

Armazenamento de Dados

Data Warehouse para Análise de Despesa com Medicamentos

João Calhau, 36764 José Pimenta, 37158

Évora, 29 de Outubro de 2017



Tabela de Conteúdo

Lista de Figuras	ii
Lista de Quadros	iii
Introdução	1
Metodologia	2
Desenvolvimento	3
Base de Dados OLTP	3
Matriz em Bus	4
Métricas	6
Dimensões	7
Descrição da Tabela de Factos	15
Descrição da Tabela de Factos Sumarizada	15
Esquema em Estrela – Tabela de Factos	16
Esquema em Estrela – Tabela de Factos Sumarizada	17
Dimensões Role Playing	18
Hierarquias	18
Sistema de Metadados	19
Área de Estágio	19
Referências	20



Lista de Figuras

FIGURA 1 BASE DE DADOS OLTP	3
FIGURA 2 ESQUEMA EM ESTRELA DO PROCESSO DE NEGÓCIO	16
FIGURA 3 ESQUEMA EM ESTRELA DO PROCESSO DE NEGÓCIO SUMARIZADO	17
FIGURA 4 HIERARQUIA NA DIMENSÃO DATA	18



Lista de Quadros

TABELA 1 VERSÃO ORIGINAL DA MATRIZ BUS	5
TABELA 2 VERSÃO FINAL DA MATRIZ BUS	5
TABELA 3 TABELA DE MÉTRICAS	6
Tabela 4 Dimensão Localização	7
Tabela 5 Dimensão Hora	7
TABELA 6 DIMENSÃO DATA	8
TABELA 7 DIMENSÃO FARMÁCIA	9
TABELA 8 DIMENSÃO PACIENTE	10
Tabela 9 Dimensão Médico	11
TABELA 10 DIMENSÃO LABORATÓRIO	11
TABELA 11 DIMENSÃO MEDICAMENTO	
TABELA 12 DIMENSÃO SEMESTRE	14
TABELA 13 TABELA DE FACTOS	15
TABELA 14 TABELA DE FACTOS SUMARIZADA	15



Introdução

A revolução tecnológica sentida no século XX trouxe muitos benefícios ao ser humano. A partir desta data o ser humano desenvolveu bastante dados, de registar passando de tradicionalmente seriam registados em papel e arquivados numa estante, acessíveis a um número restrito de pessoas, passando para um método tecnológico que permite uma facilidade de distribuição, organização e acessibilidade de dados, as bases de dados. O modo como estes dados são guardados foi uma mais valia para verificar certo tipo de dados e avaliação dos mesmos, mas mesmo assim ainda há um caminho longo para se percorrer de modo a tornar a análise dos mesmos muito mais acessível.

Desse modo, pretendemos realizar este trabalho e criar uma data warehouse para análise de despesa com medicamentos pelo Estado. Pretende-se assim que esta data warehouse esteja dotada da informação necessária para que se analise os vários tipo de despesas suportadas pelo estado com a venda de medicamentos, entre outro tipo de análises de vendas consoante o tipo de análise pretendida, seja a análise por zona, tempo, especialidade médica, farmácia, entre outros.



Metodologia

A elaboração desta primeira parte do trabalho foi feita com a ajuda do Sublime Text para arquitetar o sistema relacional de SQL da base de dados criada primeiramente. Essa base de dados foi então usada para podermos ter uma ideia de quais as dimensões que iriamos utilizar no trabalho. De seguida utilizámos o Google Docs para elaborar a Fase 1 da entrega do trabalho e o Google Sheets para elaborar a matriz em bus.

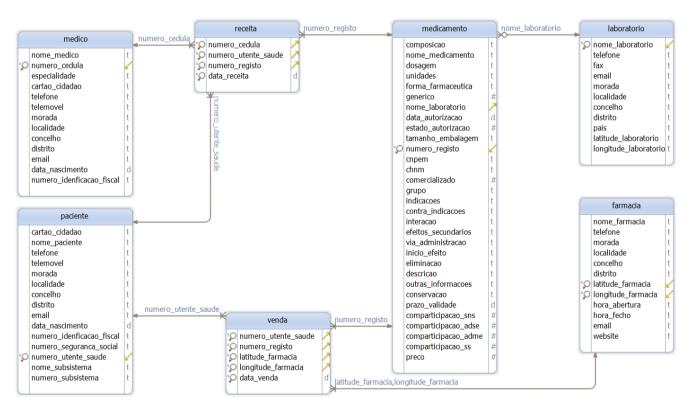


Desenvolvimento

Base de Dados OLTP

Antes de começar o trabalho propriamente dito, começámos por desenvolver uma base de dados relacional na linguagem postgreSQL. Para tal tivemos de pensar quais seriam as tabelas mais adequadas para o registo de medicamentos e a área da qual fazem parte, e assim criámos as seguintes tabelas como se verifica na Figura 1:

- laboratorio
- medicamento
- farmacia
- medico
- paciente
- receita
- venda



Generated using DbSchema

Figura 1 Base de dados OLTP



Matriz em Bus

Para a realização da matriz em bus, teve-se em conta o tipo de negócio que se pretendia realizar, neste caso, a análise das vendas de medicamentos comparticipados pelo estado. Neste tipo de análise, o grão mais atómico que se pretende verificar é a comparticipação da venda de uma caixa de medicamentos.

Tendo em conta o processo de negócio escolhido, foi altura de serem escolhidas algumas métricas para a devida análise do negócio escolhido, e para tal, decidimos escolher as seguintes métricas:

- Valor Comparticipado
- Valor Comparticipado por Subsistema de Saúde
- Valor Comparticipado por Especialidade Médica
- Valor Comparticipado por Laboratório
- Valor Comparticipado por Farmácia
- Valor Comparticipado por Medicamento
- Valor Comparticipado por Paciente

Para tal, inicialmente, verificámos a necessidade da criação de várias dimensões para a nossa data warehouse. Inicialmente criámos uma tabela com as dimensões:

- Farmácia
- Laboratório
- Medicamento
- Médico
- Paciente
- Data
- Receita
- Venda



A matriz bus que criámos com estes dados, Tabela 1, tinha alguns problemas, pelo que mais tarde, e com a ajuda do professor nas aulas, verificámos que a existência das tabelas Receita e Venda não eram necessárias como Dimensão, pois a tabela Venda acaba por ser a tabela de factos e a Receita é uma dimensão degenerada, pois retirando todos os atributos que se encontram simultaneamente nas outras tabelas, ficámos apenas com o atributo da referência da receita. Após as devidas alterações, e alguns debates nas aulas entre alunos e com o professor, verificámos que algumas tabelas teriam de ser acrescentadas, tais como:

- Local
- Hora

Finalmente, e após as devidas alterações terem sido implementadas, chegámos à matriz bus final, tal como podemos verificar na Tabela 2.

	Dimensões	Laboratório	Medicamento	Farmácia	Médico	Paciente	Receita	Venda	Data	Dimensionalidade
Processo	Análise de vendas de medicamento comparticipados pelo Estado	х	х	х	X	х	Х	Х	Х	8

Tabela 1 Versão Original da Matriz Bus

	Dimensões	Laboratório	Medicamento	Farmácia	Médico	Paciente	Local	Data	Hora	Dimensionalidade
Processo	Análise de vendas de medicamento comparticipados pelo Estado	Х	х	Х	Х	Х	Х	Х	х	8

Tabela 2 Versão Final da Matriz Bus



Métricas

O nosso processo pode ser avaliado por um variado tipo de métricas. Por exemplo, em relação à data, podemos avaliar por diferentes quantidades de tempo, i.e. por hora, dia, semana, mês, semestre, em relação à localização, podemos avaliar por concelho, distrito, entre outros.

Para tal, temos na seguinte figura, Tabela 3, o exemplo de algumas métricas a avaliar neste trabalho.

	Valor Comparticipado por Subsistema de Saúde	
	Valor Comparticipado por Especialidade Médica	
Métricas	Valor Comparticipado por Laboratório	
Wiethiode	Valor Comparticipado por Farmácia	
	Valor Comparticipado por Medicamento	
	Valor Comparticipado por Paciente	

Tabela 3 Tabela de Métricas



Dimensões

Com base na matriz em bus realizada, e com algumas alterações decididas posteriormente, as dimensões escolhidas para este trabalho foram:

 Dimensão Localização - dimensão localização com atributos referentes à localização.

DIMENSÃO LOCALIZAÇÃO	EXEMPLO
PK Chave de Localização	12012102
Coordenadas Geográficas	38°34'19.4"N 7°55'51.1"W
Latitude	38°34'19.4"N
Longitude	7°55'51.1"W
Número da Porta	1º Esquerdo, nº 13
Rua	Rua dos Amores
Código Postal	7000-402
Freguesia	Malagueira
Concelho	Évora
Distrito	Évora
Região do País	Alentejo
País	Portugal

Tabela 4 Dimensão Localização

A chave primária da Dimensão Localização é a surrogate key "Chave da Localização". Essa surrogate key é apenas um número inteiro natural.

• Dimensão Hora - dimensão hora composta por hora, minuto e segundo.

DIMENSÃO HORA	EXEMPLO
PK Chave da Hora	121020
Hora Completa	12:10:20
Descrição da Hora	12 horas, 10 minutos e 20 segundos
Horas	12
Minutos	10
Segundos	20

Tabela 5 Dimensão Hora



A chave primária da Dimensão Hora é a surrogate key "Chave da Hora". Essa surrogate key é especial, e é definida do modo HHMMSS. No caso de a hora estar em falta, uma surrogate key possível é por exemplo 400000, pois não existe hora 40.

• Dimensão Data - dimensão que contém as características da data.

DIMENSÃO DATA	EXEMPLO
PK Chave da Data	20170806
Data Completa	06/08/2017
Descrição da Data	6 de Agosto de 2017
Número do Dia da Semana	1
Número do Dia do Mês	6
Número do Dia do Trimestre	37
Número do Dia do Semestre	37
Número do Dia do Ano	218
Número do Dia Absoluto	310121
Número da Semana	32
Número do Mês	8
Número do Trimestre	3
Número do Semestre	2
Número do Ano	2017
Nome do Dia da Semana	Domingo
Nome do Mês	Agosto
É dia Útil	Não é dia útil
É Feriado	Não é feriado
É Ano Bissexto	Não é Ano bissexto
Nome da Estação do Ano	Verão
Data do Início da Semana	06/08/2017
Data do Fim da Semana	12/08/2017
Data do Início do Mês	01/08/2017
Data do Fim do Mês	31/08/2017
Data do Início do Trimestre	01/07/2017
Data do Fim do Trimestre	30/07/2017
Data do Início do Semestre	01/07/2017
Data do Fim do Semestre	31/12/2017

Tabela 6 Dimensão Data



A chave primária da Dimensão Data é a surrogate key "Chave da Data". Essa surrogate key é especial, pelo que é definida do modo YYYYMMDD. No caso de a data estar em falta ou ser data de algo que não aconteceu, por exemplo, uma surrogate key possível para cada um dos casos é por exemplo 40004040 e 40004141, pelo que nunca haverá problema com esta surrogate key, devido a não existirem meses 40 e dias 40.

• Dimensão Farmácia - dimensão que contém os atributos da farmácia.

DIMENSÃO FARMÁCIA	EXEMPLO
PK Chave da Farmácia	24012
Nome da Farmácia	Farmácia Branco
Número de Telefone	266705362
Hora de Abertura	08:30:00
Hora de Fecho	19:30:00
E-mail	Não tem e-mail disponível
Website	farmaciabrancoevora.pt
Coordenadas Geográficas	38°33'58.5"N 7°54'10.5"W
Número da Porta	9
Rua	R. Dr. António José de Almeida
Código Postal	7005-297
Freguesia	Nossa Sra. da Saúde
Concelho	Évora
Distrito	Évora
Região	Alentejo

Tabela 7 Dimensão Farmácia

A chave primária da Dimensão Farmácia é a surrogate key "Chave da Farmácia". Essa surrogate key é apenas um número inteiro natural.



• **Dimensão Paciente** - dimensão que contém atributos específicos referentes ao paciente.

DIMENSÃO PACIENTE	EXEMPLO
PK Chave do Paciente	1
Nome do Paciente	João Pedro Figueira Galhardo Calhau
Cartão de Cidadão	14914698
Número de Documento	1 ZY7
Número de Identificação Fiscal	263009858
Número de Segurança Social	12037474855
Número de Utente de Saúde	494667293
Nome do Subsistema de Saúde	ADSE
Nacionalidade	Portuguesa
Data de Nascimento	11-07-1995
Idade	22
Sexo	Masculino
Altura	1,77 metros
Filiação	Franklim Manuel Nunes Condeço Galhardo Calhau * Florinda da Conceição da Silva Figueira
Número de Telefone	266731872
Número de Telemovel	967157369
E-mail	joaocalhau6@live.com.pt
Número da Porta	5
Rua	Rua das Melgas
Código Postal	7000-365
Freguesia	Malagueira
Concelho	Évora
Distrito	Évora
Região	Alentejo

Tabela 8 Dimensão Paciente

A chave primária da Dimensão Paciente é a surrogate key "Chave do Paciente". Essa surrogate key é apenas um número inteiro natural.



• **Dimensão Médico** - dimensão que contém atributos específicos referentes ao médico.

DIMENSÃO MÉDICO	EXEMPLO
PK Chave do Médico	5
Nome do Médico	Susana Abreu
Cartão de Cidadão do Médico	12345678
Número da Cédula	12345
Nacionalidade	Portuguesa
Data de Nascimento	20-10-1980
Idade do Médico	37
Sexo	Feminino
Número de Telefone	213456789
Número de Telemove	961234567
Especialidade	Ortopedia
E-mail	susana.abreu@gmail.com

Tabela 9 Dimensão Médico

A chave primária da Dimensão Médico é a surrogate key "Chave do Médico". Essa surrogate key é apenas um número inteiro natural.

• Dimensão Laboratório - dimensão que contém os atributos dos laboratórios.

DIMENSÃO LABORATÓRIO	EXEMPLO
PK Chave do Laboratório	6743
Nome do Laboratório	Meda Pharma
Número de Telefone	218420300
Número de Fax	218492042
E-mail	Geral@medapharma.pt
Coordenadas Geográficas	38°45'23.7"N 9°08'29.8"W
Número da Porta	13
Rua	R. Centro Cultural
Código Postal	1700-051
Freguesia	Alvalade
Concelho	Lisboa
Distrito	Lisboa
Pais	Portugal

Tabela 10 Dimensão Laboratório



A chave primária da Dimensão Laboratório é a surrogate key "Chave do Laboratório". Essa surrogate key é apenas um número inteiro natural. A Dimensão Laboratório é uma dimensão outrigger que se liga ao medicamento, devido a que o medicamento esteja associado ao laboratório que o criou.

• **Dimensão Medicamento** - dimensão que contém os atributos que definem um medicamento.

DIMENSÃO MEDICAMENTO	EXEMPLO
PK Chave do Medicamento	337293291323
Nome do Medicamento	Betadine
Número de Registo do Medicamento	8436014
Composição do Medicamento	Iodopovidona
Dosagem do Medicamento	40 mg/ml
Unidades do Medicamento	Frasco - 1 unidade(s) -
Forma Farmacêutica do Medicamento	Espuma cutânea
É Medicamento Genérico	Não é genérico
Nome do Laboratório do Medicamento	Meda Pharma - Produtos farmacêuticos, S.A.
Localização do Laboratório do Medicamento	1221156645
Data da Autorização do Medicamento	19751127
Estado da Autorização do Medicamento	Autorizado
Tamanho da Embalagem do Medicamento	500 ml
Código Nacional para a Prescrição Eletrónica de Medicamentos	50003666
Código Hospitalar Nacional do Medicamento	10034474
É Comercializado	É comercializado
Classificação Farmacoterapêutica - Grupo	13 Medicamentos usados em afecções cutâneas.
Classificação Farmacoterapêutica - Subgrupo	13.1 Antiinfecciosos de aplicação na pele
Classificação Farmacoterapêutica - Tipo	13.1.1 Anti-sépticos e desinfectantes.
Indicações do Medicamento	Desinfecção e higiene da pele e mucosas. Adjuvante no tratamento ou profilaxia da infecção em Micoses interdigitais (ex: pé de atleta) Adjuvante no tratamento da infecção na Dermatite seborreica. Profilaxia da Dermatite seborreica. Úlceras da perna: terapêutica adjuvante como desinfectante Em cirurgia: Lavagem das mãos antes e depois das intervenções cirúrgicas, como produto bactericida. Como sabão líquido para uso no consultório. Para os cuidados de limpeza e assepsia local do doente antes das intervenções cirúrgicas. Em obstetrícia-ginecologia: Assepsia ginecológica Desinfecção da vulva, períneo e face interna das coxas antes do parto.



Contra-Indicações do Medicamento	Este medicamento não deve ser utilizado nas seguintes situações: Antecedentes de alergia a qualquer dos constituintes, em particular à povidona. Não existem reacções cruzadas com os produtos de contraste iodados. As reacções de intolerância (reacções anafilactoides) aos produtos de contraste iodados ou a anafilaxia aos mariscos não constituem uma contra indicação à utilização de Betadine Espuma Cutânea. Desinfecção de material médico-cirúrgico. 1º trimeste da gravidez No recém-nascido até 1 mês De forma prolongada durante o 2º e 3º trimestre da gravidez. No aleitamento está contra indicado no caso de tratamento prolongado. Não utilizar este medicamento em associação com anti-sépticos derivados de mercúrio
Interações do Medicamento	Tendo em conta possíveis interacções (antagonismo, inactivação) a utilização em simultâneo ou sucessiva com outros anti-sépticos é de evitar. Possível interferência com exames à função da tiróide. Associações desaconselhadas: Anti-sépticos mercuriais: formação de um complexo cáustico no caso de utilização concomitante de anti-sépticos iodados e mercuriais.
Efeitos Secundários do Medicamento	No caso de administração repetida e prolongada, pode produzir-se uma sobredosagem de iodo susceptível de provocar disfunção da tiróide, nomeadamente no prematuro e em situações de grandes queimaduras. Excepcionalmente foram descritas reacções de hipersensibilidade: urticária, edema de Quincke, choque anafiláctico, reacção anafilactóide. No caso de aplicação em grandes superfícies e em grandes queimaduras, foram observados efeitos sistémicos: alterações da função renal com acidose metabólica, hipernatrémia. Podem produzir-se reacções cutâneas locais: dermatite cáustica e eczema de contacto.
Vias de Administração do Medicamento	Uso cutâneo
Início de Efeito do Medicamento	Instantaneo
Eliminação do Medicamento	A sua eliminação faz-se principalmente por via urinária. A polividona não dá lugar à passagem sistémica.
Descrição do Medicamento	Trata-se de um anti-séptico na forma de espuma cutânea
Prazo de Validade do Medicamento	20191010
Comparticipação do Serviço Nacional de Saúde (SNS)	0%
Comparticipação na Assistência na Doença aos Servidores Civis do Estado (ADSE)	0%
Comparticipação na Assistência na Doença aos Militares (ADME)	0%
Comparticipação da Segurança Social	0%
Preço do Medicamento	6.75€
FK Chave do Laboratório	6743
T-1-1-44 D:	mansão Madicamanto

Tabela 11 Dimensão Medicamento



A chave primária da Dimensão Medicamento é a surrogate key "Chave do Medicamento". Essa surrogate key é apenas um número inteiro natural. A Dimensão Medicamento tem uma chave estrangeira para a Dimensão Laboratório que é outrigger.

• **Dimensão Semestre** - dimensão que contém as características de um semestre.

DIMENSÃO SEMESTRE	EXEMPLO
PK Chave do Semestre	20171
Número do Ano	2017
Número do Semestre	1
Descrição do Semestre	1º Semestre de 2017
Data de Início do Semestre 01/01/2017	
Data de Final do Semestre	30/06/2017

Tabela 12 Dimensão Semestre

A chave primária da Dimensão Semestre é a surrogate key "Chave do Semestre". Essa surrogate key é especial, sendo definida como YYYYS.



Descrição da Tabela de Factos

A tabela de factos, na Tabela 13, representa o processo de negócio que pretendemos analisar. A métrica mais atómica que é possível analisar graças a esta tabela é a comparticipação do estado na venda de uma embalagem de medicamento.

Esta tabela é composta por 7 chaves primárias, sendo cada chave a chave primária de cada uma das dimensões que se ligam a esta tabela de factos. A tabela de factos contém ainda dois atributos, a referência da factura resultante da venda, e ainda o número de referência da receita, sendo esta uma dimensão degenerada.

Venda de Medicamentos		
Comparticipados Pelo Estado		
PK,FK1 PK,FK2 PK,FK3 PK,FK4 PK,FK5 PK,FK6 PK,FK7	Chave da Data Chave da Hora Chave da Localização Chave da Farmácia Chave do Médico Chave do Paciente Chave do Medicamento	
	Número de Referência da Factura Número de Referência da Receita (DD)	

Tabela 13 Tabela de Factos

Descrição da Tabela de Factos Sumarizada

Para a criação de uma tabela de factos sumarizada, Tabela 14, tivemos como base a tabela de factos da Tabela 13, sendo removido neste caso a Dimensão Data e Hora, e sido acrescentada uma dimensão Semestre, podendo verificar a comparticipação pelo estado das vendas de um medicamento por Semestre. É de notar que esta tabela tem ainda um atributo extra, que é o montante total pago no semestre.

Venda de Medicamentos Comparticipados Pelo Estado por		
Semestre		
PK,FK1 PK,FK2 PK,FK3 PK,FK4 PK,FK5 PK,FK6	Chave do Semestre Chave da Localização Chave da Farmácia Chave do Médico Chave do Paciente Chave do Medicamento	
	Montante total Pago	

Tabela 14 Tabela de Factos Sumarizada



Esquema em Estrela – Tabela de Factos

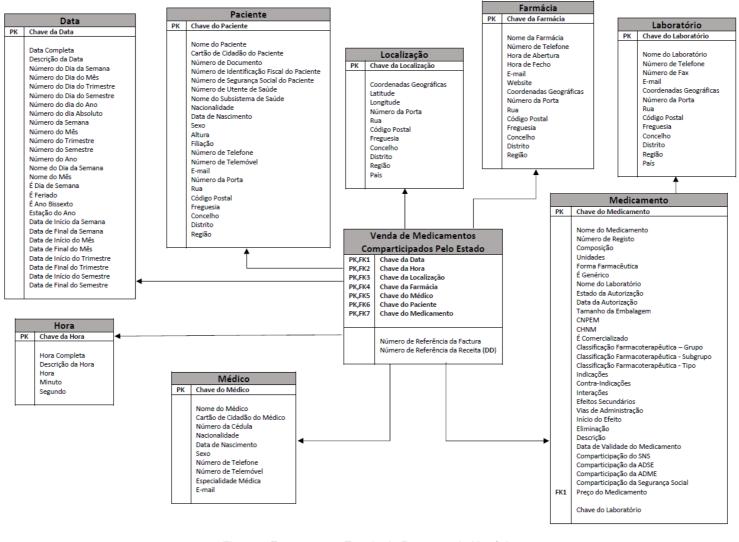


Figura 2 Esquema em Estrela do Processo de Negócio



Esquema em Estrela – Tabela de Factos Sumarizada

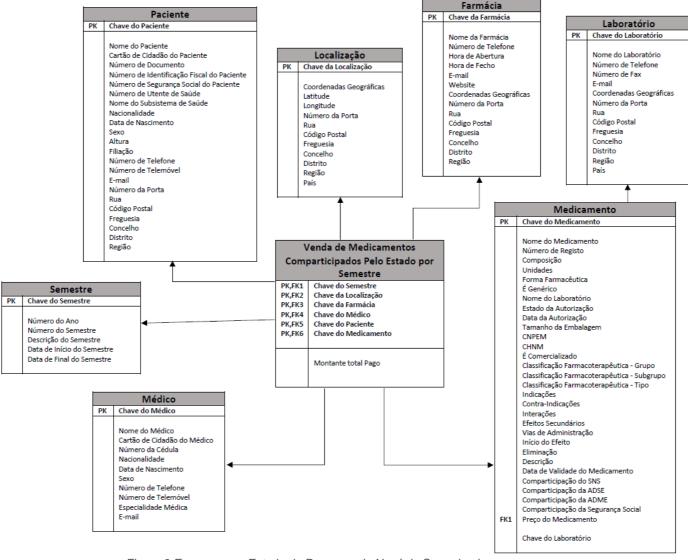


Figura 3 Esquema em Estrela do Processo de Negócio Sumarizado



Dimensões Role Playing

No desenvolvimento do nosso trabalho, não foi verificada a necessidade de serem criadas views de tabelas de dimensão, e como tal, não existem dimensões role playing.

Hierarquias

No trabalho verificam-se algumas hierarquias.

Um caso é na dimensão data, que se verifica a hierarquia da Figura 5, com o Ano, Semestre, Trimestre e Mês.

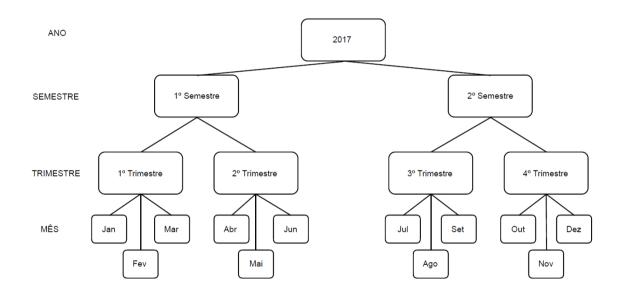


Figura 4 Hierarquia na Dimensão Data

Na dimensão Localização há uma hierarquia da localização, indo do mais abrangente, País, para o menos abrangente, Freguesia. A hierarquia é composta por País, Região, Distrito, Concelho e Freguesia. De notar ainda que por vezes esta hierarquia encontra-se também noutras dimensões, entre estas a dimensão Farmácia, Laboratório, Paciente.

Na dimensão Medicamento verifica-se a hierarquia da Classificação Farmacoterapêutica. Estar hierarquia é composta por Grupo Farmaceutico, o Subgrupo do medicamento e o Tipo de medicamento. Esta informação encontra-se disponível no Despacho n.º 4742/2014, de 21 de março da Infarmed, no seguinte link:

http://www.infarmed.pt/documents/15786/1072289/110-AB6 Desp 4742 2014 VF.pdf



Sistema de Metadados

O sistema de metadados foi realizado com base no template fornecido pelo professor. O sistema de metadados encontra-se no seguinte link: https://docs.google.com/spreadsheets/d/1sWh1G-4D3o4Svani-fOum-kG2jdEZntLAMZ3fhdeocA/edit?usp=sharing

Área de Estágio

Decidimos fazer a área de estágio deste trabalho manualmente, ou seja, decidimos fazer o procedimento de passar o código SQL para dimensões e tabela de factos à mão. Como tinhamos as tabelas criadas em PostgreSQL, e como, no Tableau, podemos connectar-nos diretamente ao servidor de PostgreSQL, então decidimos criar outro ficheiro SQL e mudar as tabelas para ficarem com o "aspecto" de uma dimensão. Para além disso, recorremos à ajuda de algumas funções que são executadas automaticamente quando um trigger é activado, para ajudar a adicionar campos que possam ser calculados a partir de um outro campo já existente, como por exemplo a idade do Paciente que pode ser calculada a partir da data de nascimento com a query "SELECT (current_date – data_nasc)/365 from Paciente".

As ferramentas utlizadas neste caso foram Tableau, PostgreSQL, PL/Pgsql, Sublime Text (onde fizemos o código SQL).



Referências

Caldeira, C. (2012). *Data Warehousing. Conceitos e modelos*. 2ª Edição. Edições Sílabo. ISBN: 978-972-618-696-0