

DEIS - Departamento de Engenharia Informática e de Sistemas

ISEC - Instituto Superior de Engenharia de Coimbra

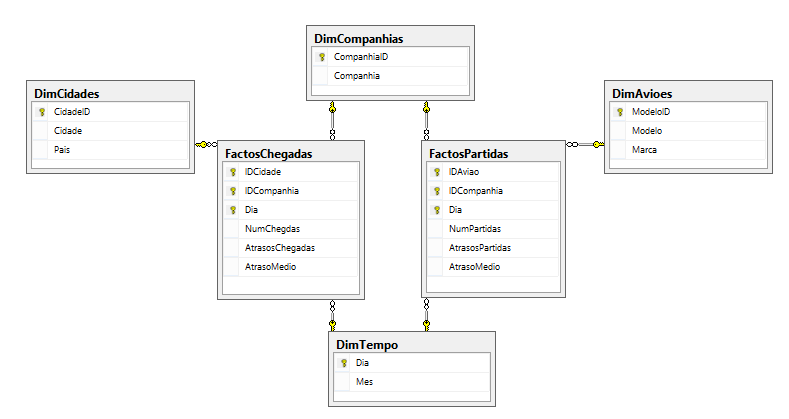
Sistemas de Informação II

Exercícios Ficha 1 - Datawarehouses

**1.** Um aeroporto pretende implementar uma datawarehouse com vista ao apuramento das seguijtes informações:

* Número de chegadas e acumulado dos atrasos nessas chegadas cidade origem, país origem, companhia aérea, dia e mês;
* Número de partidas e acumulado dos atrasos nessas partidas por modelo do avião, marca do avião, companhia aérea, dia e mês.

Desenhe uma data warehouse adequada a estes objetivos. Identifique as chaves primárias e externas, as medidas e as hierarquias que considerar.

****

Hierarquias: Dia -> Mês

CidadeID -> País

ModeloID -> Marca

**2.** A seguinte datawarehouse destina-se a um aeroporto que pretende criar um sistema OLAP para determinar

* O número de chegadas, acumulado dos atrasos nessas chegadas e atraso médio, por cidade origem, país origem, companhia aérea, dia e mês;
* O número de partidas, acumulado dos atrasos nessas partidas e atraso médio, por modelo de avião, marca de avião, companhia aérea, dia e mês.

A informação encontra-se inicialmente registada no sistema OLTP que também se representa em seguida.

Para a carga inicial da DW num projeto ETL, escreva os queries destinados a carregar as tabelas DimAvioes e FactosChegadas.

DimAvioes:

select Modelos.ModeloID, Modelo, Marca

from Modelos, Marcas

where Modelos.marcaID = Marcas.MarcaID

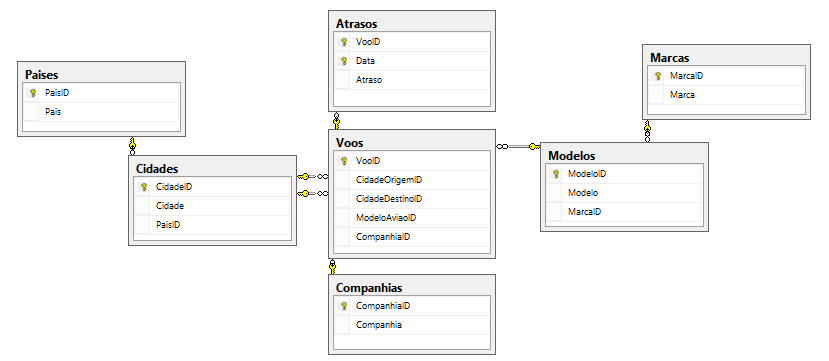
FactosChegadas:

select Voos.CidadeOrigemID, Voos.CompanhiaID, Data, sum(Atraso) as SomaAtrasos, count(Voos.VooID) as NumVoos

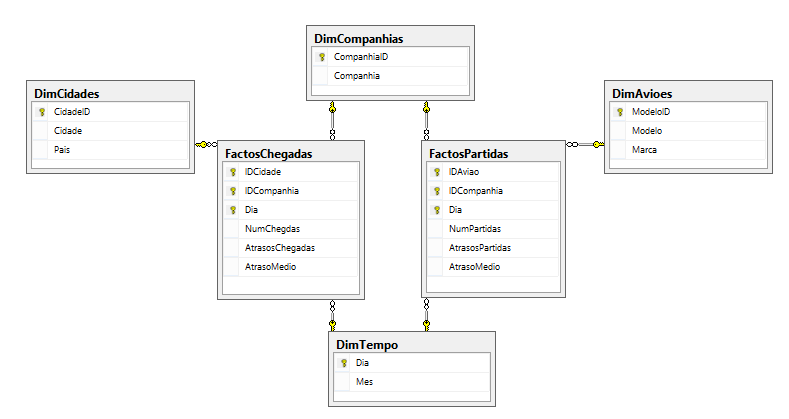
from Voos, Atrasos, Companhias

where Voos.VooID=Atrasos.VooID

group by Voos.CidadeOrigemID, Voos.CompanhiaID, Atrasos.Data

**OLTP **

**DW**

****

**3.** A empresa X dispõe de N lojas distribuídas pelo país. Cada loja dispõe do seu próprio ERP onde se pretendem implementar *datamarts* iguais em todas as lojas para apuramento dos seguintes valores e definição do seguinte KPI:

* VQ = Vendas em quantidade
* VV = Vendas em valor
* STI = Stock disponível, em quantidade, no início de cada período de tempo
* CQ = Compras em quantidade
* CV = Compras em valor
* PMC = Preço Médio de Custo = CV / CQ
* MLB = Margem de Lucro Bruta = (VV – (VQ × PMC))/VV
* KPM é um KPI para a MLB, com um *target* T, constante, de 15%

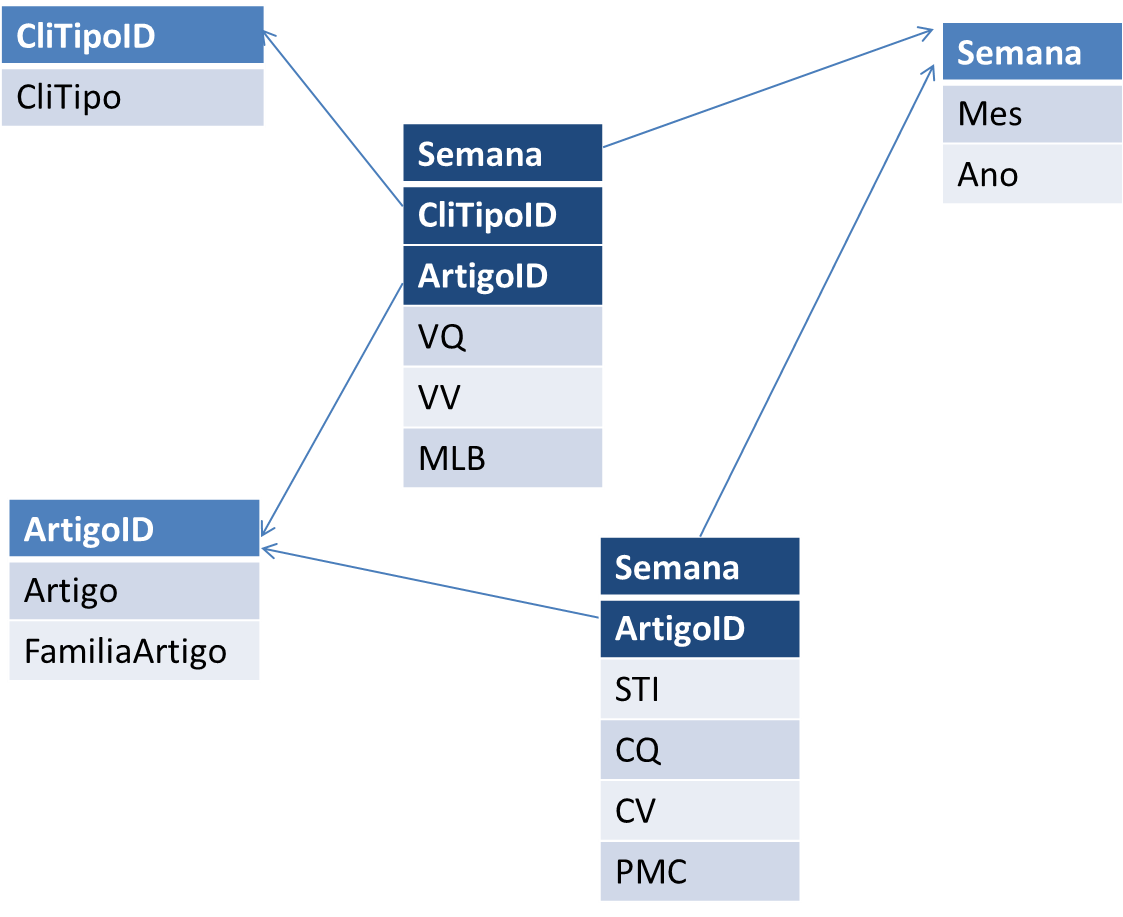
**a)** Desenhe uma datawarehouse em estrela destinada a permitir a visualização de todos os elementos constantes desta lista (incluindo KPM), por semana, mês, ano, artigo, família de artigo e tipo de cliente (A, B ou C). No diagrama assinale as PK’s, FK’s, medidas, dimensões, tipo de dimensão, e hierarquias.

**b)** Escreva, em pseudocódigo, a expressão de KPM para um indicador do tipo velocímetro (*gauge*)

**c)** Todas as medidas são aditivas? Justifique.

**a)** Como não é possível nem faz sentido saber o stock inicial nem compras (em quantidade e valor) por cliente, a DW ideal seria esta:

Vendas

****

Compras

PK’s assinaladas a cor mais escura

FK’s assinaladas pelas setas

Hierarquias:

1) Semana -> Mês

2) Semana -> Ano

3) ArtigoId -> FamiliaArtigo

As medidas calculadas (PMC e MLB) também poderiam não figurar na tabela de factos e ser definidas apenas como New Named Calculation quando da definição do cubo do servidor OLAP

**b)** KPM terá portanto de ser definido por

**KPM = MLB / 0.15**

Donde o comportamento do indicador em Gauge será do tipo:

**Caso**

**KPM >= 0.8**

**Valor\_Gauge = +1** ‘ Verde

**KPM < 0.8 E KPM > 0.5**

**Valor\_Gauge = 0** ‘ Amarelo

**KPM <= 0.5**

**Valor\_Gauge = -1** ‘ Vermelho

**Fim Caso**

A estrutura seria qq coisa deste tipo. Poderiam usar IF’s em vez de Caso. São importantes os valores -1, 0 e +1 de controlo da posição do Gauge. Limites 0.8 e os outros, poderiam ser arbitrados dentro de valores razoáveis.

**c)** As medidas são aditivas exceto:

* STI = Stock disponível, em quantidade, no início de cada período de tempo
* PMC = Preço Médio de Custo = CV / CQ
* MLB = Margem de Lucro Bruta = VV – (VQ × PMC)

STI não pode ser adicionada ao longo do tempo, mas pode ser adicionado por artigo. Por isso é semi-aditiva

PMC não pode ser adicionada ao longo de nenhuma dimensão: não aditiva

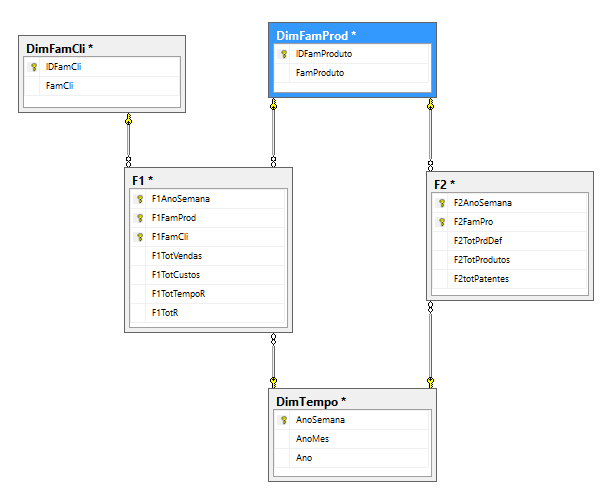
MLB não pode ser adicionada ao longo de nenhuma dimensão: não aditiva

**4.** Uma companhia fabril gere-se pelo modelo Balanced Scorecards e para isso definiu 4 KPI’s:

* ML = Margem de Lucro = Total\_de\_Vendas – Total\_de\_Custos \* 100
* Q = %Quebras = Número\_de\_Produtos\_Com\_Defeito / Número\_Total\_de\_Produtos \* 100
* NP = Número\_de\_Patentes
* TMR = TempoMédioReclamação = Total\_Tempo\_ Resposta\_a\_Reclamações / Núm.\_Total\_de\_ Reclamações
* ML e TMR devem ser obtidas por semana, mês e ano, por família de cliente e por família de produto
* Q deve ser obtida por semana, mês e ano, e por família de produto
* NP deve estar disponível por ano (pelo menos), e por família de produto

**a)** Desenhe uma datawarehouse em estrela destinada a permitir a visualização de todos os elementos constantes desta lista e também dos KPI’s acima indicados.

**b)** Quantos valores (aproximadamente) teriam de ser definidos para o indicador ML, para funcionamento da DW entre os anos de 2010 e 2015? Justifique.



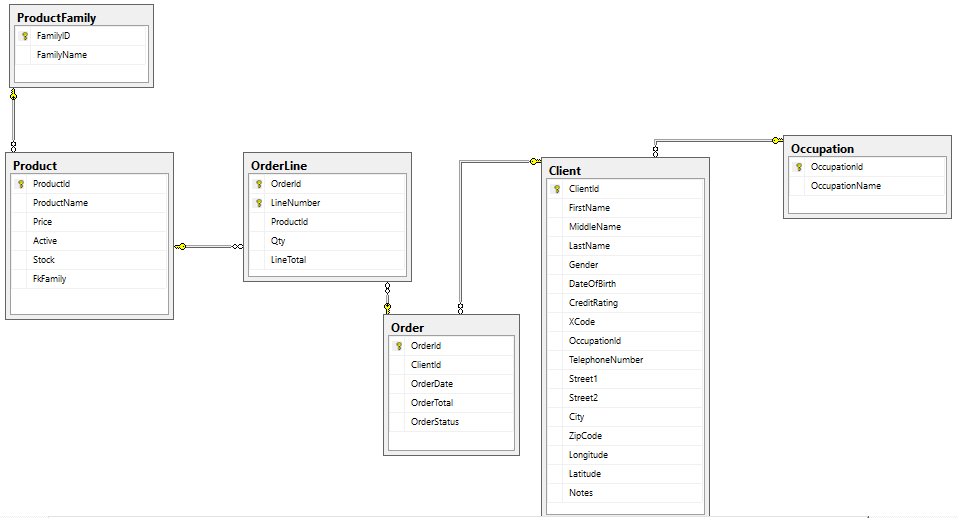
NOTA: poderiam ser 3 tabelas de factos, caso se pretendesse só e somente o número de patentes por ano.

**5.** A figura mostra a base de dados transacional de uma empresa produtora de tecidos.

**a)** Desenhe uma DW destinada a saber as vendas em quantidade e em valor, por família de produto, cliente, cidade, ocupação (do cliente), dia, mês e ano. No esquema marque as PKs e as hierarquias.

**b)** Escreva o query ou os queries que seria(m) necessário(s) para carregar a informação relativa ao cliente, nome da cidade e ocupação.

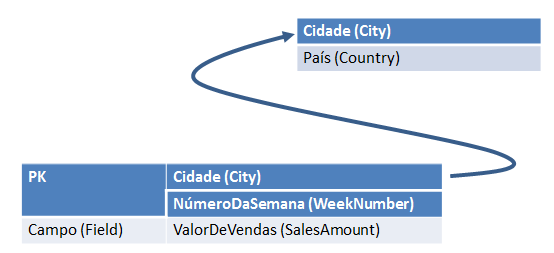
**c)** Escreva o query que seria necessário para carregar a tabela de factos



**6.** Desenhe uma datawarehouse em estrela, destinada a funcionar num hospital, e destinada a obter:

* O número de internamentos por serviço, por GDH (Grupo de Diags Homogéneos) e por ano
* O valor total das taxas moderadoras, por GDH e por ano, pagas pelos pacientes após alta
* O tempo médio de permanência no hospital, por GDH e por ano

**7.** O diagrama mostra uma datawarehouse composta apenas por uma tabela de factos e uma tabela de dimensões:



**a)** Na base de dados transacional, a Cidade está registada na tabela de facturas, e há várias facturas para o mesmo cliente. A tabela Países contém apenas o nome da cidade e o país a que a cidade pertence. Para obter dados para carregar a tabela de dimensões utilizou-se o seguinte query:

***Select Cidade, País***

***From Facturas, Paises***

***Where Facturas. Cidade = Países. Cidade***

Ao executar-se o módulo ETL de carga da tabela de dimensões, obteve-se um erro de chave primária duplicada. Corrija o código anexo.

*Select* ***Distinct*** *Cidade, País*

*From Facturas, Paises*

*Where Facturas. Cidade = Países. Cidade*

**b)** Na base de dados transacional, o Número da Semana e o Total de cada Factura estão disponíveis na tabela Facturas. Para obter dados para carregar a tabela de factos utilizou-se o seguinte query:

***Select Cidade, NúmeroDaSemana, SUM (TotalFactura) as Total***

***From Facturas***

***Group by Total***

Ao executar-se este módulo ETL obteve-se um erro. Corrija o código anexo.

*Select Cidade, NúmeroDaSemana, SUM (TotalFactura) as Total*

*From Facturas*

*Group by* ***Cidade, NúmeroDaSemana***

**c)** Esta tabela de factos pode ser actualizada diariamente? Justifique.

Poder, pode, mas exigiria operações de update do campo ValorDeVendas, o que não é usual. O mais prático, seria a actualização semanal desde que do ponto de vista dos requisitos da organização não houvesse problema em não ter o valor acumulado de cada semana à medida que ela decorre.