



## 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 análises 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*

<b>1. INTRODUÇÃO .....</b>	<b>3</b>
1.1. MELHORIAS E CORREÇÕES RELATIVAS ÀS FASES 1 E 2 DO PROJETO.....	3
<b>2. TERCEIRA E QUARTA FASES .....</b>	<b>4</b>
2.1. GERAL .....	4
2.2. DESENVOLVIMENTO DOS CUBOS OLAP .....	4
2.2.1. Construção dos cubos .....	4
2.2.2. Nomenclatura.....	4
2.2.3. Tipo de Armazenamento cubos .....	5
2.2.4. Cubo Financeiro.....	5
2.2.5. Cubo Operações .....	6
2.3. ANÁLISE .....	8
2.3.1. Volume de faturação realizado em 2012, por mês e tipo de entrega, nas freguesias da Baixa Pombalina de Lisboa .....	8
2.3.2. Número médio de Entregas por Hora do Dia (no último mês) .....	9
2.3.3. Percentagem de crescimento das entregas urgentes, no último trimestre, para as freguesias com volume de faturação superior a 1.500€.....	10
2.3.4. Estafeta que mais entregas programadas fez fora de horas no último mês	11
2.3.5. Balanço dos três últimos anos e respectivas diferenças absolutas .....	12
2.3.6. Quota-parte que cada loja tem no volume total das entregas, em cada ano	13
2.3.7. Outras análises .....	13
<b>3. ANEXOS .....</b>	<b>15</b>
<b>4. CONSIDERAÇÕES FINAIS:.....</b>	<b>16</b>
4.1. GERAL .....	16
4.2. MELHORIAS .....	17
4.3. FERRAMENTAS .....	17

## 1. Introdução

O presente relatório tem por objetivo endereçar as fases 3 e 4 do Trabalho Prático de SIAD do semestre de Verão de 2012/2013. Como resultado de uma análise crítica 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.

### 1.1. 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 uma 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 “Pré-stagging” mantemos uma tabela com a data e hora da última extração do ambiente OLTP e recorreremos 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.

## 2. Terceira e Quarta Fases

### 2.1. 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.

### 2.2. Desenvolvimento dos cubos OLAP

#### 2.2.1. 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.

#### 2.2.2. 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.

### 2.2.3. 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 híbrido.

### 2.2.4. Cubo Financeiro

#### 2.2.4.1. Business Intelligence

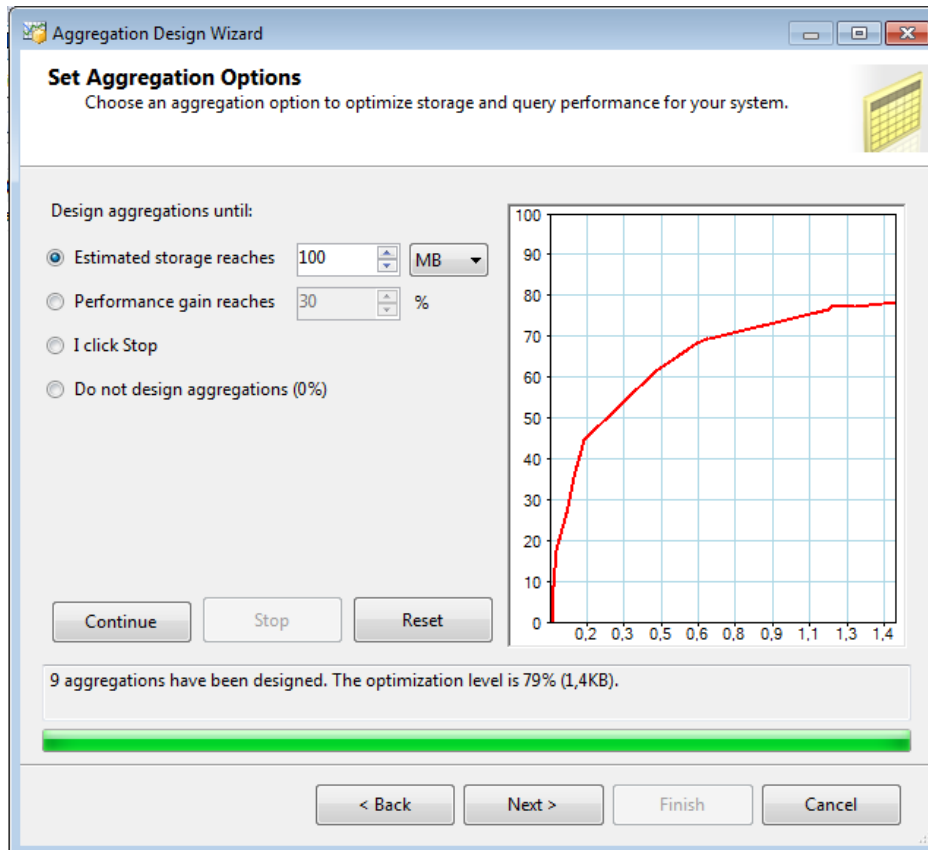
Foi adicionada *Business Intelligence* à dimensão **Rubrica**, como *Account*, em que **Custo** corresponde às despesas, e **Proveito** ao rendimento, e também à dimensão **Meses** como Tempo.

#### 2.2.4.2. Hierarquias e relações entre atributos

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

### 2.2.4.3. 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.

### 2.2.5. Cubo Operações

#### 2.2.5.1. Business Intelligence

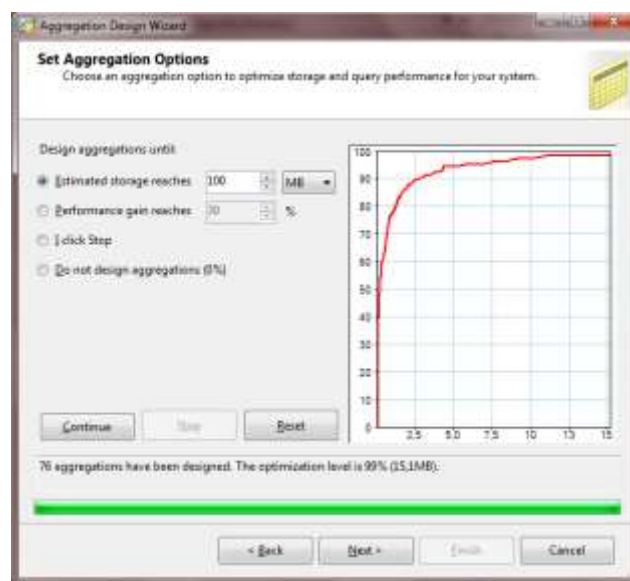
Foi adicionada *Business Intelligence* à dimensão **Data** como Tempo, à dimensão **Morada** como Geográfica, e à dimensão **Cliente** como Costumer.

### 2.2.5.2. Hierarquias e relações entre atributos

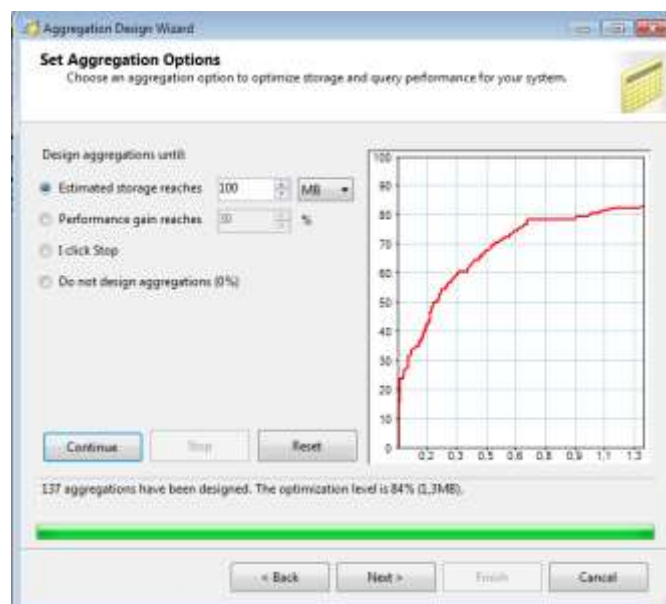
Foram definidas Hierarquias, e respetivas relações, para a dimensão de Data, Hora, Funcionário (por local e por chefe), Morada, Cliente (por designação/empresa).

### 2.2.5.3. 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:

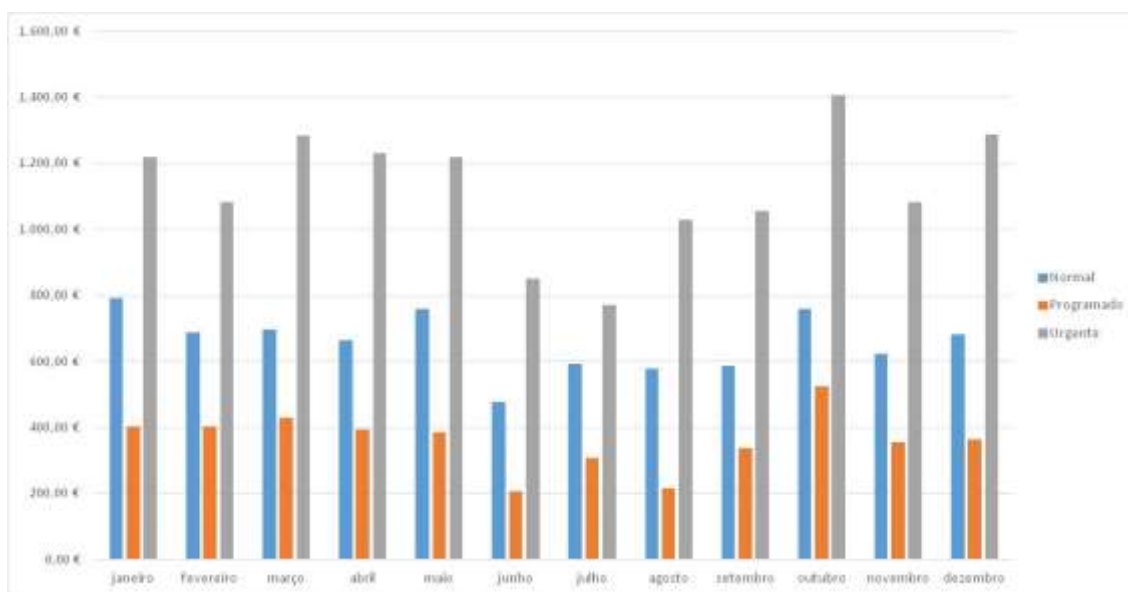


## 2.3. Análise

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

Nota: Assumi-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](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])
)
```



## 9

Versão 2.0 – 05.2013

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

	A	B	H	J	K	L	M	N
		2010	2013			2013 Entregas	2013 Crescimento %	Total Entregas
			+1	+2				
Row Labels		Entregas	Entregas	Entregas	Crescimento %			
Lisboa		2724	2194	1892	-13.76%	4086		23221
Lisboa		2724	2194	1892	-13.76%	4086		23221
Ajuda		39	32	19	-40.63%	51		285
Alcântara		44	48	41	-14.58%	89		445
Alvalade		36	35	33	-5.71%	68		308
Anjos		31	29	23	-20.69%	52		279
Beato		19	16	7	-56.25%	23		159
Benfica		27	37	24	-35.14%	61		291
Campo Grande		86	80	43	-46.25%	123		711
Campolide		69	50	47	-6.00%	97		583
Carnide		36	29	24	-17.24%	53		386
Coração de Jesus		339	276	256	-7.25%	532		2929
Encarnação		26	22	17	-45.45%	34		278
Lapa								464
Lumiar								903
Madalena								392
Mártires								437
Marvila								451
Nossa Senhora								2197
Pena		84	56	64	14.29%	120		670
Penha de França		24	24	16	-33.33%	40		148
Prazeres		26	13	10	-23.08%	23		165
Sacramento		63	61	29	-52.46%	90		548
Santa Catarina		13	7	14	100.00%	21		131

Value Filter (Freguesia)

Show items for which

Valor is greater than 1500

OK Cancel

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]
```

### 2.3.4. 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.

Tipo	Programado										
Hora Tentativa.Hora 24h	(Multiple Items)										
Contagem de Entregas	Column Labels										
	2010	2011	2012	2013					2013 Total	Grand Total	
				1	2			2 Total			
Row Labels					4	5	6				
Lisboa	54	126	107	41	18	31	19	68	109	396	
Sede	54	126	107	41	18	31	19	68	109	396	
Jacinto Pêcego	47	103	93	33	14	16	14	44	77	320	
Alarico Medina	7	23	13	8	4	13	5	22	30	73	
Miguelina Vilas Boas						2		2	2	2	
Vanda Seabra			1							1	
Porto	62	210	161	49	26	40	1	67	116	549	
Filial1	62	210	161	49	26	40	1	67	116	549	
Vasco Moita	53	175	148	37	12	9		21	58	434	
Hélder Natal	9	35	13	11	12	18		30	41	98	
Edite Festas				1	2	13	1	16	17	17	
Grand Total	116	336	268	90	44	71	20	135	225	945	

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]}}
```

### 2.3.5. 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]
```

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

Row Labels	<input type="checkbox"/> Lisboa <input type="checkbox"/> Sede	Lisboa Total	<input type="checkbox"/> Porto <input type="checkbox"/> Filial1	Porto Total	Grand Total
<b>+ 2010</b>					
Contagem de Entregas	9566	9566	2988	2988	12554
Quota-parte	76,20%	76,20%	23,80%	23,80%	100,00%
<b>+ 2011</b>					
Contagem de Entregas	30058	30058	9102	9102	39160
Quota-parte	76,76%	76,76%	23,24%	23,24%	100,00%
<b>+ 2012</b>					
Contagem de Entregas	29730	29730	9103	9103	38833
Quota-parte	76,56%	76,56%	23,44%	23,44%	100,00%
<b>+ 2013</b>					
Contagem de Entregas	15249	15249	5295	5295	20544
Quota-parte	74,23%	74,23%	25,77%	25,77%	100,00%
<b>Total Contagem de Entregas</b>	<b>84603</b>	<b>84603</b>	<b>26488</b>	<b>26488</b>	<b>111091</b>
<b>Total Quota-parte</b>	<b>76,16%</b>	<b>76,16%</b>	<b>23,84%</b>	<b>23,84%</b>	<b>100,00%</b>

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 %]
```

### 2.3.7. Outras análises

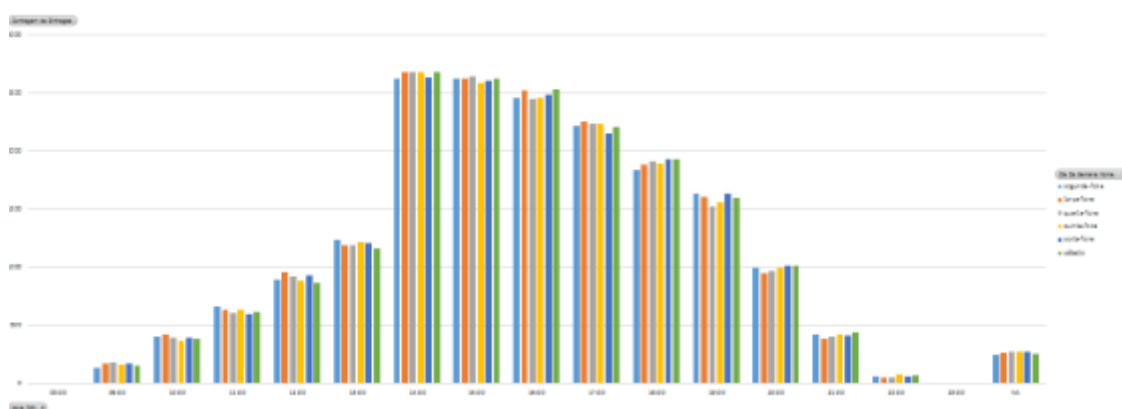
#### 2.3.7.1. Entregas por tipo de Serviço e cumprimento do SLA

Row Labels	Contagem de Serviços	Serviço Dentro Do SLA	Contagem de Entregas	Sucesso
Normal	51715	65586	67181	46937
Programado	14440	12871	13305	13305
Urgente	33264	30605	30605	30605
<b>Grand Total</b>	<b>99419</b>	<b>109062</b>	<b>111091</b>	<b>90847</b>

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

Row Labels	Número De Entregas	Concluido
Normal	67181	46937
Programado	13305	13305
Urgente	30605	30605
<b>Grand Total</b>	<b>111091</b>	<b>90847</b>

### 2.3.7.3. Evolução das Entregas



### 3. 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 característica refletida na ferramenta da Microsoft nos tenha sido difícil 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.xml`

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.xml a 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 tarefa do ETL seja executada)

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

## 4. Considerações Finais:

### 4.1. 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 particular as necessidades emergentes da utilização do sistema e compatibilizar com outras utilizações dos recursos informáticos, que poderiam ter impacto, por exemplo na janela de backup, 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.

## 4.2. 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.

## 4.3. 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.