Licenciatura em Engenharia Informática

Aprendizagem Organizacional

Objectivo: Utilização dos conceitos de Data Warehouse e Business Intelligence usando SQL Server 2008

Índice

Apresentação	 1
Objetivo	 3
Modelação do sistema	 5
Criação da Estrutura Multidimensional	 10
Processar o cubo	 18
Visualizar A informação Analítica	18

Apresentação

Pretende-se com esta ficha a utilização básica de um exemplo sobre os objectivos apresentados no início deste documento. Caso o aluno considere que já possui os conhecimentos suficientes acerca dos conteúdos desta matéria, o docente sugere que consulte/implemente os tutoriais contidos nos seguintes endereços web:

• Informação Adicional:

http://msdn.microsoft.com/en-us/sqlserver/cc511477

http://www.mssqltips.com/category.asp?catid=17

http://msftisprodsamples.codeplex.com/

http://www.mssqltips.com/sqlservertip/2450/ssis-package-deployment-model-in-sql-server-denali-part-1-of-2/

http://msdn.microsoft.com/en-us/library/bb522859.aspx

• SQL Server 2008 Books Online:

 $\underline{\text{http://msdn.microsoft.com/en-us/library/bb543165(sql.100).aspx}}$

Virtual Labs

http://msdn.microsoft.com/en-us/sqlserver/cc138238

• WebCasts:

http://www.microsoft.com/latam/sql/2008/learning/webcasts.mspx

• Videos:

http://www.ssistutorial.com/video 685ce7 SSIS-SQL-Server-Business-Intel.php

• Como instalar exemplos do SQL Server 2008:

http://msftdbprodsamples.codeplex.com/wikipage?title=Installing%20Databases



Licenciatura em Engenharia Informática

Aprendizagem Organizacional

http://www.ssas-info.com/analysis-services-faq/29-mgmt/242-how-install-adventure-works-dw-database-analysis-services-2005-sample-database

Pretende-se com esta ficha a utilização básica de um exemplo sobre os objectivos apresentados no início deste documento. Caso o aluno considere que já possui os conhecimentos suficientes acerca dos conteúdos desta matéria, o docente sugere que consulte/implemente os tutoriais contidos nos seguintes endereços web:

Microsoft Training Kits:

http://www.microsoft.com/download/en/details.aspx?id=16281

How to Install SQL Server 2008:

http://msdn.microsoft.com/en-us/library/ms167593.aspx

http://www.sqlserverclub.com/essentialguides/how-to-install-sql-server-2008-step-by-step-guide.aspx

Manage the Data Warehouse:

http://www.sql-server-performance.com/2008/management-data-warehouse/

http://siddhumehta.blogspot.com/2009/07/sql-server-2008-data-warehousing-video.html

http://www.ssas-info.com/VidasMatelisBlog/55 sql-server-2008-management-data-warehouse

An introduction to SQL Server data warehousing concepts

http://searchsqlserver.techtarget.com/video/An-introduction-to-SQL-Server-data-warehousing-concepts

Building a data warehousing and BI solution

http://searchsqlserver.techtarget.com/video/Building-a-data-warehousing-and-Bl-solution?videoId=e9e5802b3e7f6210VgnVCM1000000d01c80aRCRD

Adventure Works Sample Data Warehouse

http://technet.microsoft.com/en-us/library/ms124623.aspx

Microsoft SQL Server Community Projects & Samples

http://sqlserversamples.codeplex.com/

Training Series

http://www.sqlservercentral.com/Training/

Building Your First Cube with SQL Server 2008 R2 Analysis Services:

http://msdn.microsoft.com/en-us/SQL10R2BYFBI-TrainingCourse SQL10R2BYFBI05-HOL-01

Po

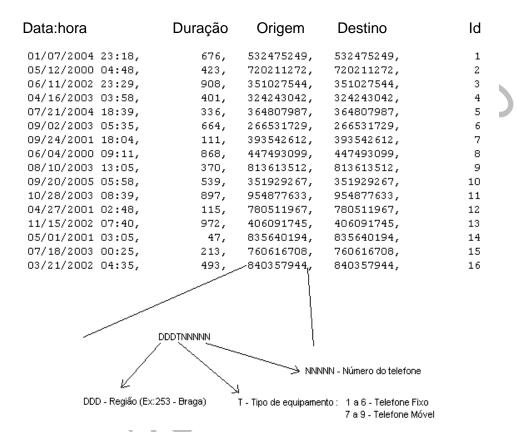
Tutorial n.º 4

Licenciatura em Engenharia Informática

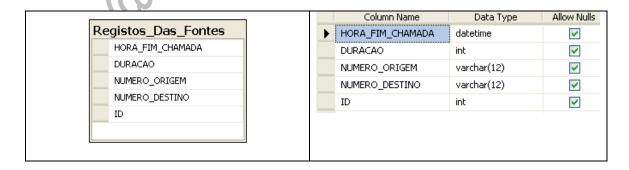
Aprendizagem Organizacional

Objetivo

Considere um cenário de aplicabilidade dos Sistemas de Informação na área das operadoras telefónicas. A nível de fontes de dados, existe uma única fonte com um milhão de registos com a seguinte estrutura:



Em termos de estrutura relacional ela é a seguinte:



Algumas regras em relação ao número origem e destino:



Licenciatura em Engenharia Informática

Aprendizagem Organizacional

- 1. Se o terceiro dígito estiver compreendido entre 1 e 6 podemos considerar como um aparelho que utiliza a rede fixa;
- 2. Caso contrário, ou seja, se o terceiro dígito for 7, 8 ou 9 podemos considerar como um aparelho que utiliza a rede móvel.

Outras informações são obtidas a partir do número origem, como o nome da operadora ou informações mais especificas da região.

Os dados auxiliares são os seguintes:

```
1,DOMINGO
Dias da semana:
                                                                                  SEGUNDA-FEIRA
                                                                            3, TERÇA-FEIRA
                                                                            4.QUARTA-FEIRA
                                                                           5, QUINTA-FEIRA
                                                                           6,SEXTA-FEIRA
7,SÁBADO
Regiões:
                                                                                     ","BRAGA","","253"
","LISBOA","","100"
","ALMADA","10 ","101"
","SETUBAL","10 ","102"
","FAMALICAO","1 ","252"
","GUIMARAES","1 ","254"
","BARCELOS","1 ","255"
                                                                       "1
"10
"11
"12
"2
"3
"4
"5
                                                                                                                                                                                                 <-- Pertence à região 1 - Brag
<-- Pertence à região 1 - Brag
<-- Pertence à região 1 - Brag
                                                                                             "BARCELOS","1 ","2
"VIANA DO CASTELO",
"PORTO","","200"
                                                                                                                                                                   ","256"
                                                                          "6
"7
                                                                                             "PORTO","","200"
"VILA NOVA DE GAIA","6
"MATOSINHOS","6 ","202"
"RIO TINTO","6 ","203"
                                                                                                                                                                    ","201" <-- Pertence à região 6 - Porto
" <-- Pertence à região 6 - Porto
<-- Pertence à região 6 - Porto
                                                                          "8
"9
                                                                                         - Fixo PORTÁTIL",10
- Fixo NAO PORTÁTIL",10
- Fixo RDIS 2 Linhas",10
- Fixo RDIS 4 Linhas",10
- Fixo CENTRAL 100 Linhas",10
- Fixo CENTRAL 500 Linhas",10
- Fixo CENTRAL 500 Linhas",10
- Telemóvel TMN",11
- Telemóvel OPTIMUS",11
- Telemóvel VODAFONE",11
Fixo".
                                                                             Ο,
Tipos de telefone:
                                                                            7,"/ - lele
8,"8 - Tele
<mark>9,"9 - Tele</mark>
10,"Fixo",
11,"Móvel",
```

Durante o processo de exploração de informação, várias observações foram observadas pelos utilizadores/agentes de decisão. Depois de uma entrevista (elaboração de inquéritos) aos agentes de decisão, algumas dessas questões são as seguintes:



Licenciatura em Engenharia Informática

Aprendizagem Organizacional

- " É preciso visualizar quantos minutos foram gastos em ligações telefónicas feitas de telemóveis para telemóveis móveis ao fim de semana."
- " Qual é o tempo médio de ligações, separadas por região, durante os dias da semana?"
- " Após a meia noite, quantas ligações de telefones fixos de uma cidade foram feitas para telefones móveis de outra cidade ?"
- " É possível criar um gráfico comparando a quantidade de minutos de todas as segundas-feiras só com os telefones fixos ? E com os telefones móveis e fixos ?"

Tendo em consideração estas e outras informações, considerou-se que os utilizadores precisavam de maior flexibilidade na exploração de dados, nomeadamente para:

- Facilitar a visualização da informação referente às ligações;
- Comparação da informação com vários anos anteriores;
- Analisar quantos minutos foram gastos em ligações telefónicas efetuadas por telemóveis para telemóveis nos fins de semana dos últimos 4 anos;
- Depois de uma determinada hora (Ex: 1 hora da manhã) quantas ligações de telefone fixo da região de Braga foram efetuadas para telefones móveis ou fixos para a região do Porto.
- Analisar por região, o tempo médio de ligações durante a semana e fins de semana;
- Apresentar um gráfico comparando o número de minutos de todos os sábados, com telefones fixos, com telefones fixos e móveis.

Modelação do sistema

Depois do levantamento de requisitos, juntamente com os utilizadores, foi definido que, para as informações/questões levantadas dever-se-ia criar uma estrutura multidimensional com sete dimensões e três medidas:

Tipo	Identificador	Descrição
	DIM_ID	Identificador do Registo
Chave de		Esta dimensão vai permitir que o utilizador
Dimensão	DIM_TEMPO	faça o filtro por qualquer dia, hora ou minuto do mês.



Licenciatura em Engenharia Informática

Aprendizagem Organizacional

		Esta dimensão tem como objetivo permitir
	DIA CEMANIA	ao utilizador escolher sobre qual(ais) dia(s)
	DIA_SEMANA	da semana gostaria de consultar as
		medidas.
	REGIAO_NUM_ORIGEM	Esta dimensão vai permitir filtrar de
		qual(ais) região(ões) as ligações se
		originam.
	REGIAO_NUM_DESTINO	Esta dimensão vai permitir filtrar para
		qual(ais) regiões(s) as ligações estão feitas.
	TIPO_NUM_ORIGEM	Esta dimensão permite que se filtrem as
		medidas pelo tipo de telefone que originou a
		ligação. Os tipos possíveis são telefones
		móveis ou fixos, de acordo com a regra
		apresentada anteriormente.
		Esta dimensão permite que se filtrem as
	TIPO_NUM_DESTINO	medidas pelo tipo de telefone de destino da
		ligação.
	DURAÇÃO	Este medida vai trazer todos os minutos
		somados de acordo com os filtros aplicados
		pelos valores escolhidos nas dimensões.
Medidas	LIGAÇÕES	Este medida apresenta a quantidade de
		ligações de acordo com os filtros aplicados
		pelos valores escolhidos nas dimensões
	70,	Esta medida apresenta o TML, Tempo
		Médio de Ligações efetuadas, de acordo
	V	com os filtros aplicados pelos valores
	TML (Tempo médio de	escolhidos nas dimensões. O TML é obtido
	ligações)	dividindo-se o valor da medida DURAÇÃO
		pelo valor da medida LIGAÇÔES para o
		conjunto de dados resultando dos filtros
		aplicados pelos valores das dimensões.

Crie esta estrutura e importe os dados diretamente dos seguintes ficheiros:

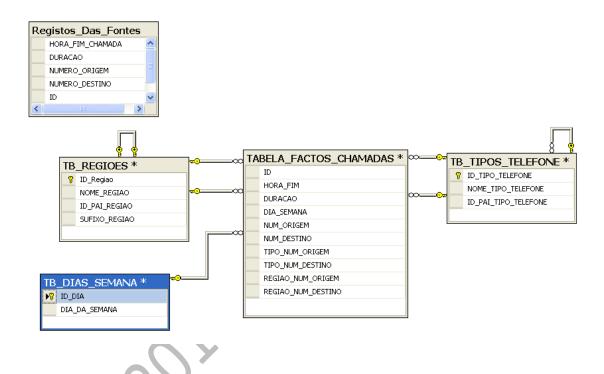


Licenciatura em Engenharia Informática

Aprendizagem Organizacional

CAB-Tutorial-7-Criar_Tabela_Factos_passo1
CAB-Tutorial-7-Tabela_DIAS_SEMANA
CAB-Tutorial-7-Tabela_REGIOES
CAB-Tutorial-7-Tabela_TIPOS_TELEFONE
CAB-Tutorial-7-Registos_Das_Fontes

criando as seguintes tabelas:



No sentido de auxiliar a separação de vários campos a interligar a um outro (neste caso da tabela de factos), crie as seguintes vistas:

```
CREATE VIEW VW_TIPO_NUM_origem
AS
SELECT ID_TIPO_TELEFONE, NOME_TIPO_TELEFONE, ID_PAI_TIPO_TELEFONE
FROM TB_TIPOS_TELEFONE

GO

CREATE VIEW VW_TIPO_NUM_destino
AS
SELECT ID_TIPO_TELEFONE, NOME_TIPO_TELEFONE, ID_PAI_TIPO_TELEFONE
FROM TB_TIPOS_TELEFONE
```



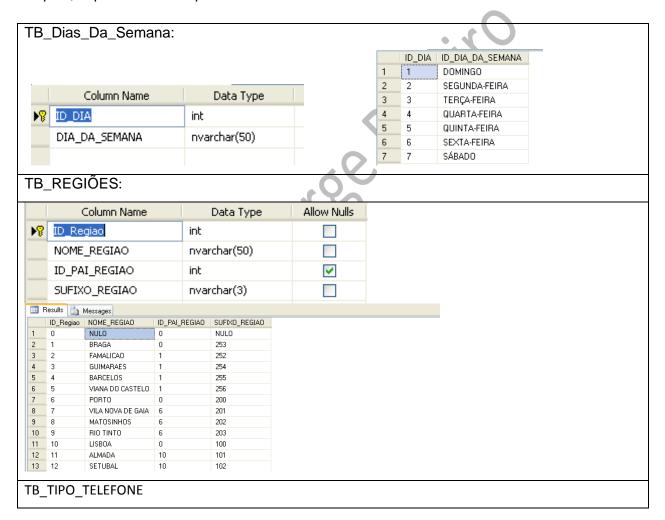
Licenciatura em Engenharia Informática

Aprendizagem Organizacional

```
CREATE VIEW VW_TIPO_Regiao_origem
AS
SELECT ID_Regiao,NOME_REGIAO,ID_PAI_REGIAO,SUFIXO_REGIAO
FROM TB_REGIOES

GO
CREATE VIEW VW_TIPO_Regiao_destino
AS
SELECT ID_Regiao,NOME_REGIAO,ID_PAI_REGIAO,SUFIXO_REGIAO
FROM TB_REGIOES
```

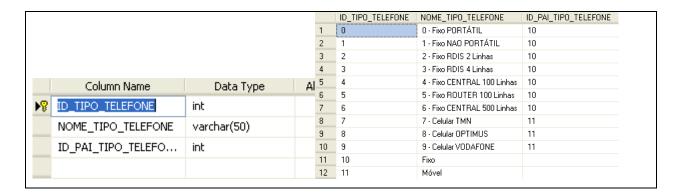
Depois, importe os dados para as tabelas associadas:





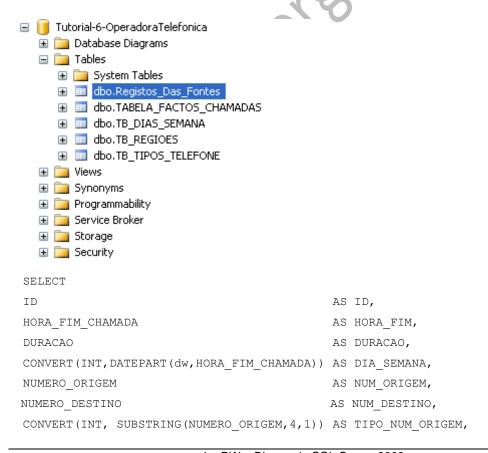
Licenciatura em Engenharia Informática

Aprendizagem Organizacional



Importe os dados do ficheiro "Registos_Das_Fontes.txt" para a tabela Registos_Das_Fontes.

Crie a tabela de factos através do seguinte comando T-SQL e, iremos acrescente depois as medidas





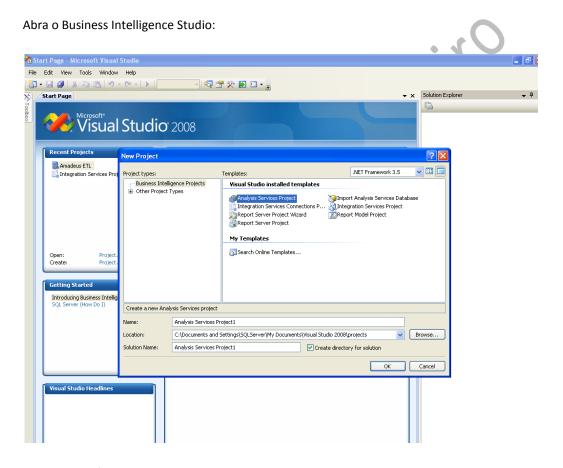
Licenciatura em Engenharia Informática

Aprendizagem Organizacional

CONVERT(INT, SUBSTRING(NUMERO_DESTINO,4,1)) AS TIPO_NUM_DESTINO,
CONVERT(NVARCHAR(3), SUBSTRING(NUMERO_ORIGEM,1,4)) AS REGIAO_NUM_ORIGEM,
CONVERT(NVARCHAR(3), SUBSTRING(NUMERO_DESTINO,1,4)) AS REGIAO_NUM_DESTINO

INTO TABELA_FACTOS_CHAMADAS
FROM Registos_Das_Fontes;

Criação da Estrutura Multidimensional

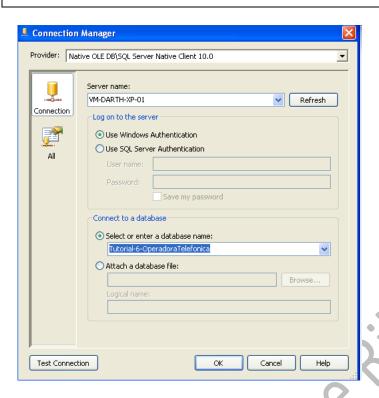


Crie a ligação à base de dados:

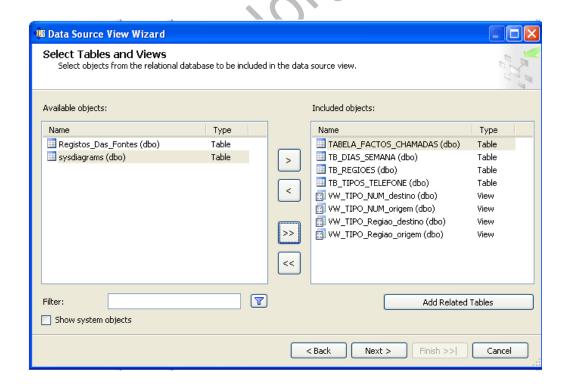


Licenciatura em Engenharia Informática

Aprendizagem Organizacional



Crie o Datasource view:

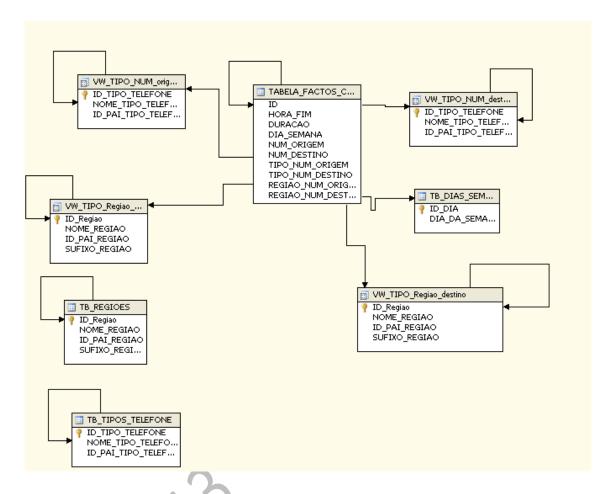




Licenciatura em Engenharia Informática

Aprendizagem Organizacional

Crie a ligação entre as tabelas:



Vamos criar agora as dimensões dia_da_semana:

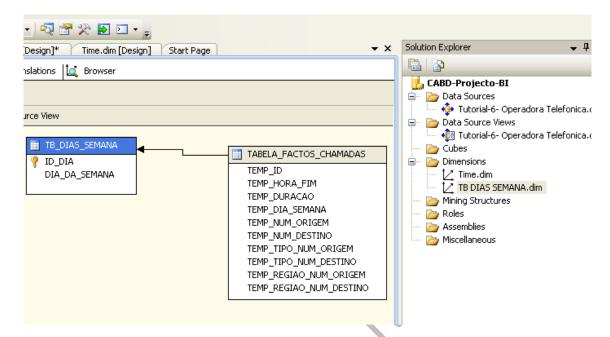




Licenciatura em Engenharia Informática

Aprendizagem Organizacional

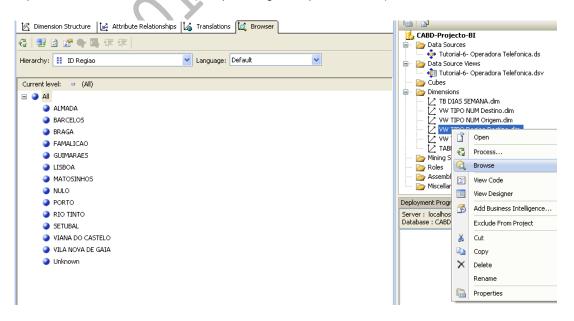
Visualize a tabela de factos e a tabela de dimensão TB_Dias_Semana:



Crie as dimensões:



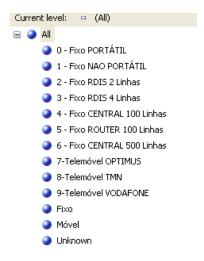
Explore os dados das dimensões (ex: tipo de região e tipo de telefone):





Licenciatura em Engenharia Informática

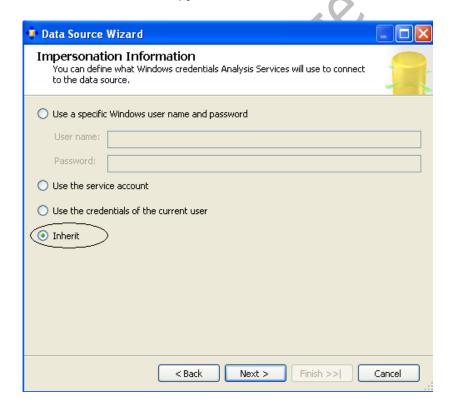
Aprendizagem Organizacional



Se lhe aparecer um erro de "impersonate mode", terá de eliminar o link do datasource:



e criar um novo, mas com a opção "inherit":

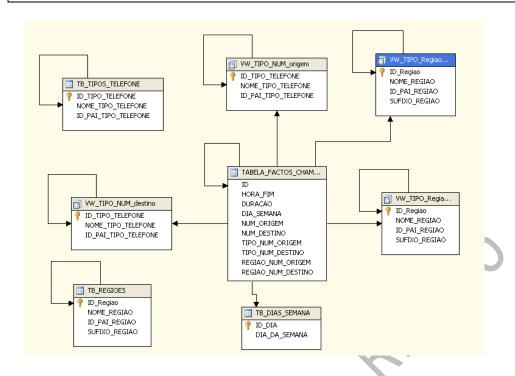


Proceda aos relacionamentos das dimensões: Vistas regiões e tipo de telefone (origem e destino):



Licenciatura em Engenharia Informática

Aprendizagem Organizacional



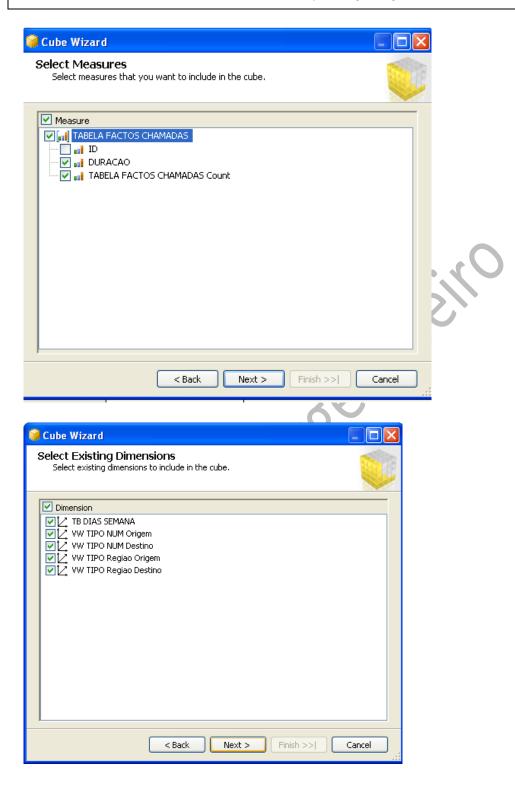
Vamos agora configurar uma medida. Selecicone "Cube-New Cube":



Selecione as medidas:



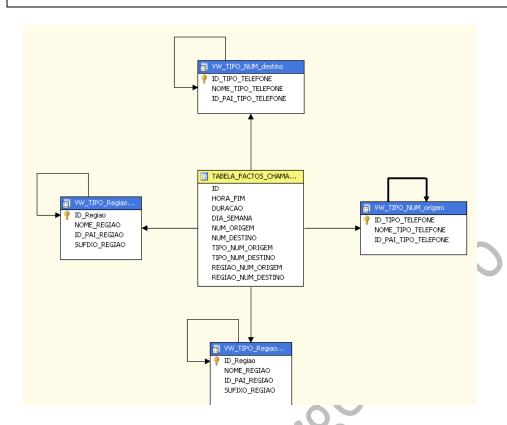
Licenciatura em Engenharia Informática





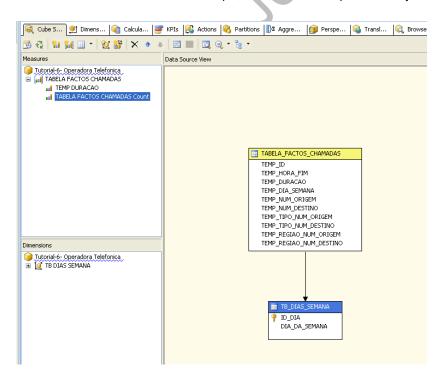
Licenciatura em Engenharia Informática

Aprendizagem Organizacional



Vamos criar a medida "Ligações" a partir da informação já existente:

Por defeito é criada a medida "Count" que, neste caso corresponde ao objectivo da medida "ligações":

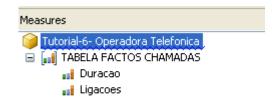


Renomeie para:

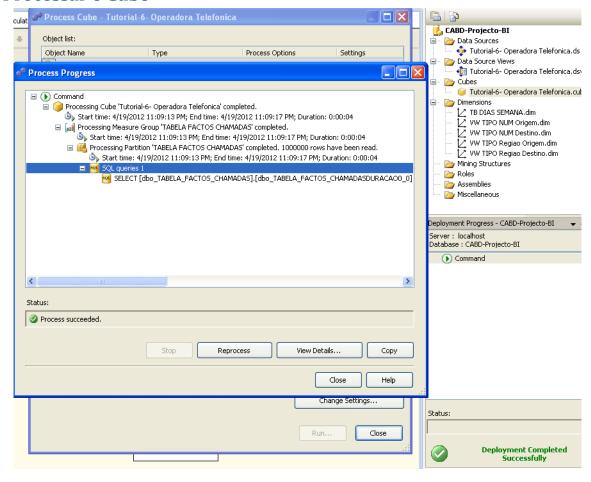


Licenciatura em Engenharia Informática

Aprendizagem Organizacional



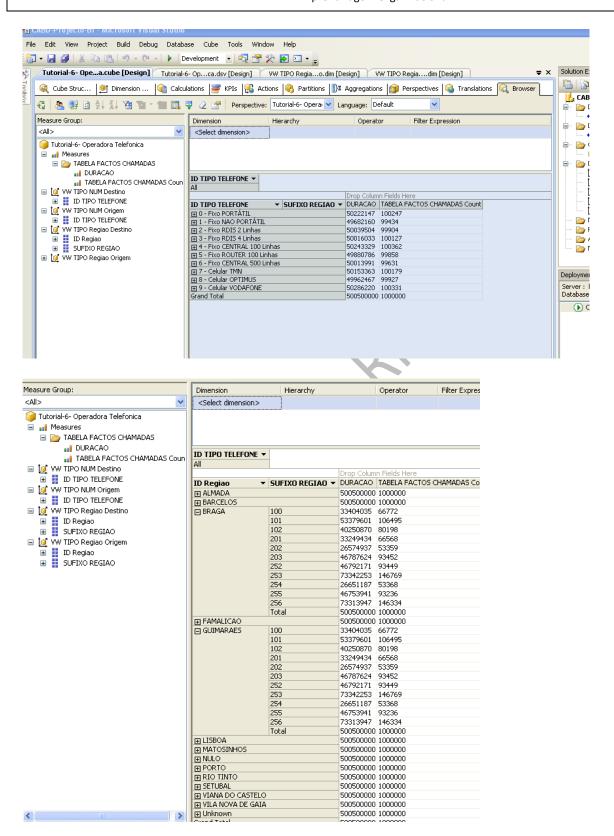
Processar o Cubo



Visualizar a informação analítica



Licenciatura em Engenharia Informática

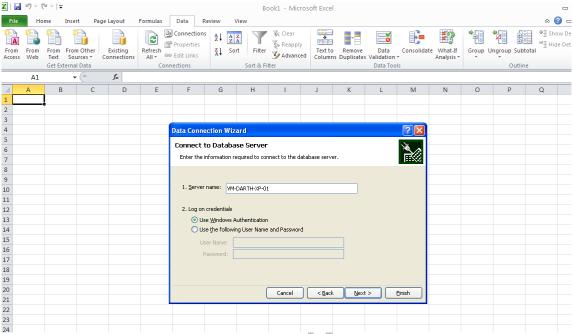


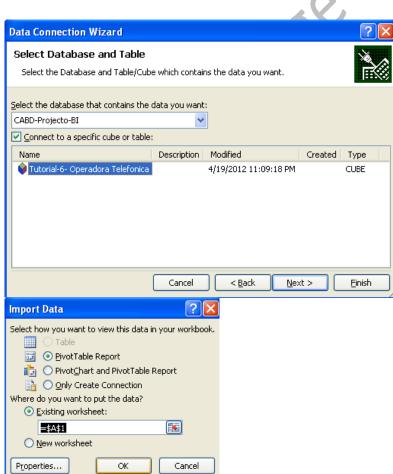


Licenciatura em Engenharia Informática

Aprendizagem Organizacional

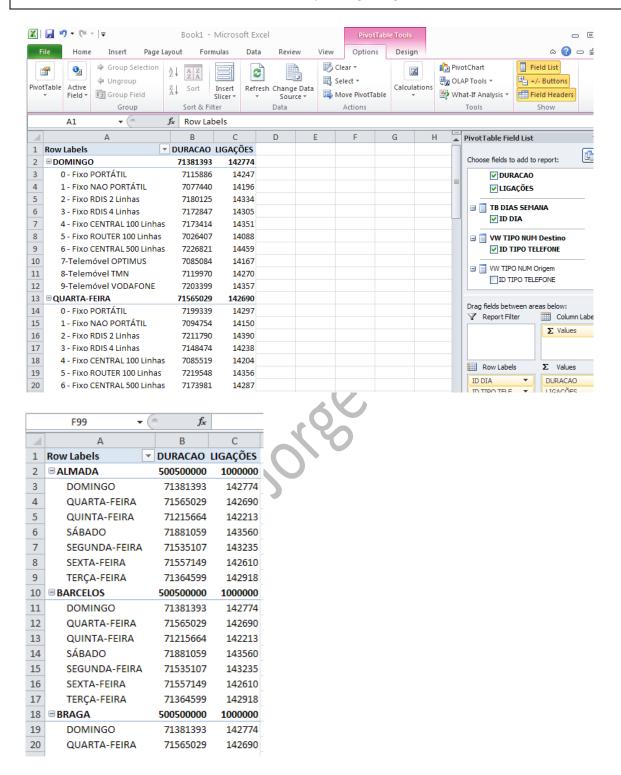
Utilizando o Microsoft EXCEL:







Licenciatura em Engenharia Informática



Licenciatura em Engenharia Informática

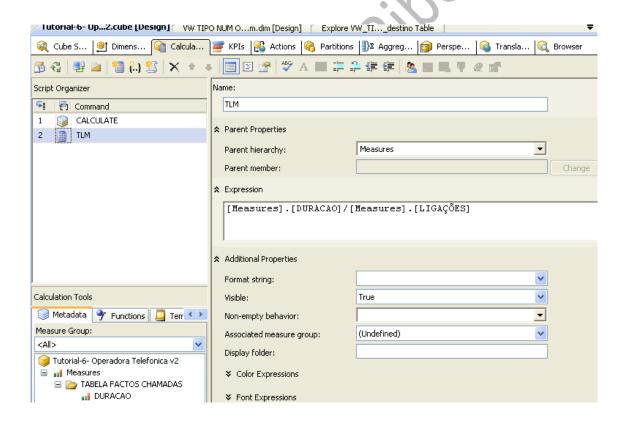
Aprendizagem Organizacional

Tarefa:

- a) Analise os requisitos do projecto com os resultados obtidos. Verifique em que pontos pode melhorar o desempenho de utilização para assegurar os requisitos.
- b) Crie agora a medida TML Tempo Médio de Ligações efetuadas. Relembre que o objectivo desta medida é, de acordo com os filtros aplicados pelos valores escolhidos nas dimensões. O TML é obtido dividindo-se o valor da medida DURAÇÃO pelo valor da medida LIGAÇÔES para o conjunto de dados resultando dos filtros aplicados pelos valores das dimensões. Pesquise na literatura como criar medidas calculadas usando o SQL Server 2008.

Ajuda: Consulte os seguintes sítios sobre "Expressões Calculáveis":

http://technet.microsoft.com/en-us/library/ms166568.aspx

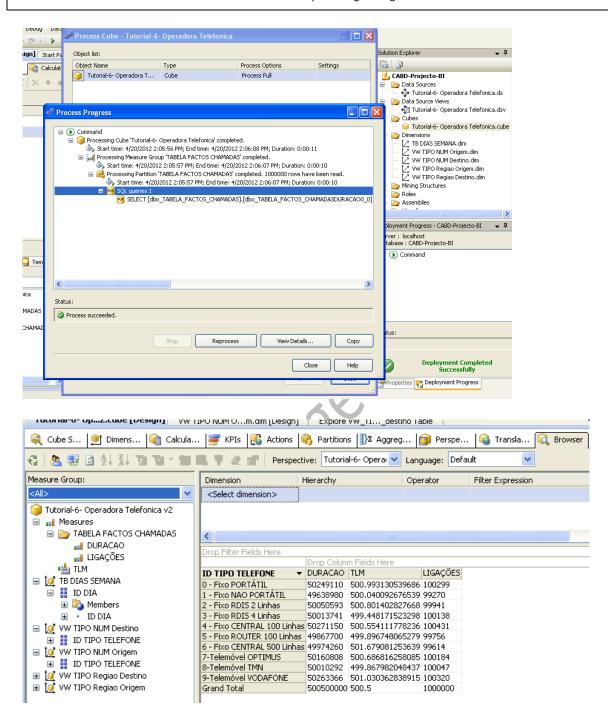


Selecione o separador Browser e selecione "Reconnect", de modo a actualizar os dados no servidor:





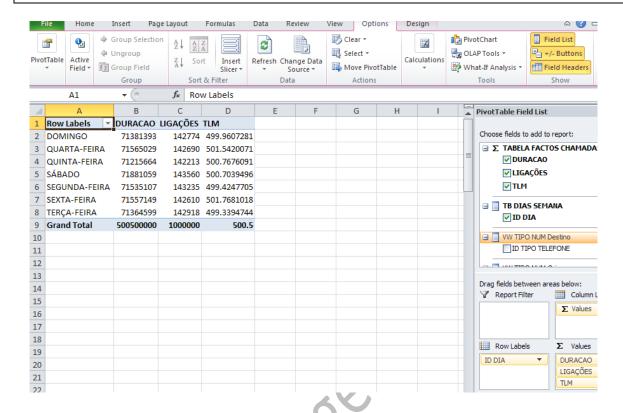
Licenciatura em Engenharia Informática





Licenciatura em Engenharia Informática

Aprendizagem Organizacional



Vamos criar uma dimensão tempo. Para entender melhor o objectivo veja o seguinte vídeo:

http://channel9.msdn.com/Blogs/rdoherty/Demo-Developing-a-Time-Dimension-with-SQL-Server-2008-R2-Analysis-Services?format=html5

Para criar uma dimensão tempo gerando os dados consulte:

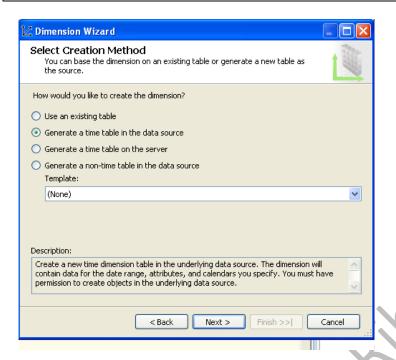
http://msdn.microsoft.com/en-us/library/ms174832.aspx

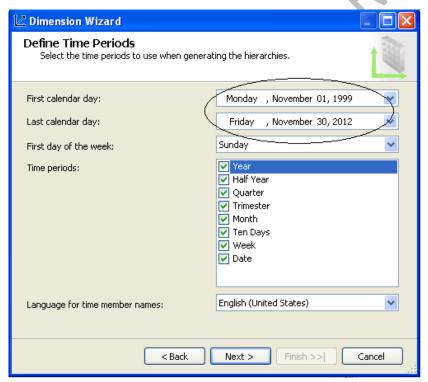
http://www.mssqltips.com/sqlservertip/1454/built-in-time-dimension-and-time-intelligence-in-sql-server-analysis/



Licenciatura em Engenharia Informática

Aprendizagem Organizacional

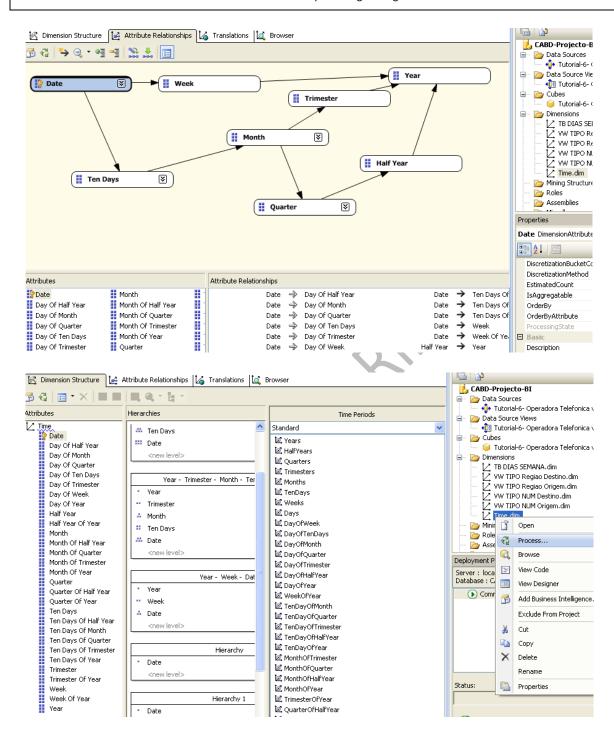




Analise os relacionamentos desta dimensão:



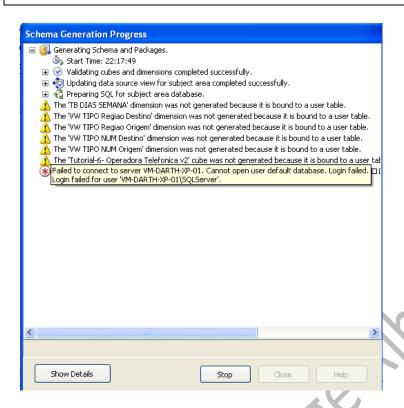
Licenciatura em Engenharia Informática





Licenciatura em Engenharia Informática

Aprendizagem Organizacional

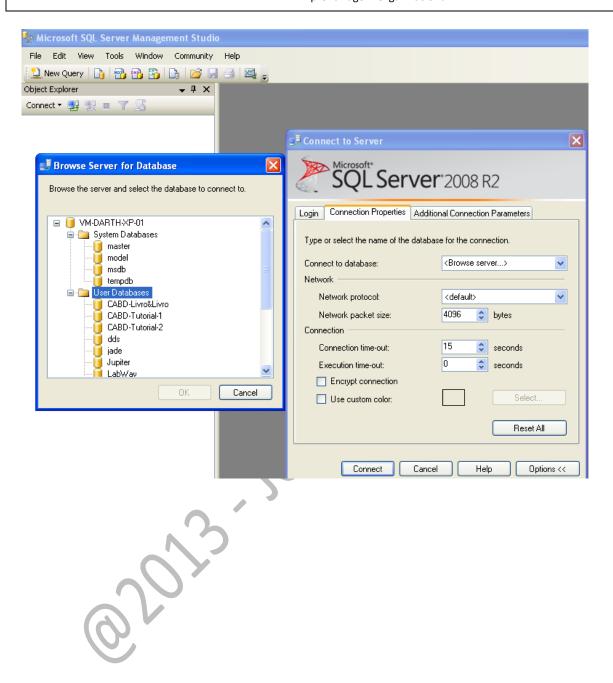


Para resolver este erro de acesso à base de dados por defeito do SQL Server, aceda ao SQL Server e no item "opções"



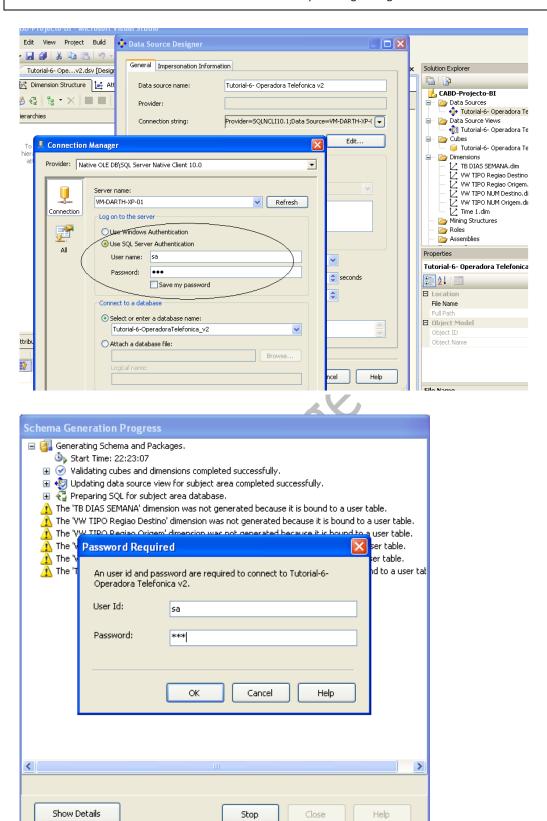


Licenciatura em Engenharia Informática





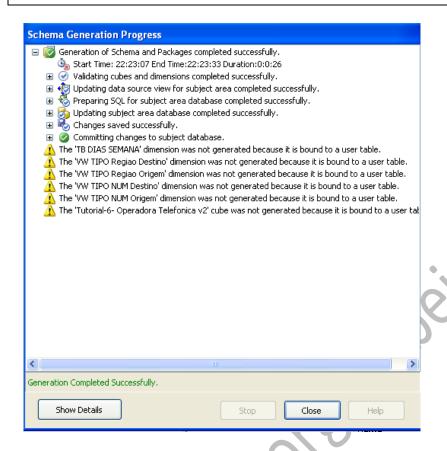
Licenciatura em Engenharia Informática



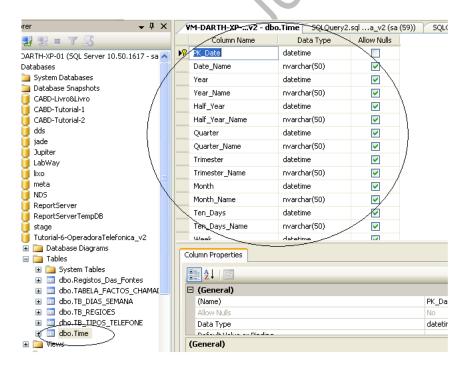


Licenciatura em Engenharia Informática

Aprendizagem Organizacional



Consulte o SQL Server 2008:

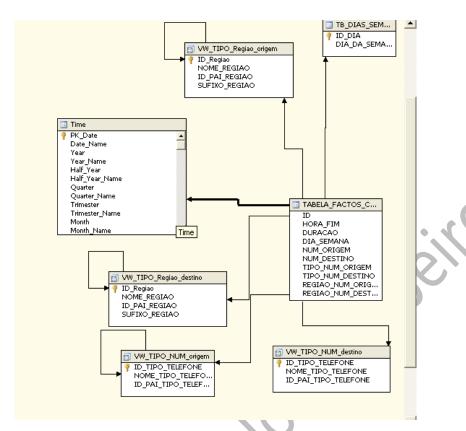




Licenciatura em Engenharia Informática

Aprendizagem Organizacional

Crie o relacionamento entre a "Time" e o atributo de dimensão "hora_fim":



Adicione a dimensão "time" á estrutura do cubo de dados:

