Mestrado em Engenharia Informática e de Computadores

SIAD - Sistemas de Informação para Apoio à Decisão



|  |  |
| --- | --- |
| **Assunto:** | **Trabalho de SIAD – BIT (Bikes In Transit) - Parte 2 (Fases 3 e 4)** |
| **Departamento:** | **DEETC** |
| **Última Revisão:** | **2013/06/21** |



Resumo:

Nesta segunda parte da implementação de um sistema de apoio à decisão para o sistema de dados BIT, partindo do desenvolvimento efetuado na primeira parte, foram construídos e completamente configurados os cubos OLAP, e foram diversas analises sobre esses mesmo cubos.

**Controlo de versões:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grupo nº** | | **7066** | **Frederico Ferreira** | | |
| **14940** | **Angelo Borges** | | |
| **32342** | **Rui Miranda** | | |
| **Versão** | **Autor** | | **Revisão** | **Data** | **Observações** |
| 1.0 | **Grupo** | | **Grupo** | **2013/22/06** |  |
| 1.1 | **Grupo** | | **Grupo** | **2013/23/06** |  |
| 1.2 | **Grupo** | | **Grupo** | **2013/24/06** | Análises |

Índice

[1. Introdução 3](#_Toc359874467)

[1.1. Melhorias e correções relativas às fases 1 e 2 do projeto 3](#_Toc359874468)

[2. Terceira e Quarta Fases 4](#_Toc359874469)

[2.1. Geral 4](#_Toc359874470)

[2.2. Desenvolvimento dos cubos OLAP 4](#_Toc359874471)

[2.2.1. Construção dos cubos 4](#_Toc359874472)

[2.2.2. Nomenclatura 4](#_Toc359874473)

[2.2.3. Tipo de Armazenamento cubos 5](#_Toc359874474)

[2.2.4. Cubo Financeiro 5](#_Toc359874475)

[2.2.5. Cubo Operações 6](#_Toc359874476)

[2.3. Análise 8](#_Toc359874477)

[2.3.1. Volume de faturação realizado em 2012, por mês e tipo de entrega, nas freguesias da Baixa Pombalina de Lisboa 8](#_Toc359874478)

[2.3.2. Número médio de Entregas por Hora do Dia (no último mês) 9](#_Toc359874479)

[2.3.3. Percentagem de crescimento das entregas urgentes, no último trimestre, para as freguesias com volume de faturação superior a 1.500€ 10](#_Toc359874480)

[2.3.4. Estafeta que mais entregas programadas fez fora de horas no último mês 11](#_Toc359874481)

[2.3.5. Balanço dos três últimos anos e respectivas diferenças absolutas 12](#_Toc359874482)

[2.3.6. Quota-parte que cada loja tem no volume total das entregas, em cada ano 13](#_Toc359874483)

[2.3.7. Outras análises 13](#_Toc359874484)

[3. Anexos 15](#_Toc359874485)

[4. Considerações Finais: 16](#_Toc359874486)

[4.1. Geral 16](#_Toc359874487)

[4.2. Melhorias 17](#_Toc359874488)

[4.3. Ferramentas 17](#_Toc359874489)

# Introdução

O presente relatório tem por objetivo endereçar as fases 3 e 4 do Trabalho Prático de SIAD do semestre de Verão de 2012/213. Como resultado de uma análise critica ao trabalho do grupo das fases 1 e 2, resultou a necessidade de produzir um relatório mais simples e direto.

Identificamos, também alguns erros de implementação que foram corrigidos nesta versão.

## Melhorias e correções relativas às fases 1 e 2 do projeto

As melhorias e correções implementadas ao trabalho das fases 1 e 2 do projeto foram as seguintes:

* Cálculos salariais: os cálculos estavam incorretos pois não consideravam os trabalhadores ativos no mês de agrupamento. A versão atual já considera essa correção;
* Entregas “fora do SLA”: o cálculo realizado para este campo estava incorreto e apresentava um resultado errado quanto ao dentro e fora do SLA. Foi corrigido no ETL e já apresenta resultados corretos;
* Carregamento incremental dos dados provenientes do OLTP não estava a ser considerada.

Para evitar interferir com o OLTP não consideramos inicialmente a extração incremental dos dados. Embora viável para este sistema de “testes” esta aproximação não seria razoável num sistema de produção. Alterámos por isso o processo, criando um zona de “Pré-stagging” no Data Warehouse, para onde transferimos apenas os “novos” dados.

Os novos dados são os produzidos no OLTP desde o último carregamento até ao início do corrente, para as tabelas que tenham campo com data de alteração e cuja dimensão justifique compense a utilização de um filtro (select ... where) versus a transferência total dos dados.

Na área de”Pre-stagging” mantemos uma tabela com a data e hora da última extração do ambiente OLTP e recorremos ao uso de Linked Data para implementar o query remoto.

Esta solução permitiu manter o restante processo já que tinha sido implementado e testado anteriormente.

# Terceira e Quarta Fases

## Geral

A segunda parte do trabalho é composta pelas Terceira e Quarta Fases, que incluem fundamentalmente a construção de cubos e a geração de análises, por forma a responder aos requisitos do enunciado. O relatório apresenta fundamentalmente os resultados da análise, que implicam a boa construção do Cubos OLAP incluídos no código produzido.

## Desenvolvimento dos cubos OLAP

### Construção dos cubos

Os cubos foram implementados com base no modelo dimensional implementado na fase anterior. A informação a disponibilizar serve 2 processos de negócio e áreas, para além de terem dados com granularidade distinta,pelo que a opção foi por criar 2 cubos: Financeiro e Operações.

### Nomenclatura

Foram renomeados objetos (dimensões, e atributos das dimensões) a partir do Data Source View, para que a sua interpretação durante a análise seja imediata.

### Tipo de Armazenamento cubos

O caso presente (e pensamos que mesmo numa aplicação real para um negócio semelhante) tem um volume relativamente pequeno de dados, pelo que a nossa opção foi para a otimização das pesquisas e optamos por MOLAP para tipo de armazenamento dos cubos.

No caso de maiores volumes de dados, alterações de fundo mais frequentes ou janelas de “atualização” mais pequenas teríamos de considerar o modelo ROLAP ou hibrido.

### Cubo Financeiro

#### Business Intelligence

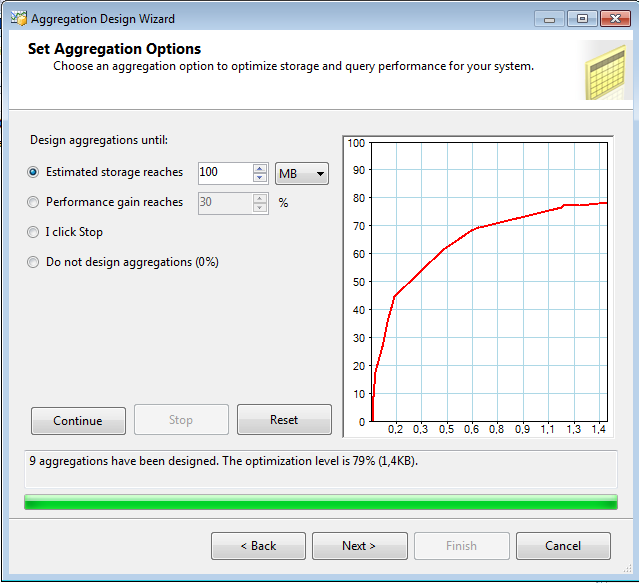
Foi adicionada *Business Intelligence* à dimensão **Rubrica**, como *Account*, em que **Custo** corresponde às despesas, e **Proveito** ao rendimento, e também à dimensão **Meses** como Tempo.

#### Hierarquias e relações entre atributos

Foram definidas Hierarquias, e respetivas relações, para a dimensão de Meses, e para a dimensão de Rubrica.

#### Desenho do pré-agregado

Para a medida Resultados do cubo, foi desenhado e implementado o pré-agregado com a seguinte resultado de performance:



Para a medida O grau de otimização de 79% é conseguido com 9 agregações e com 1,4KB. O gráfico mostra um ganho bastante residual a partir dos 1,1KB o que se em presença de um sistema com limitação de armazenamento nos poderia levar a optar por um número menor de agregações.

### Cubo Operações

#### Business Intelligence

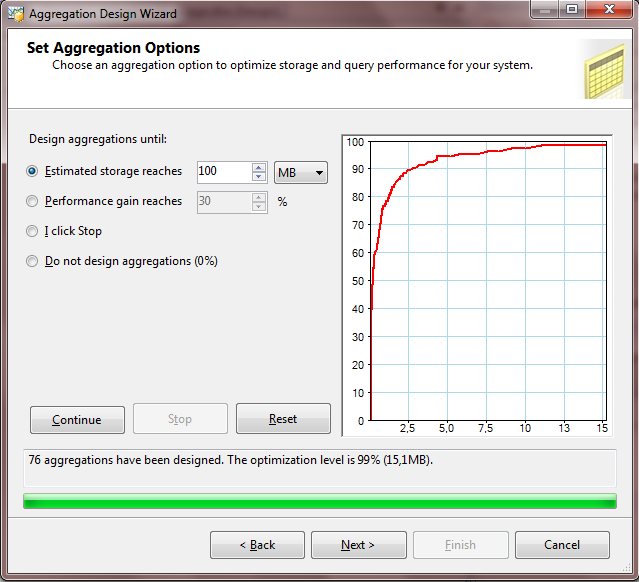
Foi adicionada *Business Intelligence* à dimensão **Data** como Tempo, à dimensão **Morada** como Geográfica, e à dimensão **Cliente** como Costumer.

#### Hierarquias e relações entre atributos

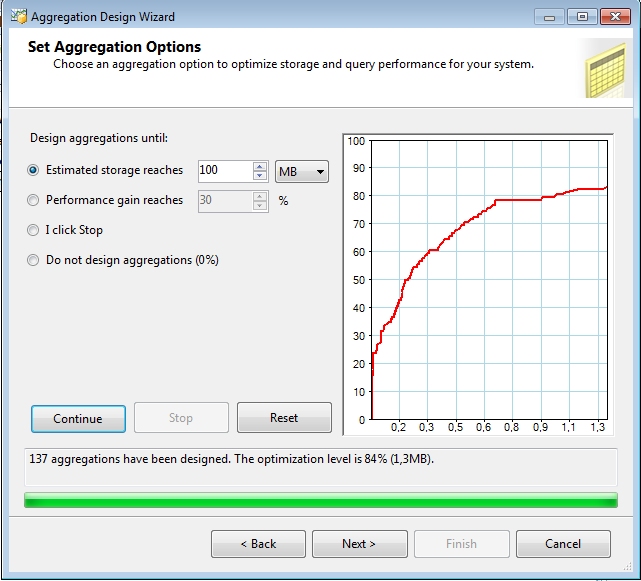
Foram definidas Hierarquias, e respetivas relações, para a dimensão de Data, Hora, Funcionário (por local e por chefe), Morada, Cliente (por designação/empresa).

#### Desenho do pré-agregado

Para a medida Servicos do cubo, foi desenhado e implementado o pré-agregado com a seguinte resultado de performance:



Para a medida Entregas dos Servicos do cubo, foi desenhado e implementado o pré-agregado com a seguinte resultado de performance:

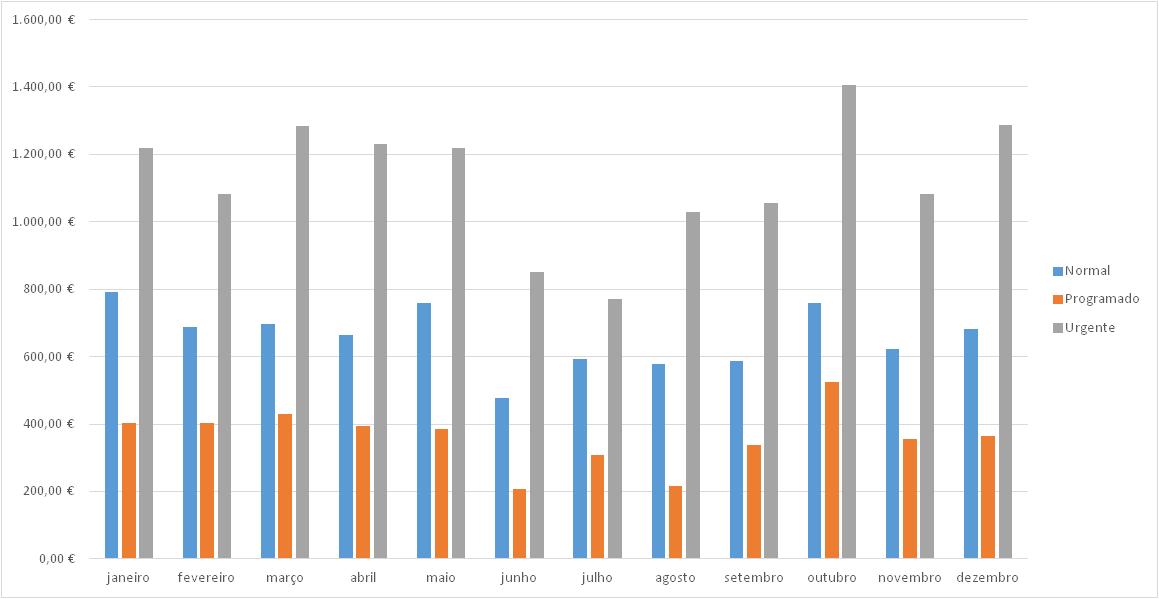


## Análise

### Volume de faturação realizado em 2012, por mês e tipo de entrega, nas freguesias da Baixa Pombalina de Lisboa

Nota: Assumiu-se como baixa pombalina as freguesias de Lisboa: São Nicolau, São Cristóvão, Santa Justa, Sacramento, Madalena, Mártires.

(fonte: <http://pt.wikipedia.org/wiki/Baixa_de_Lisboa>)



Query MDX:

select

NON EMPTY {([Tipo Servico].[Tipo].members)} on columns

,NON EMPTY {([Data Requisicao].[Ano - Dia].[Ano].&[2012],[Data Requisicao].[Mês Nome].members)} on rows

from [Servicos apenas]

where

(

{

[Morada Facturacao].[Distrito - Número Da Porta].[Freguesia].&[Lisboa]&[Lisboa]&[São Nicolau]

,[Morada Facturacao].[Distrito - Número Da Porta].[Freguesia].&[Lisboa]&[Lisboa]&[São Cristóvão]

,[Morada Facturacao].[Distrito - Número Da Porta].[Freguesia].&[Lisboa]&[Lisboa]&[Santa Justa]

,[Morada Facturacao].[Distrito - Número Da Porta].[Freguesia].&[Lisboa]&[Lisboa]&[Sacramento]

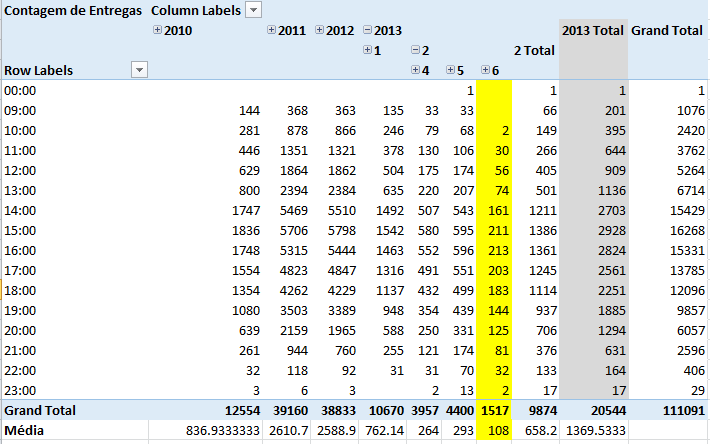
,[Morada Facturacao].[Distrito - Número Da Porta].[Freguesia].&[Lisboa]&[Lisboa]&[Madalena]

,[Morada Facturacao].[Distrito - Número Da Porta].[Freguesia].&[Lisboa]&[Lisboa]&[Mártires]

},([Measures].[Valor])

)

### Número médio de Entregas por Hora do Dia (no último mês)



Query MDX:

WITH MEMBER [Hora Tentativa].[Hora 24h].[O número de entregas médio por horas do dia no último mês] AS

ROUND(AVG ([Hora Tentativa].[Hora 24h].[Hora 24h],[Measures].[Contagem de Entregas]),0)

select

([Data Tentativa].[Ano].[Ano],TAIL(NONEMPTY ({[Data Tentativa].[Ano - Dia].[Mês Número].members}))) on columns

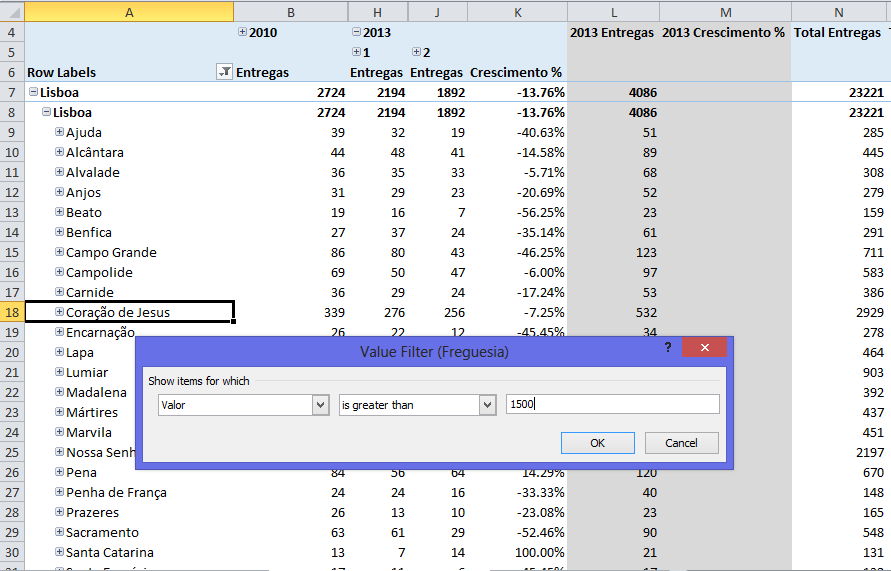
, [Hora Tentativa].[Hora 24h].[O número de entregas médio por horas do dia no último mês] on rows

from [Operacoes]

where

[Measures].[Contagem de Entregas]

### Percentagem de crescimento das entregas urgentes, no último trimestre, para as freguesias com volume de faturação superior a 1.500€



Query MDX:

WITH MEMBER [Hora Tentativa].[Hora 24h].[O número de entregas médio por horas do dia no último mês] AS

ROUND(AVG ([Hora Tentativa].[Hora 24h].[Hora 24h],[Measures].[Contagem de Entregas]),0)

select

([Data Tentativa].[Ano].[Ano],TAIL(NONEMPTY ({[Data Tentativa].[Ano - Dia].[Mês Número].members}))) on columns

, [Hora Tentativa].[Hora 24h].[O número de entregas médio por horas do dia no último mês] on rows

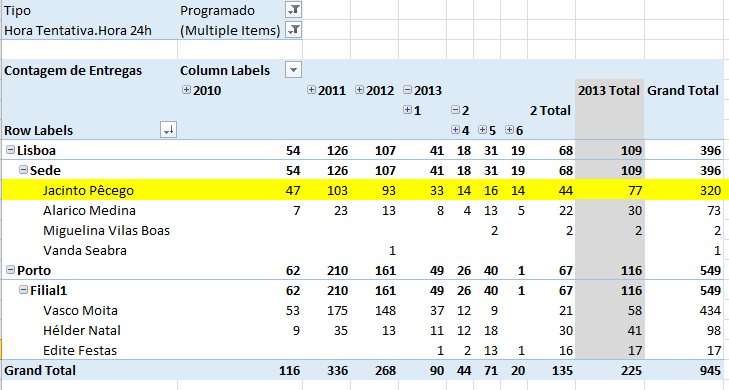
from [Operacoes]

where

[Measures].[Contagem de Entregas]

### Estafeta que mais entregas programadas fez fora de horas no último mês

Nota: Assumiu-se como “fora de horas” os períodos entre as 0 horas e as 8h, e as 18h e as 23h.



Query MDX:

select

([Data Tentativa].[Ano].[Ano],TAIL(NONEMPTY ({[Data Tentativa].[Ano - Dia].[Mês Número].members}))) on columns

,NONEMPTY([Funcionário].[Nome].[Nome],[Measures].[Contagem de Entregas]) on rows

from [Operacoes]

where

(([Measures].[Contagem de Entregas])

,[Tipo Servico].[Tipo].&[Programado]

,{[Hora Tentativa].[Hora 24h].&[00:00]:[Hora Tentativa].[Hora 24h].&[08:00]

,[Hora Tentativa].[Hora 24h].&[18:00]:[Hora Tentativa].[Hora 24h].&[23:00]})

### Balanço dos três últimos anos e respectivas diferenças absolutas

Query MDX:

with member [Measures].[Diferença]

as'

IIF(

IsEmpty(ParallelPeriod([Meses].[Ano - Mês Número].[Ano],1,[Meses].[Ano - Mês Número].CurrentMember))

,"NA"

,(

[Measures].[Valor]

- ([Measures].[Valor],ParallelPeriod([Meses].[Ano - Mês Número].[Ano],1,[Meses].[Ano - Mês Número].CurrentMember))

)

)'

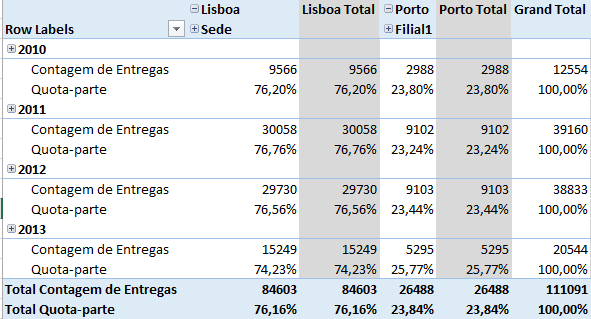
select

LastPeriods(3,[Meses].[Ano - Mês Número].[Ano].&[2012]) on columns

,{([Measures].[Valor]),([Measures].[Diferença])} on rows

from [Financeiro]

### Quota-parte que cada loja tem no volume total das entregas, em cada ano



Query MDX:

with member [Measures].[Loja Quota %]

as

'[Measures].[Contagem de Entregas]

/Sum([Funcionário].[Localização - Loja].[Loja],[Measures].[Contagem de Entregas])'

,format\_string='Percent'

select

NON EMPTY({[Funcionário].[Localização - Loja].[Loja]}) on columns ,

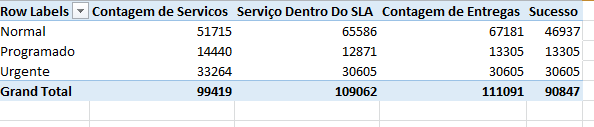
NON EMPTY({[Data Tentativa].[Ano].[Ano]}) on rows

from [Operacoes]

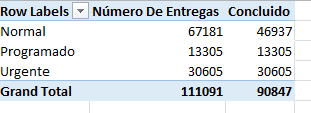
where [Measures].[Loja Quota %]

### Outras análises

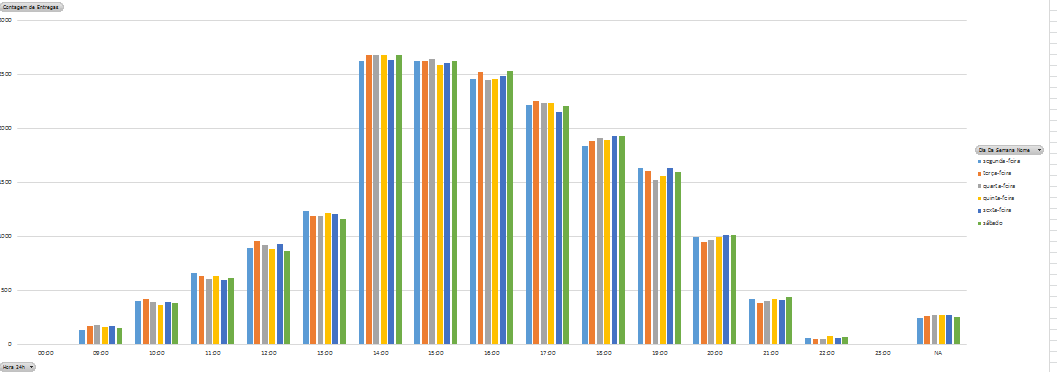
#### Entregas por tipo de Serviço e cumprimento do SLA



#### Número de Entregas e de Serviços concluídos por tipo de serviço



#### Evolução das Entregas



# Anexos

Um Data Warehouse e os seus datamarts são soluções construídas à medida para uma empresa, que levam em conta os seus processos de negócio, a forma com está organizada e a sua necessidade de análises e reports, pelo que a re-utilização, ainda que parcial não será possível.

Talvez por esta caraterística refletida na ferramenta da Microsoft nos tenha sido dificil criar uma forma fácil de deployment do ETL noutros sistemas e noutras configurações (neste caso os postos individuais de cada membro do grupo). Apesar do conjunto de indicações abaixo sempre foi necessário no Visual Studio proceder a alterações (refresh das ligações) para resolver todos os erros do ETL.

O processo geral e as alterações a fazer antes de abrir a solucao com.isel.siad.sv1213.bit.sln, consiste no fazer das seguintes configuracoes:

1) Correr batch RUNME.cmd (Run As Administrator)

2) Através da linha de comandos executar batch RUNMESQL.cmd passando como parâmetro a instância onde ficará o Data Wharehouse

3) Configurar os servidores alvos, e directorias locais nos ficheiros:

\com.isel.siad.sv1213.bit.ssis\BitMainPackage.Connections.dtsConfig

\com.isel.siad.sv1213.bit.ssis\BitMainPackage.Variables.dtsConfig

\com.isel.siad.sv1213.bit.ssis\FinacialPackage.Connections.dtsConfig

\com.isel.siad.sv1213.bit.ssis\FinacialPackage.Variables.dtsConfig

\com.isel.siad.sv1213.bit.ssis\OperationalPackage.Connections.dtsConfig

\#XMLA\02-create-com isel siad sv1213 bit ssas.xmla

BIT será a conexão para a base de dados, no DW, onde se encontra os dados do OLTP sobre a qual correrá o ETL

BITDW será a conexão para o servidor sql, DW, onde se encontra a base de dados de Data Staginng

BITDW.DataStaging/Data Staging BIT será a conexão para a base de dados, DW, de Data Staging

*FinacialPackagePath*: caminho completo para o ficheiro \com.isel.siad.sv1213.bit.ssis\FinacialPackage.dtsx

OperationalPackagePath: caminho completo para o ficheiro \com.isel.siad.sv1213.bit.ssis\OperationalPackage.dtsx

4) Executar o script \#XMLA\02-create-com isel siad sv1213 bit ssas.xmla sobre o Analysis Services do DataWharehouse

Ao abrir solução "com.isel.siad.sv1213.bit.sln" e durante a utilização, se for pedida alguma password colocar "siad\_bit\_dev" (sem aspas).

Se for perguntando para actualizar a connection string fechar essa janela de diálogo, para que a mesma não seja actualizada e mantenha os valores do ficheiros de configuração.

Ignorar erros (apenas surgiram na 1ª abertura/execução, porque é necessário que a primeira tarega do ETL seja executada)

Executar pacote BitMainPackage.dtsx (este pacote chamará os outros pacotes).

# Considerações Finais:

## Geral

Os modelos criados tinham por objetivo responder aos requisitos expressos no enunciado do trabalho, onde incluímos alguns requisitos adicionais (nomeadamente análises) que nos pareceram relevantes. Num projeto real, seria necessário enquadrar os requisitos dos utilizadores finais, em paticular as necessidades emergentes da utilização do sistema e compatibilizar com outras utilizações dos recursos informaticos, que poderiam ter impacto, por exemplo na janela de bacup, ou eventualmente em merges com outras companhias ou áreas de negócios.

Um projeto de Data Warehouse, parece-nos por isso um trabalho mais continuado que não se esgota com a primeira versão (por muito boa e analisada que possa ter sido).

No caso das análises, as tabelas de dimensões integram um conjunto diversificado de atributos, ou colunas, pelos quais os indicadores de negócio considerados nas tabelas de factos poderão ser avaliados e analisados. Estes atributos integram, habitualmente descrições que permitem contextualizar as métricas em análise.

## Melhorias

Neste processo de aprendizagem de implementação do modelo Dimensional, tomámos algumas decisões, que fomos repensando durante o projeto algumas das quais levaram às implementações entre as fases atrás identificadas.

## Ferramentas

Neste projecto foram usadas as ferramentas da Microsoft, cuja utilização beneficia com a fluência e prática de utilização, pelo que algumas das dificuldades encontradas se deverão dever à nossa (falta de) experiência. Contudo as ferramentas Microsoft parecem também ter as suas limitações ou contrariedades, quando alguma coisa não é feita bem “à primeira” e um processo de debug nem sempre claro ou esclarecedor. Por diversas vezes a solução encontrada passou, pelo refazer do código/ processos montados. A generalização das packages não está fácil ou “ergonómica” em particular parece haver situações em que perde a ligação definida e que é necessário revisitar os links para as reestabelecer.

Do ponto de vista funcional não encontramos limitações. De forma geral usar a ferramenta por default costuma correr bem e ser fácil, ao contrário parece difícil de expandir a ferramenta ou o seu uso.

Para sincronização do desenvolvimento em grupo foi utilizada a ferramenta/serviço de controlo de versão [github](https://github.com/), estando todo o repositório disponível: <https://github.com/RASMiranda/com.isel.siad.sv1213.bit>

Com a parte de data mining fizemos uma aula prática e por si parece ser uma área para exploração futura. Parece-nos que a área de Data Mining por si só merecia uma cadeira semestral, por precisar de um conhecimento muito aberto e muito ligado à experimentação.