR_QUOTA_OPRE,	0,07

```

--CONCELHO:Lisboa

```

```
SELECT *
FROM dmG_agg_ap_quota_cc_ml
WHERE cod_mes = 200902 AND cod_concelho = 1106
```

COD_MES,	200902
COD_CONCELHO,	1106
QUOTA_OPRA_DDCC,	61,64
QUOTA_OPRB_DDCC,	4,46
QUOTA_OPRC_DDCC,	24,61
QUOTA_OPRD_DDCC,	8,85
QUOTA_OPRE_DDCC,	0,36
QUOTA_OPRF_DDCC,	0,08
VAR_LIQ_OPRA,	296.160
VAR_LIQ_OPRB,	21.424
VAR_LIQ_OPRC,	118.246
VAR_LIQ_OPRD,	42.497
VAR_LIQ_OPRE,	1.741
VAR_LIQ_OPRF,	382
VAR_QUOTA_OPRA,	61,64
VAR_QUOTA_OPRB,	4,46
VAR_QUOTA_OPRC,	24,61
VAR_QUOTA_OPRD,	8,85
VAR_QUOTA_OPRE,	0,36
VAR_QUOTA_OPRF,	0,08

--ANTENA:Lisboa,Lisboa

```
SELECT cod_mes, cod_antena, val_latitude, val_longitude,
quota_opra_ant
FROM dmG_agg_apura_quota_ml
WHERE cod_mes = 200902 AND cod_antena = 1020
```

COD_MES	200902
COD_ANTENA	1020
VAL_LATITUDE	38,7242
VAL_LONGITUDE	-9,15306
QUOTA_OPRA_ANT	45,68

--DISTRITO:Lisboa

**LISBOA** (Apuramento de Quotas por Distrito) △ (X)

DSG_DISTRITO	LISBOA
COD_MES	200902
QUOTA_OPRA_DD	64,97
QUOTA_OPRB_DD	4,48
QUOTA_OPRC_DD	22,02
QUOTA_OPRD_DD	8,12
QUOTA_OPRE_DD	0,35
QUOTA_OPRF_DD	0,07
VAR_LIQ_OPRA	954693
VAR_LIQ_OPRB	65778
VAR_LIQ_OPRC	323644
VAR_LIQ_OPRD	119266
VAR_LIQ_OPRE	5117
VAR_LIQ_OPRF	1010
VAR_QUOTA_OPRA	64,97

GEOOPRA\_DM GEO\_QUOTAS\_PT > Apuramento de Quotas por Distrito  
[Add to Results](#)

Figura 36 – Apuramento de quotas por distrito

```
--CONCELHO:Lisboa
```

**LISBOA** (Apuramento de Quotas por Concelho) △ (X)

DSG_CONCELHO	LISBOA
COD_DISTRITO	11
COD_MES	200902
QUOTA_OPRA_DDCC	61,64
QUOTA_OPRB_DDCC	4,46
QUOTA_OPRC_DDCC	24,61
QUOTA_OPRD_DDCC	8,85
QUOTA_OPRE_DDCC	0,36
QUOTA_OPRF_DDCC	0,08
VAR_LIQ_OPRA	296160
VAR_LIQ_OPRB	21424
VAR_LIQ_OPRC	118246
VAR_LIQ_OPRD	42497
VAR_LIQ_OPRE	1741
VAR_LIQ_OPRF	387

GEOOPRA\_DMGEO\_QUOTAS\_PT > Apuramento de Quotas por Concelho  
[Add to Results](#)

Figura 37 – Apuramento de quotas por concelho

```
--ANTENA:Lisboa,Lisboa
```

**i Identify** ? X

Identify from: Quotas de mercado da OPRA

Field	Value
Class value	13
Pixel value	46
OBJECTID	47
COUNT	1103

Location: -9,152984 38,724490 Decimal Degrees

Figura 38 – Apuramento de quota por antena

#### Caso de teste 4

O serviço e a aplicação do serviço ArcGIS Server encontram-se criados de acordo com as instruções fornecidas pela CheckList e estando a funcionar em pleno.

*--passo 1*

**Manage Services**

|

Services in:  Manage Folders

|  |  |  |

Name	Type	Status	Instances (In Use/Running)	Permissions	Edit
<input type="checkbox"/> GEOCRM	Map Service	Started	0/1		
<input type="checkbox"/> GEOCRM4	Map Service	Stopped	0/0		
<input type="checkbox"/> <b>GEOCRM4_DMGEQ_QUOTAS_PT</b>	Map Service	Started	0/1		
Description: Source: <input type="text" value="DMGEO_QUOTAS.mxd"/> Capabilities: Map Service, KML Pooling: Pooled Startup: Automatic					
<input type="checkbox"/> IndicadoresGeo	Map Service	Started	0/1		
<input type="checkbox"/> PortugalPTC	Map Service	Started	0/1		

<< 1 - 5 of 5 >>

Figura 39 – Validação da criação do serviço passo 1

*--passo 2*

|  |  |  |  |

Map Document:

Type in the location of the resource. If you want to browse to a location, only shared drives appear in the list.

Data Frame:

Output Directory:

Virtual Output Directory:

Supported Image Return Type:

Specify cache directory

Server cache directory:

Figura 40 – Validação da criação do serviço passo 2

--passo 3

The screenshot shows the 'Capabilities' tab in the ArcGIS configuration interface. The 'Map Service' checkbox is checked, and the text 'No properties to configure' is displayed. Other capabilities like WCS, WMS, WFS, Mobile Data Access, KML, and Network Analysis are unchecked. The 'Enable web access' checkbox is also checked. The URL field contains 'http://[redacted]/ArcGIS/service'. Under 'Operations allowed', the checkboxes for 'Map', 'Query', and 'Data' are all checked.

Figura 41 – Validação da criação do serviço passo 3

--passo 4

The screenshot shows the 'Caching' tab in the ArcGIS configuration interface. The 'Draw this map service' section has 'Using tiles from its cache' selected. The 'Scales' list contains '1:16,000,000.0000000000'. The 'Origin(x,y) in map units' section has X: -400 and Y: 400. The 'Image Settings' section has 'Tile Format' set to PNG8, 'Height' set to 512 pixels, 'Width' set to 512 pixels, and 'Resolution' set to 96 DPI. The 'Smooth line and label edges (anti-aliasing)' checkbox is unchecked. The 'Cache directory' is set to 'c:\arcgisserver\arcgiscache'. There are buttons for 'Add', 'Delete', 'Suggest...', and 'Advanced Options'.

Figura 42 – Validação da criação do serviço passo 4

### Caso de teste 5

Desktop: [http://\[redacted\]/QUOTASMERCADO/Default.aspx](http://[redacted]/QUOTASMERCADO/Default.aspx)

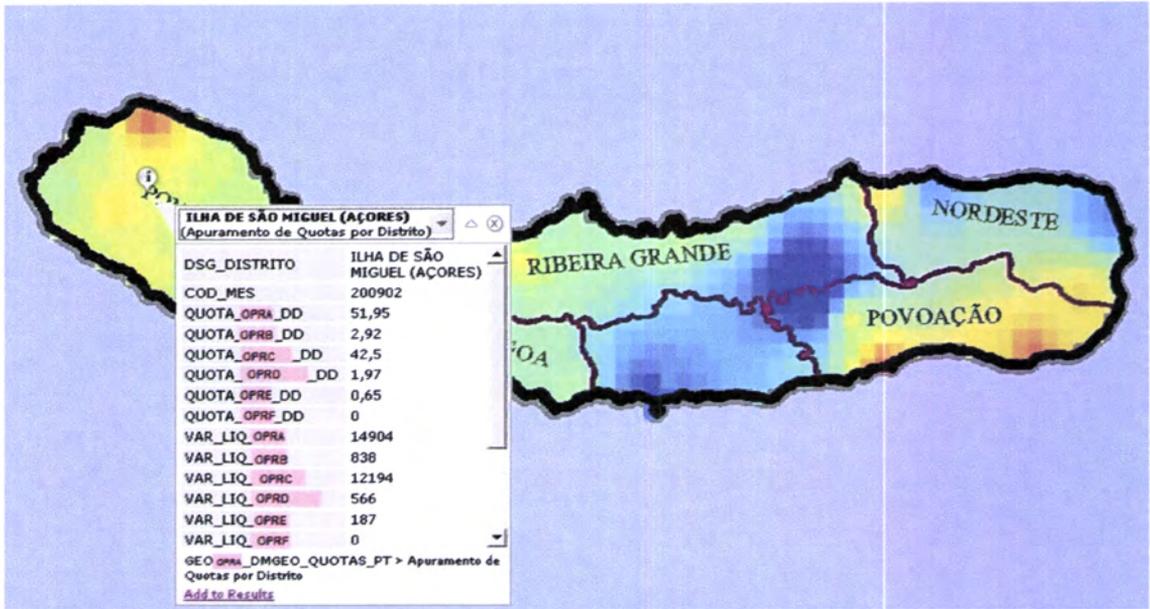


Figura 43 – Validação da mxd no desktop

Server: [http://\[redacted\]/QUOTASMERCADO/](http://[redacted]/QUOTASMERCADO/)

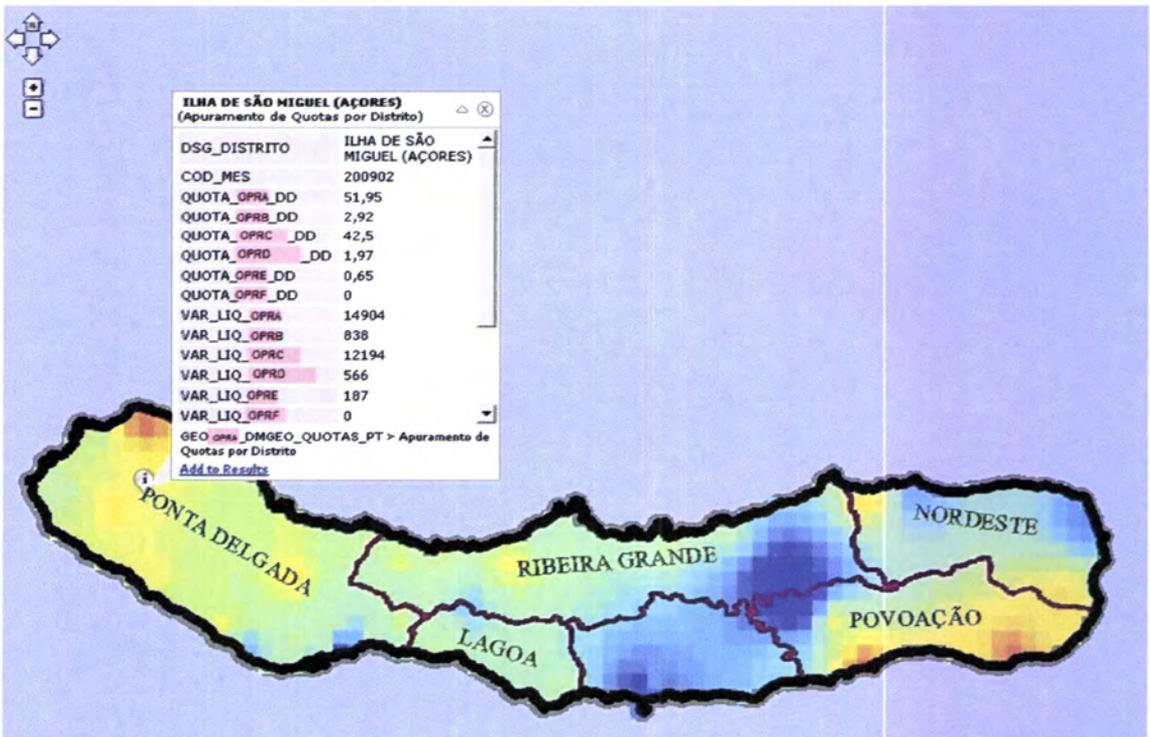


Figura 44 – Validação da mxd no ArcGIS Server