

# Business Intelligence I

2017/2018

## PÓS-GRADUAÇÃO EM SMART CITIES

### Projecto:

Solução de Business Intelligence para empresa de importação e venda de automóveis, localizada em Angola – Cars&Co.

### Turma prática: P5

### Grupo: 11

### Elementos do grupo:

M20170202 – Jorge Santos

M20170532 – Carlos Lopes

M20170732 – Mafalda Monteiro

Janeiro 2018

## ÍNDICE

1. INTRODUÇÃO .....	4
1.1. Caracterização da empresa.....	4
1.2. Caracterização do problema .....	5
1.3. Benefícios de uma solução Business Intelligence para a empresa .....	5
1.4. Dados estatísticos .....	6
1.4.1. Performance dos produtos e vendas.....	6
1.4.2. Performance dos stands .....	10
1.4.3. Carteira de clientes.....	11
2. METODOLOGIA DE ABORDAGEM DO TRABALHO DE B.I. ....	12
3. DATA SOURCE.....	14
3.1. Data Source existente .....	14
3.2. Complemento da Data Source .....	14
3.3. MODELO RELACIONAL DA DATA BASE.....	15
3.3.1. Identificação e relação entre as entidades.....	15
3.3.2. Esquema relacional.....	16
4. A ARQUITECTURA DA DATA WAREHOUSE.....	18
4.1. Modelo dimensional .....	18
4.1.1. Abordagem Kimball .....	19
4.1.2. Abordagem Moody e Kortink .....	19
4.2. GALAXY SCHEMA.....	20
4.2.1. Tabelas de factos (Fact table).....	20
4.2.2. Tabela de dimensões (Dimensional table) .....	21
4.2.3. Esquema relacional.....	21

4.3. Problemas encontrados .....	23
5. RESULTADOS OBTIDOS.....	24
5.1. Staging Area .....	24
5.2. Data Warehouse .....	27

## 1. INTRODUÇÃO

### 1.1. CARACTERIZAÇÃO DA EMPRESA

A empresa **Cars&Co Lda** está presente no mercado Angolano desde 2000. Dedicar-se actualmente à importação e venda a retalho de automóveis ligeiros e pesados de passageiros e de mercadorias e camiões.

O negócio cresceu na década de 2000 o que lhe permitiu abrir diversos pontos de venda, sobretudo em Luanda e no Lobito, junto dos principais portos marítimos. As viaturas são importadas directamente das fábricas e transportadas para Angola por via marítima.

A comercialização é feita directamente aos clientes - particulares e empresas - nos centros-automóvel. Os compradores deslocam-se de diversos pontos do País até esses entrepostos localizados nas províncias de Luanda, Benguela, Huambo e Cabinda, para fazerem as suas aquisições. É nesses locais que a empresa faz exposição de mercadoria de serviço. São os serviços centrais que contactam o fornecedor e que fazem a *stockagem* da mercadoria. As aquisições são realizadas, em princípio, mediante o sucesso das vendas.



Fig. 1 Esquema da relação comercial da empresa com fornecedores e clientes.

## 1.2. CARACTERIZAÇÃO DO PROBLEMA

Devido à crise financeira que se instalou em Angola por causa da desvalorização do petróleo e que catapultou a crise de divisas no País, a empresa atravessa alguns problemas, como sejam:

- Preço de referência das importações fixado em Euros e contabilizado em Kwanzas nas vendas;
- Riscos cambiais decorrentes da desvalorização do Kwanza em relação ao Euro;
- Custos de importação, incluindo taxas alfandegárias, e de transporte, elevados;
- Outros riscos, nomeadamente, a inflação interna;
- Pouca rotatividade de *stock* de alguns artigos;
- Grande redução nas vendas.

## 1.3. BENEFÍCIOS DE UMA SOLUÇÃO BUSINESS INTELLIGENCE PARA A EMPRESA

A empresa Cars&Co tem registos da sua actividade desde 2004. No período 2010-2014, o negócio estava em crescimento. Aos *stands* iniciais em Luanda e Benguela (Lobito) foram adicionados novos stands no Huambo e em Cabinda. A sua estratégia de negócio foi baseada no crescimento das vendas.

Contudo, neste momento a empresa não dispõe de uma base de comparação da performance das vendas ao longo do tempo e que reflecta, também, os problemas até agora detectados. Mas, sobretudo, que lhe permita orientar a sua estratégia de negócio. Pretende responder, nomeadamente, a:

- No âmbito de Marketing e Comercial:
  - Qual a quantidade e qual o valor e margem (em Kwanzas) das vendas mensais, por segmento de mercado, por marca/modelo e por *stand*, nos últimos 5 anos?
  - Qual foi a variação da sazonalidade das vendas, por segmento de mercado, por marca/modelo e por *stand*, nos últimos 5 anos?
  - Quais são os segmentos de mercado alvo, a nível empresarial e particular, e como são caracterizados, em particular quanto ao perfil socio-económico, poder de compra e preferências por tipo de veículos?
  - Qual a gama de produtos que será melhor do ponto de vista da maximização de lucros?
  - Como ajustar a sua representação a nível de *stands* aos vários segmentos de mercado alvo e à sua diversidade entre províncias?
  - Quais os *stands* mais eficientes? Deve concentrar pontos de venda nesses locais ou procurar outros pontos estratégicos?

– No âmbito de Logística:

- Qual a evolução de *stock* mensal, por segmento/ marca/ modelo, por stand, nos últimos 5 anos?
- Como melhorar a rotatividade de *stocks* em geral e por stand?
- Qual o prazo médio de rotação de *stock*, por segmento/ marca/ modelo, por stand e por Província, nos últimos 5 anos?

– No âmbito Financeiro:

- De que forma a crise de divisas afectou o negócio? Deve manter um nível de *stock* mínimo, só importar após encomenda sinalizada?

Para obviar o problema, a empresa Cars&Co recorreu a um serviço de Business Intelligence (B.I.) para organizar a sua informação de forma coerente e verdadeira, numa só versão uniformizada e que, assim, permita tomar decisões estratégicas de curto e de longo prazos. Para poder reajustar mais rapidamente a sua estratégia de negócio precisa, também, que o intervalo de tempo habitualmente longo entre o fecho semestral de vendas e a emissão do respectivo relatório seja muito reduzido.

O serviço contratado permitir-lhe á criar uma estrutura de *Data Warehouse*, segundo um modelo de orientação por objectos, integrada, com incorporação de informação no tempo e não-volátil, que é alimentada por diversas fontes de dados da empresa os quais por intermédio de *Cubos* e *Data Marts* são convertidos em *reports* de fácil leitura com vista ao apoio à decisão.

A empresa optou por criar essa estrutura no sistema SQL Server 2017 de modo a reduzir os custos de consultoria e por este programa conter as ferramentas necessárias à organização e extracção de dados e, ainda, à exportação e produção de relatórios de consulta simples.

## 1.4. DADOS ESTATÍSTICOS

A nível de marcas a empresa iniciou actividade com a Nissan e Nissan Diesel. Em 2005 surgiram os primeiros Renault mas só a partir de 2006 é que começou a comercializar vários modelos dessa marca. Em 2008 adicionou ao seu catálogo de marcas a Renault Trucks.

### 1.4.1. PERFORMANCE DOS PRODUTOS E VENDAS

Entre 2004 e 2014 a empresa apresentou um crescimento do número de viaturas vendidas, apesar de ter tido uma quebra de vendas em 2009 e 2010. O ano durante o qual atingiu o máximo de vendas com 3796 viaturas foi 2014. Na sequencia da crise económica que afecta o país desde 2015 a vendas de todas

as marcas reduziu de forma muito acentuada, entre 2014 e 2016 a quebra no número de vendas de veículos foi de 90% (Fig. 2).

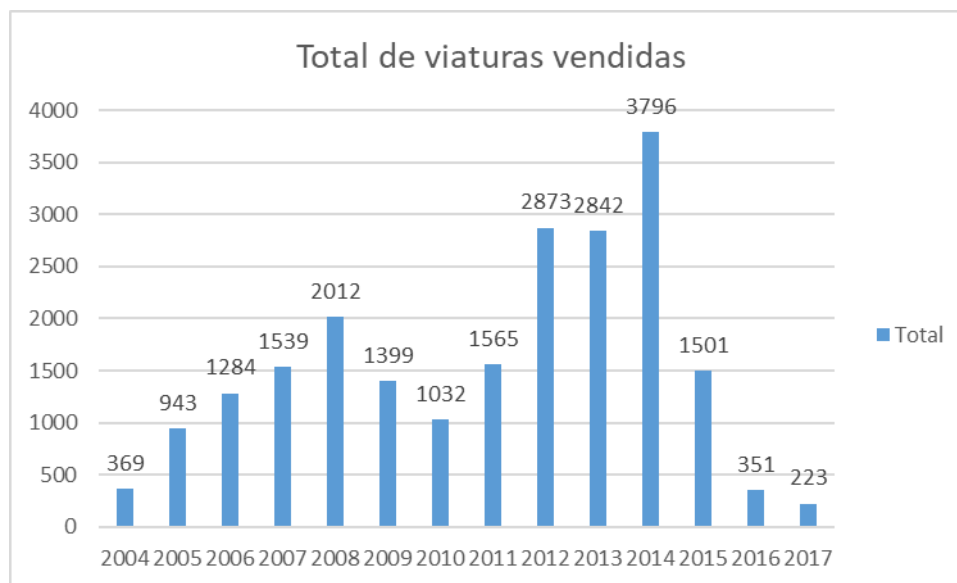


Fig. 2 Total de viaturas vendidas anualmente, entre 2004 e Setembro de 2017.

Até 2011 a marca que liderava o número de vendas era a Nissan. Em 2012 as vendas foram repartidas com a Renault que liderou o número de chassis vendidos a partir de 2013, ano que se deixou comercializar a marca NISSAN DIESEL passando a ser integrada na NISSAN. A partir de 2015 em consonância com a crise económica a Renault Trucks registou um volume de chassis vendidos muito residual que traduz a ausência de investimento das empresas em meios de produção (Fig. 3).

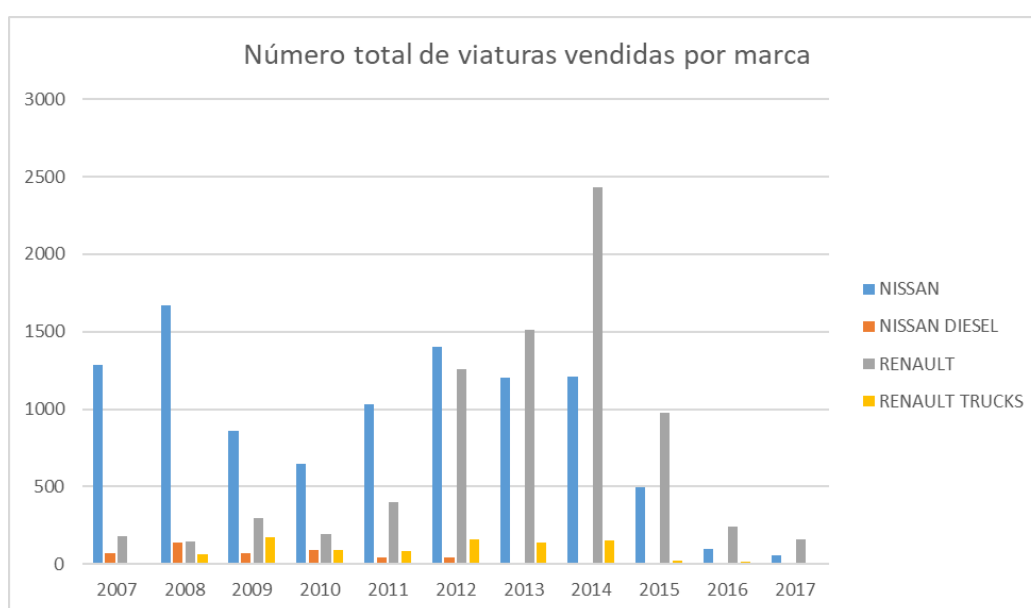


Fig. 3 Número total de viaturas vendidas por marca comercializada, entre 2007 e Setembro de 2017.

O modelo *best seller* dos últimos 5 anos é o Duster da Renault, seguido da Hardbody da Nissan e do Renault Sandero (Fig. 4).

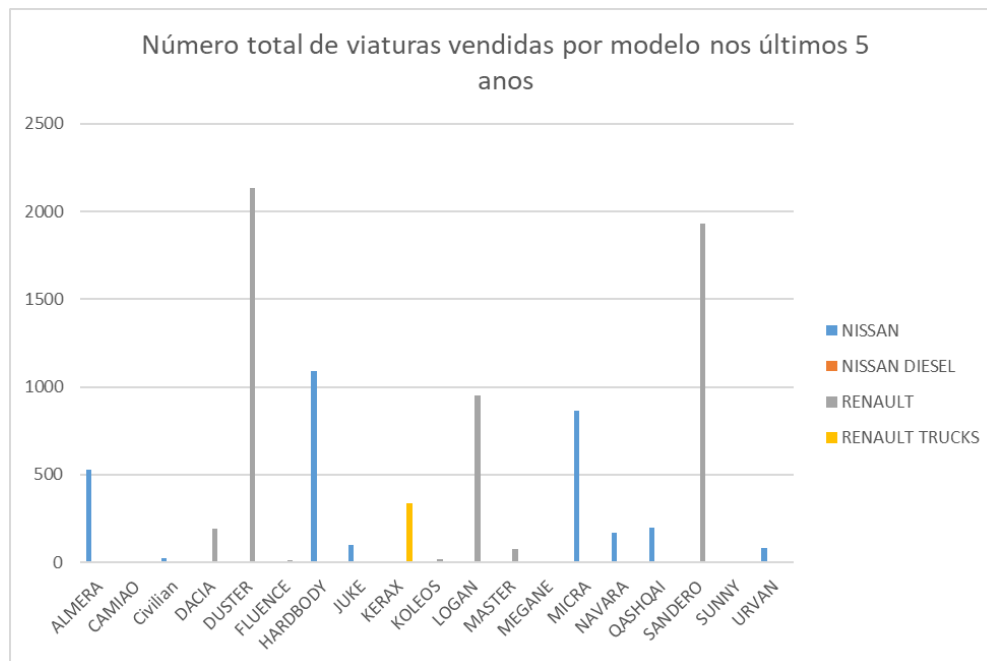


Fig. 4 Número total de viaturas vendidas por modelo comercializado, entre 2013 e Setembro de 2017.

O volume de vendas da empresa teve uma quebra de 88% em apenas 2 anos entre 2014 e 2016 (Fig. 5).

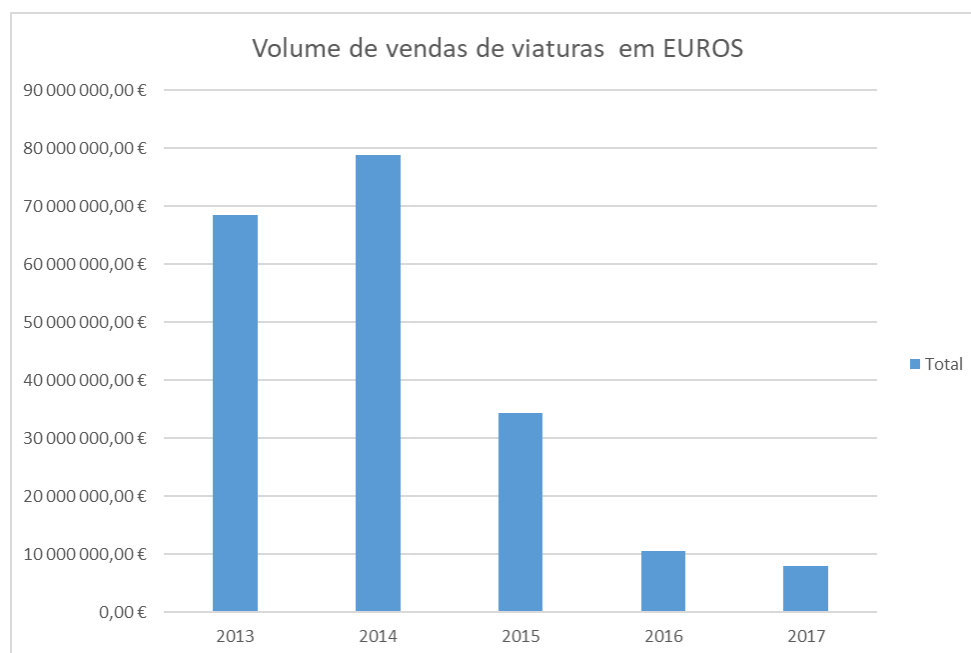


Fig. 5 Volume de vendas de viaturas entre 2013 e Setembro de 2017.



O modelo que mais contribui para o volume de negócios em 2013 e 2014 são os Renault Kerax contribuindo para isso o seu elevado valor unitário, segue-se e depois Renault Duster e o Renault Sandero. A partir de 2015 o modelo com maior volume de vendas em tempo de crise é o Renault Duster (Fig. 6).

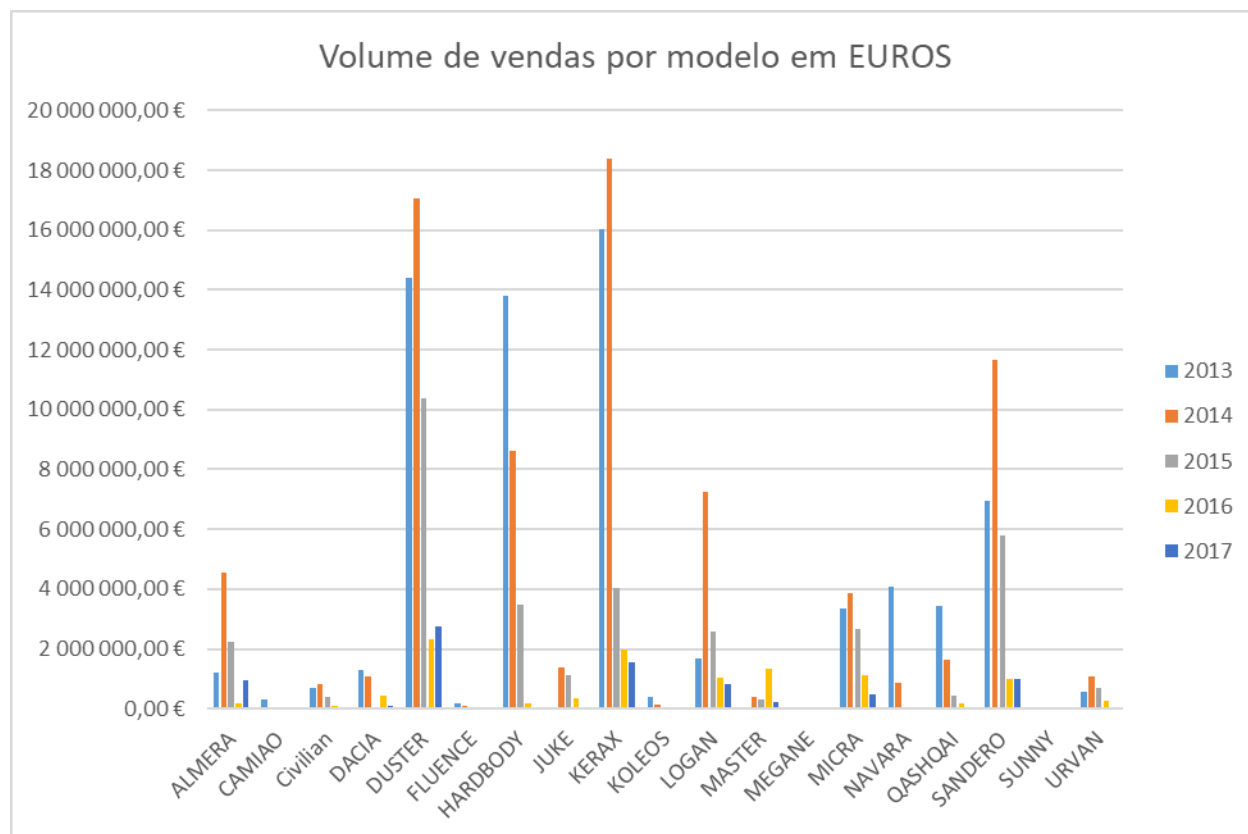


Fig. 6 Volume de vendas por modelo de 2013 até Setembro de 2017

Em relação aos stocks a 30/09/2017 o modelo que tem mais unidades disponíveis em armazém é o Renault Duster com 74 viaturas seguindo-se o Renault Sandero com 58 viaturas, sendo estas as que menos tempo estão em armazém o Duster está em média á 138 dias em armazém e o Sandero á 315 dias, estes valores estão de acordo com o modelos mais vendidos, no entanto face á media das unidades vendidas em 2016 e 2017 tem *stocks* para os próximos dois anos, o que poderá ser um pouco em excesso face aos riscos de alteração do modelo, roubo, deterioração da viatura, além do capital empatado (Fig. 7).

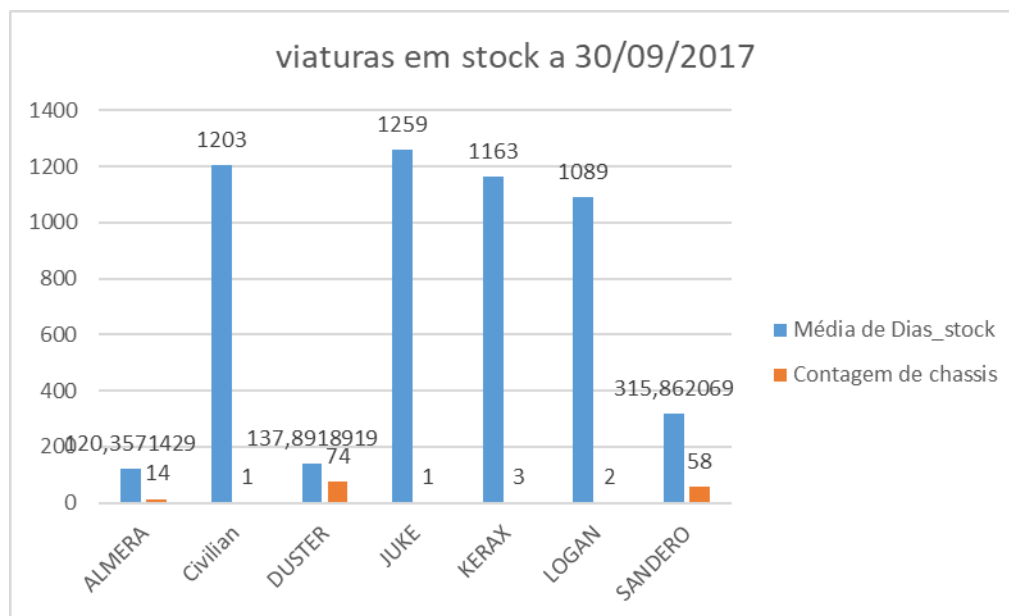


Fig. 7 Viaturas em armazém a 30/09/2017.

#### 1.4.2. PERFORMANCE DOS STANDS

O stand de Talatona onde está localizada a sede e a armazenagem de viaturas é o que representa maior volume de vendas (24,55%), seguindo-se Benguela 7,76%, São Pedro (7,1%) e Cacuaco (2,26%) é o stand com menor volume de vendas (Fig. 8).

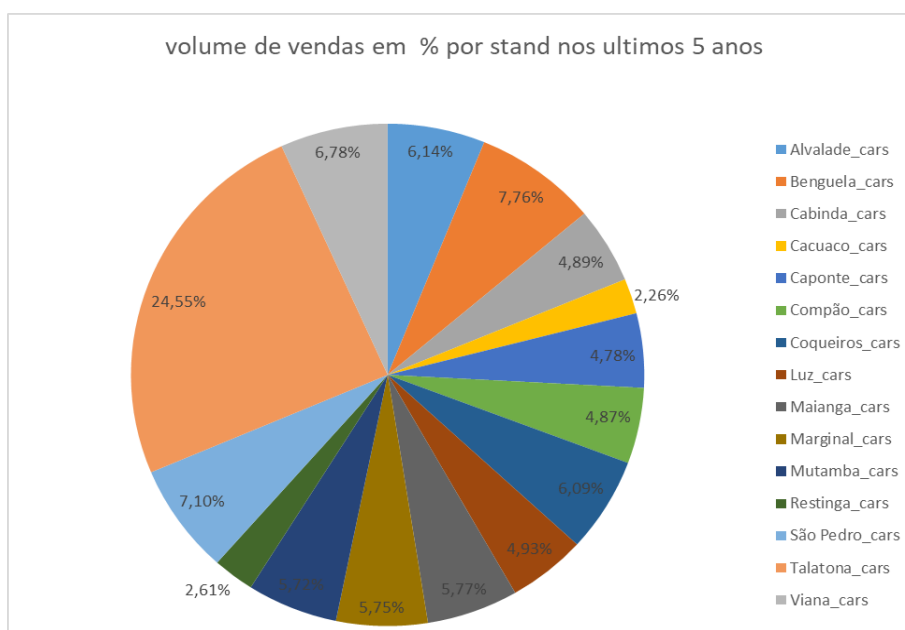


Fig. 8 Volume de vendas em % por stand de 2013 a Setembro 2017.

### 1.4.3. CARTEIRA DE CLIENTES

O mercado empresarial representa 70% da quota de mercado. Dada a sua importância, apenas o mercado empresarial foi alvo de análise. O top dos 10 principais clientes empresariais é liderado por BCI do Zaire seguindo-se BFA Kwanza Sul (Fig. 9).

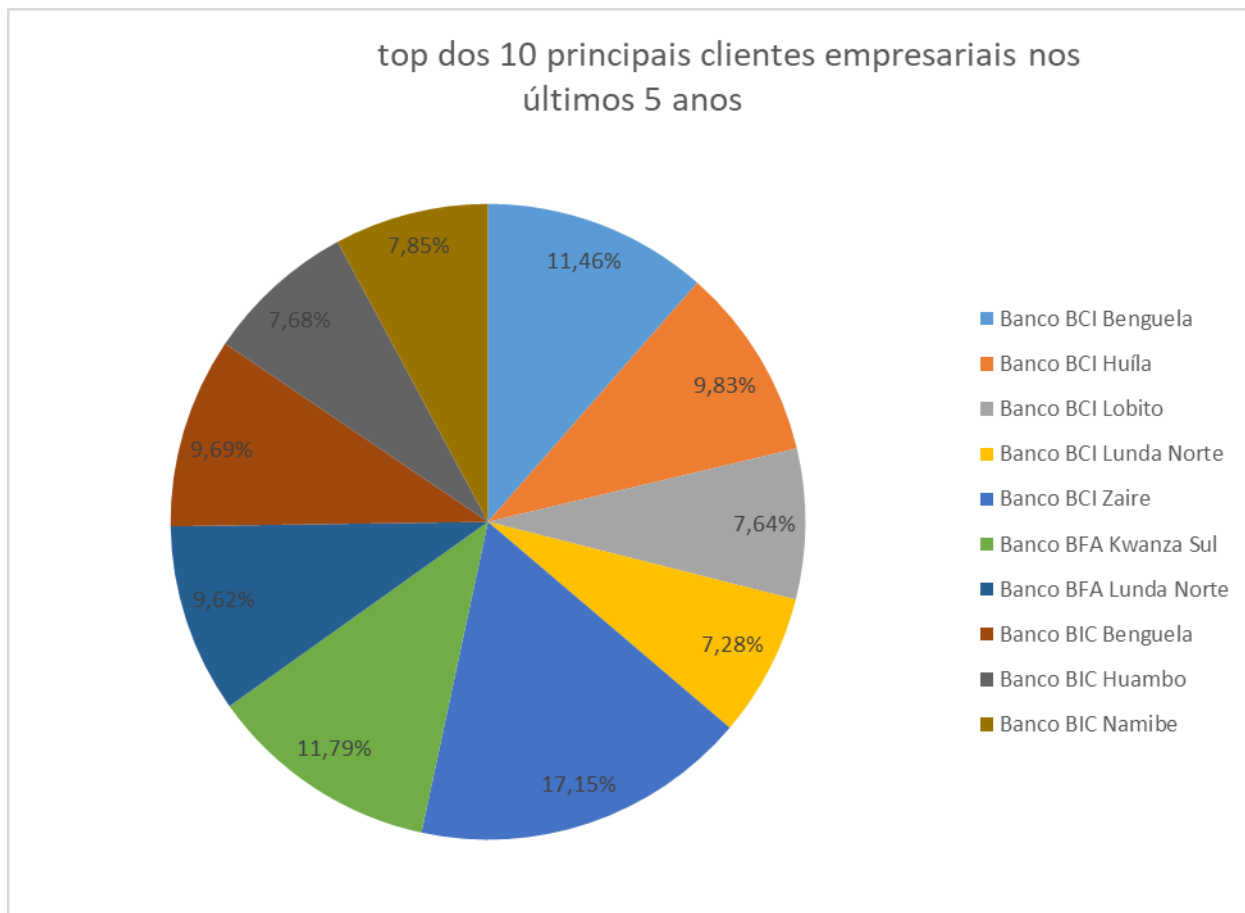


Fig. 9 Top dos 10 principais clientes empresariais de 2013 a Setembro 2017.

## 2. METODOLOGIA DE ABORDAGEM DO TRABALHO DE B.I.

---

A metodologia adoptada na abordagem do problema compreendeu os seguintes passos:

1. Escolha de uma empresa fictícia baseada em dados reais, na medida do possível;
2. Análise crítica dos dados existentes (originais):
  - Quanto à sua viabilidade em termos de informação;
  - Quanto à coerência de dados;
  - Quanto às lacunas a preencher de modo a conter informação analisável do ponto de vista estratégico e do processo de decisão;
  - Quanto à possibilidade de se converter em *Data Source*;
3. Identificação de elementos de pesquisa;
4. Complemento da *Data Source*:
  - Pesquisa sobre elementos estatísticos que enquadrem a actividade da empresa;
  - Pesquisa sobre produtos de comercialização e diversificação dos dados originais;
  - Análise de sensibilidade dos dados “acrescentados” quanto à sua coerência em relação ao historial da empresa e do comportamento conhecido do sector de mercado automóvel;
  - Compilação de registos cambiais no período em que se tem registos;
5. Criação do Modelo Relacional da *Data Base* que compõe a *Data Source*:
  - Caracterização das entidades e ligações;
  - Definição do esquema relacional do modelo;
  - Criação do modelo físico da *Data Base* em SQL Server;
6. Definição da estrutura da *Staging Area* e da *Data Warehouse* (DW):
  - Discussão do esquema que melhor representa a actividade da empresa;
  - Definição das tabelas dimensionais e factuais;
  - Criação do modelo físico da *Staging Area* (Cars\_Co\_SA) e da DW (Cars\_Co\_DW) em SQL Server, por supressão de campos da Cars\_Co\_SA referentes a *business keys*;
7. Criação do processo ETL para carregamento da *Staging Area*:
  - Criação de variáveis para controlo do processo de extracção e de transformação e controlo dos dados carregados - ETL\_name, row\_count, UltimaDataCompra, UltimaDataVenda;
  - Criação de parâmetro para simplificação da gestão de *flat files* (InputDataFile):

- Criação de parâmetro de escolha para *incremental loading* (ParamClearDimensionTables) e execução do processo ETL - Apagar Dim Tables?
  - = Yes – elimina registos das tabelas de Cars\_Co\_SA
  - = No – mantêm registos nas tabelas Dim e Fact e incrementa registos correspondentes a alterações por via de SCD ou de introdução de registos com datas mais recentes que as últimas existentes na Cars\_Co\_SA;
- Extracção e transformação de dados provenientes da DataSource SQL Cars\_Co e de *flat files* Data\_csv.csv e cambios.csv
- Carregamento da Cars\_Co\_SA pelo processo ETL (Extract, Transform and Load process).
- A extracção de dados é feita por inner join de tabelas da DataSource SQL Cars\_Co, *merge join* com *flat files*, *split* de resultados, etc;
- Existem 2 Slowly Changing Dimensions (SCD) tipo 2 aplicadas às tabelas Dim\_Fornecedores (com variação de cidade) e Dim\_Produto (com variação de versao).
- A transformação de dados inclui a conversão de dados e o cálculo de medidas.

#### 8. Carregamento da DW:

- Definição do processo de Extracção ETL da Cars\_Co\_SA para a Cars\_Co\_DW:
  - Extracção dos dados da *Staging Area*;
  - Existem SCD tipo 1 aplicadas às tabelas Dim\_Fornecedores (com variação de cidade) e Dim\_Produto (com variação de versao) e às Fact Compras e Fact Vendas;
  - As restantes tabelas Dim apresentam SCD tipo 0;
- Observações ao processo ETL, problemas e soluções adoptadas;

#### 9. Realização de processos OLAP e análise crítica de resultados:

- Verificação da coerência dos resultados;
- Eventuais ajustes necessários na *Data Source*, correcção dos dados adicionados aos dados originais, para afeiçoamento a um problema “real”;

#### 10. Apresentação dos dados finais consolidados e dos resultados.

Foram utilizadas as seguintes ferramentas informáticas:

- Microsoft EXCEL para tratamento da *Data Source*;
- SQL Server Data Tools 2015 (visual studio) para desenho da BD e da DW e definição do processo ETL;

- Microsoft SQL Server Management Studio 17 para gestão da BD, da SA e da DW.

### 3. DATA SOURCE

---

#### 3.1. DATA SOURCE EXISTENTE

O conjunto de dados obtidos para a criação da *Data Source* consiste em diversos campos estruturados em **formato xls** com mais de 32 900 registos. Contudo diversos registos têm informação omissa, mal introduzida, com caracteres inapropriados e informação espúria.

O histórico de dados obtidos pelas datas de encomenda ao fornecedor e de factura ao cliente abrange, principalmente, o intervalo de tempo de 2004 até 2017, com maior concentração de vendas entre 2010 e 2016, sendo que em 2017 ocorreu queda abrupta das vendas devido à crise.

Face às omissões, incoerência de diversos registos e à necessidade de saneamento de valores adulterados optou-se por restringir o intervalo da DW ao período 2010 a 2017, em que se verifica maior amplitude na variação das vendas, e ao qual correspondem cerca de 20.000 registos. Considerou-se que este período é suficiente para analisar a tendência de evolução do mercado.

Os campos da *Data Source* são referentes a stockagem, compra e venda de veículos e os valores monetários estão expressos em Kwanzas (AOA). Por este facto, os preços de compra de veículos iguais variam diariamente consoante o mercado internacional e as alterações cambiais. Assim, o registo de compra (importação) de cada veículo na *Data Source* tem, efectivamente, valores únicos.

Cada compra e cada venda correspondem à transacção de um veículo, isto é, as facturas têm todas quantidade =1.

#### 3.2. COMPLEMENTO DA DATA SOURCE

De modo a criar mais fontes de análise de dados, foi corrigido e acrescentado à tabela inicial:

- Detalhe quanto ao produto-veículo – modelo, versão, motorização, tracção, segmento de mercado;
- Informação quanto a fornecedores – nomes e locais mundiais de fabrico das marcas transaccionadas pela empresa;
- Informação quanto a vendedores (stands) – nomes e locais em Angola (cidade e província), atendendo à realidade Angolana explanada nas “páginas douradas”;
- Informação quanto a compradores – nomes e locais em Angola (cidade e província), correspondendo a uma segmentação de 70-30 (arbitrado) entre clientes do tipo empresa e particular;

- Obtenção de informação cambial mensal no período de registos da empresa, sendo essa a granularidade assumida pelo seu departamento de gestão financeira.

A partir destes dados foi estruturada a *Data Base* da empresa Cars&Co e foram definidas duas tabelas complementares de dados (*flat file* em formato xls) referentes às taxas de câmbio e às datas. Estes são os elementos da *Data Source* e que são provenientes do sistema OLTP da empresa.

### 3.3. MODELO RELACIONAL DA DATA BASE

#### 3.3.1. IDENTIFICAÇÃO E RELAÇÃO ENTRE AS ENTIDADES

Foram estruturadas as seguintes entidades e ligações:

- Tabela Produto:

- Nesta tabela estão os campos de *número de chassi* e de *status*.
- Relaciona-se 1:1 (*one-to-one*) com as tabelas de Compras / Vendas / Armazém, uma vez que os valores monetários (em AOA) e as datas de entrada/saída em armazém são únicas para cada chassi de veículo;
- Relaciona-se n-1 (*many-to-one*) com a tabela Versão (que traduz o tipo de veículo), pois há vários registos da mesma versão na tabela Produtos;
- Relaciona-se n-1 com a tabela Fornecedor, pois há vários registos de um fornecedor na tabela Produto.
- Relaciona-se n-1 com a tabela Cliente, pois apesar de um Cliente poder adquirir mais que um tipo de veículo e um tipo de veículo poder ser vendido a diferentes clientes, uma vez que o número de chassi é único a relação de Produto com esta entidade é n-1;
- Outras relações hierárquicas da tabela Produto são:
  - Tabela Versão - relaciona n-1 com Modelo / Segmento;
  - Tabela Modelo:
    - relaciona-se n-1 com a tabela Marca;
    - relaciona-se n-1 com a tabela Fornecedor. Há vários registos de um fornecedor na tabela Modelo. O inverso neste caso não existe, pois no complemento à *Data Source* optou-se por um fornecedor único para cada modelo de veículo. Admitiu-se que a empresa fez um estudo

de mercado para seleccionar o fornecedor mais económico para cada modelo e que não houve alterações no período de registos;

- Tabela Fornecedor – relaciona-se n-1 com a tabela Cidade\_Fornecedor e esta relaciona-se n-1 com a tabela Pais\_Fornecedor;
- Tabela Compras:
  - Nesta tabela estão os campos referentes à factura da compra: nº factura, valores (preço base, taxas alfandegárias, custos de transporte e diversos, em AOA) e a data de compra (quando é feita a encomenda do produto ao fornecedor);
  - Os seus valores são únicos para cada código de chassi pelo que se relaciona 1-1 com Produto;
- Tabela Vendas:
  - Nesta tabela estão os campos referentes à factura de venda: nº factura, preço final, data da venda ao Cliente;
  - Relaciona-se n-1 com a tabela Cliente, pois um Cliente pode adquirir mais do que um veículo ao longo do tempo de registos;
  - Relaciona-se n-1 com a tabela Stand, pois um Stand pode vender mais do que uma vez ao longo do tempo de registos;
  - Outras relações hierárquicas da tabela Vendas são:
    - Tabela Cliente:
      - relaciona-se n-1 com a tabela Cidade\_Cliente e esta por n-1 com a tabela Provincia\_Cliente;
      - relaciona-se n-1 com a tabela Tipo\_Cliente
    - Tabela Stand, relaciona-se n-1 com a tabela Cidade\_Stand e esta por n-1 com a tabela Provincia\_stand;

Optou-se por diferenciar as tabelas Cidade e Província entre Clientes e Stands, apesar de conterem os mesmos atributos, para evitar armazenar-se informação nula nos casos em que uma dada cidade e província não fossem comuns a Clientes e a Stands.

### 3.3.2. ESQUEMA RELACIONAL

O esquema da *Data Base* está representado na imagem seguinte.



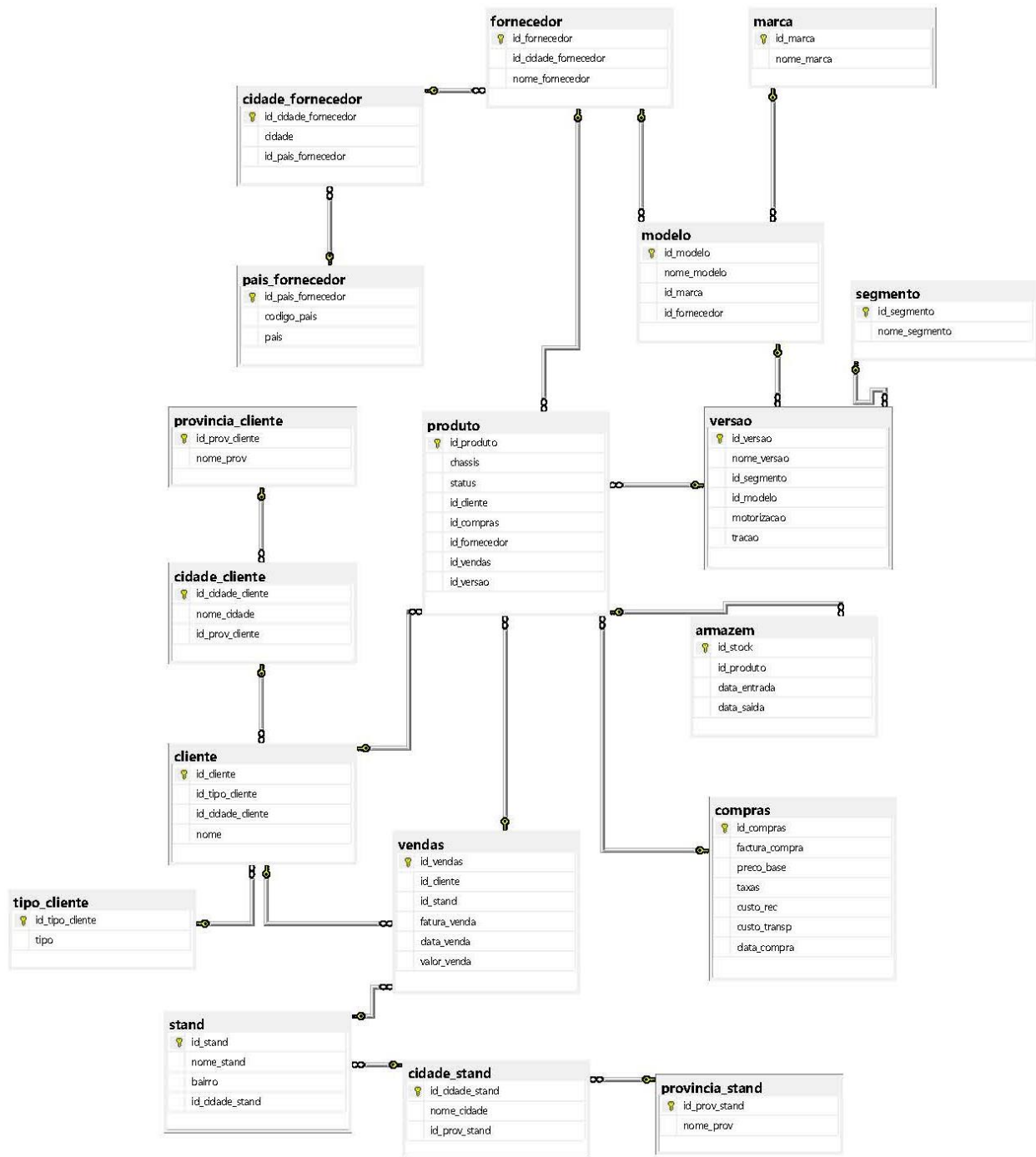


Fig. 10 – Esquema da Data Base Cars\_Co preparado para SQL Server.

## 4. A ARQUITECTURA DA DATA WAREHOUSE

A *Data Warehouse* (DW) é actualizada cada vez que é realizado o processo ETL (ETL – *Extract, Transform & Load*). Este processo consiste na extracção de dados relevantes da base de dados da empresa para um armazém temporário, onde é realizada a sua transformação, e em seguida é efectuado o seu carregamento para a DW.

Os dados relevantes, aqueles que são necessários e importantes à análise da performance da empresa Cars&Co no processo B.I., são provenientes de diferentes fontes de informação da empresa. Estas fontes pertencem à base OLTP do seu sistema (Online Transaction Processing Systems). Neste caso, são a *Data Base SQL* e as tabelas de câmbio e de datas no formato csv. A arquitectura do sistema está esquematizada na imagem seguinte.

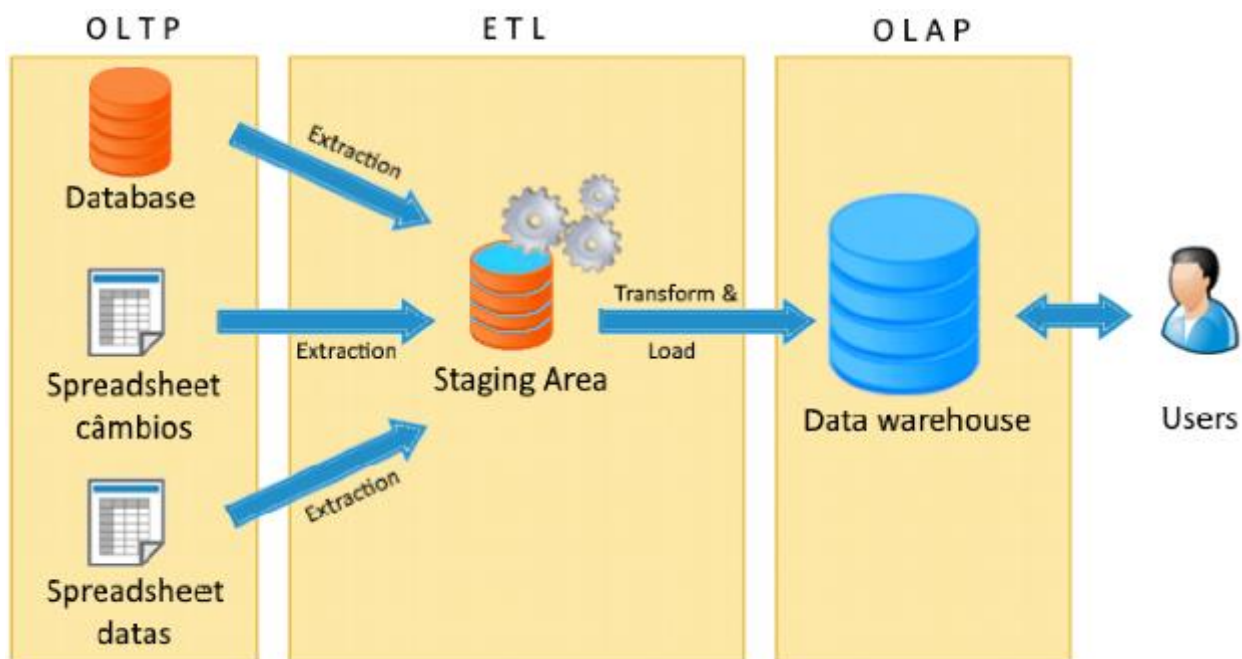


Fig. 11 – Arquitectura da Data Warehouse.

### 4.1. MODELO DIMENSIONAL

O modelo dimensional consiste na desnormalização do modelo relacional em diversas tabelas Dimensionais, de forma a otimizar o tempo de pesquisa à DW em detrimento da maior ocupação de espaço de memória, uma vez que a informação é repetida entre registos.

As tabelas Dimensionais agregam informação para cada dimensão (assunto) e contêm níveis de detalhe dessa dimensão. Os seus atributos podem conter relações hierárquicas.

As tabelas de Factos traduzem transacções ou eventos da empresa e contêm as medidas em observação para apoio à tomada de decisão. A sua granularidade (nível de detalhe) relaciona-se com a actividade e necessidades de monitorização da empresa.

#### **4.1.1. ABORGEM KIMBALL**

Por aplicação da aborgem Kimball, foram identificados os processos transaccionais, a granularidade necessária, as dimensões e as medidas.

Neste caso os assuntos organizados em dimensões são: Produto, Data, Fornecedor, Stand (vendedor) e Cliente.

As tabelas de Factos que traduzem os eventos são: Compras e Vendas. Contêm as medidas a monitorizar, que neste caso são os valores de compra e de venda cotados em AOA e em EUR. A granularidade depende das necessidades da empresa, que neste caso são as compras e vendas mensais.

#### **4.1.2. ABORDAGEM MOODY E KORTINK**

Esta abordagem consiste na classificação das entidades e na identificação das hierarquias presentes no modelo relacional.

##### Entidades

De forma a criar o modelo dimensional a partir do modelo relacional houve que categorizar as diferentes entidades: Transições / Componentes / Classificação

##### – Transições (*Transactions*)

Traduzem ocorrências no negócio da empresa Cars&Co como sejam as ordens de compra e de venda de produtos.

##### – Componentes (*Components*)

Relaciona-se com a entidade de transacção na relação 1-n (*one-to-many*). Refere-se aos detalhes da transacção, como sejam o detalhe do produto, stands e clientes.

##### – Classificação (*Classification*)

Traduzem hierarquias embebidas no modelo relacionadas por uma corrente relacional em 1-n, como sejam os fornecedores, cidades e países, as cidades e províncias, as versões os modelos a as marcas, etc.

### Hierarquias

Atendendo ao esquema representado na Fig. 10 foram identificadas as seguintes entidades e hierarquias.

Transições:	Componentes:	Classificações:		
Compras	Produto	Fornecedor	Cidade	País
Compras	Produto	Versao	Modelo	Marca
Compras	Produto	Versao	Segmento	
Compras	Produto	Armazém		
Vendas	Cliente	Cidade	Província	
Vendas	Cliente	Tipo cliente		
Vendas	Stand	Cidade	Província	

## 4.2. GALAXY SCHEMA

O modelo encontrado para descrever a actividade da empresa Cars&Co é do tipo *Galaxy Schema* derivado de *Star Schema*, pois contém duas tabelas de Factos – Compras e Vendas - às quais estão ligadas tabelas Dimensionais, sendo que partilham 2 dessas tabelas Dimensionais – Produto e Data.

Observando cada tabela de Factos isoladamente, esta relaciona-se em forma de *Star Schema* com as tabelas de Dimensão contíguas. Foi escolhido este modelo por ser simples na definição e acelerar o processo de pesquisas. Mas contém redundância de dados face a uma abordagem por *Snowflake Schema*.

O esquema que traduz a SA da empresa Cars&Co, a utilizar no processo ETL para a extracção de dados para a *Staging Area*, está representado na Fig. 12.

O esquema que traduz a DW da empresa Cars&Co, a utilizar no processo ETL para a extracção de dados da *Staging Area* para a *Data Warehouse*, está representado na Fig. 13.

### 4.2.1. TABELAS DE FACTOS (FACT TABLE)

As relações entre as entidades do modelo dimensional e do sistema OLTP está resumida nas tabelas seguintes, onde se apresenta o cálculo das métricas a usar na resolução das perguntas do negócio..

#### Tabela Fact\_Compra

Measures	OLTP Fields	Data Type	Formula
preco_base_AOA	preco_base	money	preco_base_AOA
preco_base_EUR	preco_base, taxa cambio (xls)	money	preco_base_AOA / taxa_cambio
custos_AOA	custo_rec, custo_transp	money	SOMA(custo_rec, custo_transp)_AOA
custos_EUR	custo_rec, custo_transp, taxa cambio (csv)	money	SOMA(custo_rec, custo_transp)_AOA / taxa_cambio
taxas_EUR	taxas, taxa cambio (csv)	money	Taxas_AOA / taxa_cambio
preco_total_compra_AOA	preco_base, taxas, custo_rec, custo_transp, taxa cambio (csv)	money	SOMA(preco_base, taxas, custo_rec, custo_transp)
preco_total_compra_EUR	preco_base, taxas, custo_rec, custo_transp, taxa cambio (csv)	money	preco_total_compra_AOA / taxa_cambio

#### Tabela Fact\_Venda

Measures	OLTP Fields	Data Type	Formula
valor_venda_EUR	valor_venda, taxa cambio (csv)	money	valor_venda_AOA / taxa_cambio
valor_venda_AOA	valor_venda	money	valor_venda_AOA

#### 4.2.2. TABELA DE DIMENSÕES (DIMENSIONAL TABLE)

Dimension	OLTP Entities
fornecedor	fornecedor
produto	produto
stand	stand
cliente	cliente
data	datas (xls)

#### 4.2.3. ESQUEMA RELACIONAL

Os esquemas relacionais da *Staging Area Cars\_Co\_SA* e da *Data Warehouse Cars\_Co\_DW* estão apresentados nas figuras seguintes.

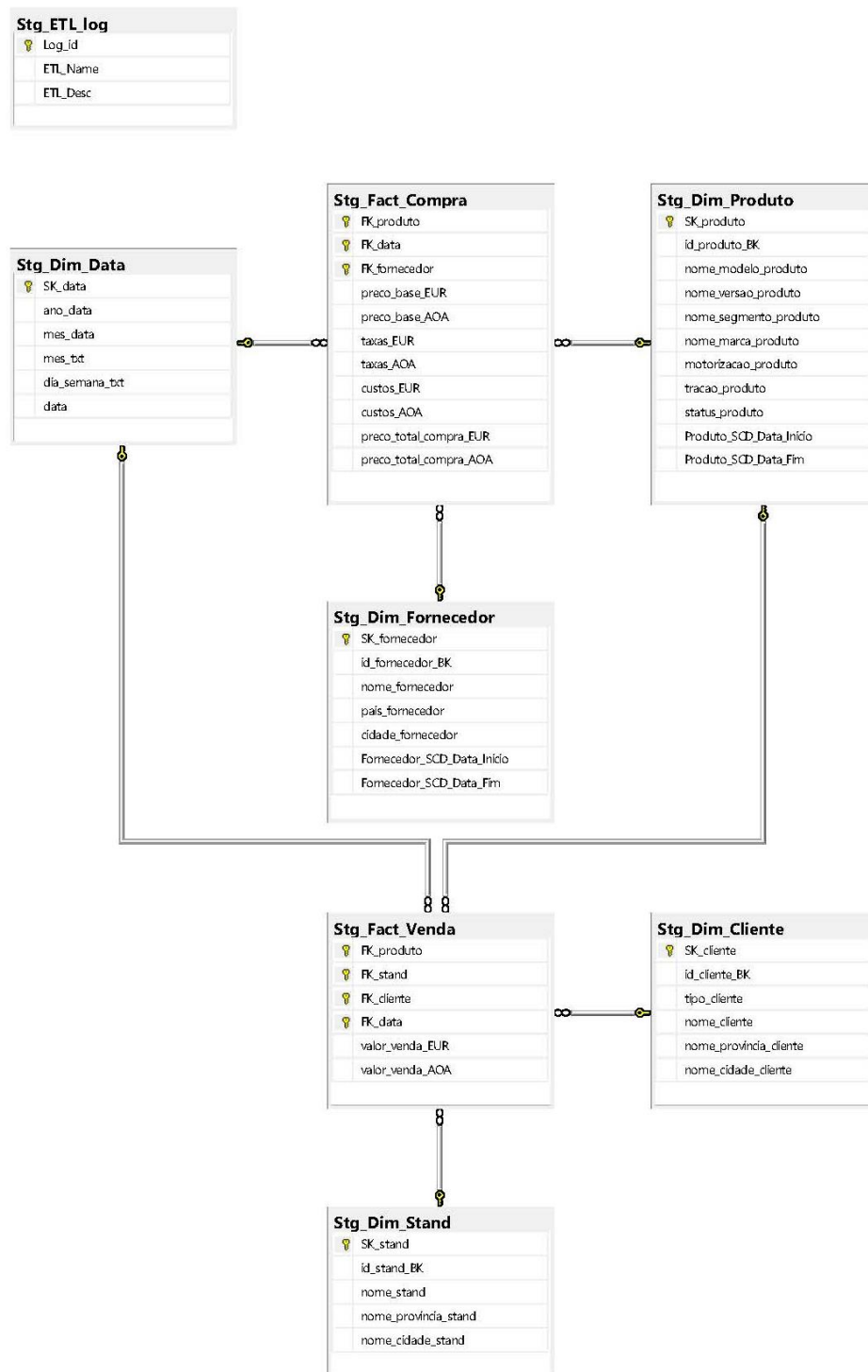


Fig. 12 – Galaxy schema da Data base Cars\_Co\_SA preparado para SQL Server usando a notação Martin IE.

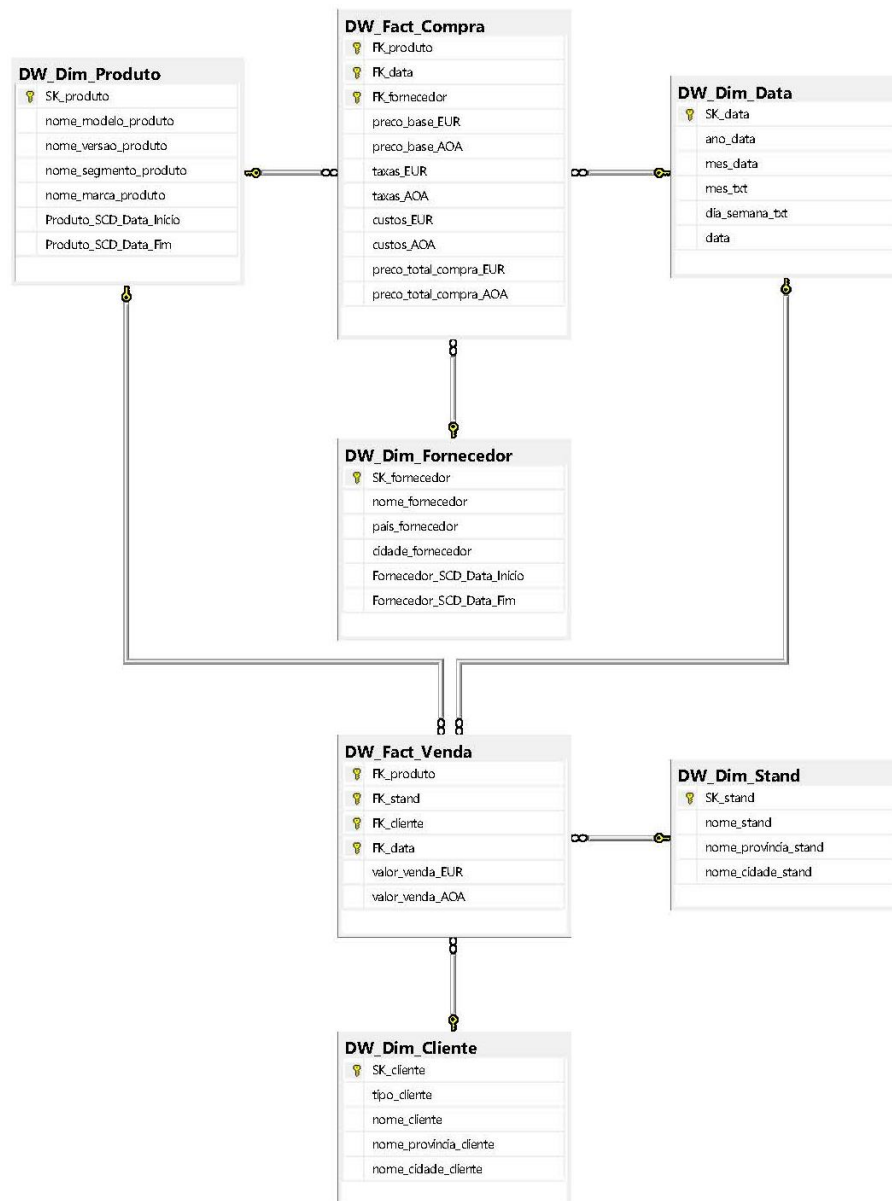


Fig. 13 – Galaxy schema da Data base Cars\_Co\_DW preparado para SQL Server usando a notação Martin IE.

### 4.3. PROBLEMAS ENCONTRADOS

#### Data Source

A criação da *Data Source* foi um grande desafio devido a inúmeros problemas de inconsistência e de incompletude dos dados, que consumiu muito tempo. Foi necessário criar dados adicionais e efectuar pesquisa sobre o sector automóvel e sobre o país, em termos demográficos, de consumo *per capita* por província e taxas cambiais, entre outros.

Uma vez que foi necessário adicionar aos elementos iniciais da Base de Dados diversos campos em falta (dados de fornecedor, vendedor e cliente) a atribuição dos respectivos registos foi realizada de forma

aleatória seguindo uma lei de distribuição normal. Posteriormente foi necessário enviesar essa distribuição de modo a que o histograma de vendas (vendedores e clientes) fosse tendencioso para as províncias mais populosas e para a aquisição por empresas. Não foi feito um acerto entre o tipo de veículo e o cliente pelo que poderá suceder que um cliente particular tenha adquirido ao longo do tempo um veículo ligeiro de mercadorias e outro pesado de passageiros, por exemplo. O resultado final apresentado para a *Data Source* em SQL é baseado numa geração pseudo-aleatória de registos.

Existem valores NULL por ausência de registos, por exemplo cliente de um produto que ainda não foi vendido.

### Staging Area

Sucederam problemas na instalação dos programas em termos de compatibilidade e ligação ao servidor local e, também, na instalação de programa de transferência de dados entre aplicações Microsoft Excel e SQL Server que muito perturbaram a prossecução do trabalho.

A instrução de *split* das datas de carregamento está condicionada à totalidade dos registos existentes...

## 5. RESULTADOS OBTIDOS

### 5.1. STAGING AREA

O controlo do processo ETL de transferência de dados da *Data Source* + *flat files* para a *Cars\_Co\_SA* pode ser observado na Fig. 14.

Results		Messages	
	Log_id	ETL_Name	ETL_Desc
179	179	ETL ID:07-01-2018 05:57:58	Início do ETL; loading Staging Area
180	180	ETL ID:07-01-2018 05:57:58	Compra mais recente: 2017-09-29
181	181	ETL ID:07-01-2018 05:57:58	Limpar e Carregar Dim tables: Yes
182	182	ETL ID:07-01-2018 05:57:58	Venda mais recente: 2017-10-02
183	183	ETL ID:07-01-2018 05:57:58	Fact tables truncadas
184	184	ETL ID:07-01-2018 05:57:58	Dim tables apagadas
185	185	ETL ID:07-01-2018 05:57:58	Dim tables carregadas
186	186	ETL ID:07-01-2018 05:57:58	Carregada Fact Compra
187	187	ETL ID:07-01-2018 05:57:58	Numero linhas carregadas na FACT compra: 21882
188	188	ETL ID:07-01-2018 05:57:58	Carregada Fact Venda
189	189	ETL ID:07-01-2018 05:57:58	Numero linhas carregadas na FACT Venda: 21695
190	190	ETL ID:07-01-2018 05:57:58	Concluido ETL; Staging Area carregada

Fig. 14 – Extracto do ETL\_log produzido durante o carregamento da *Cars\_Co\_SA*.



O processo de ETL pode ser visualizado na árvore de tarefas de *Control Flow* representada da Fig. 15.

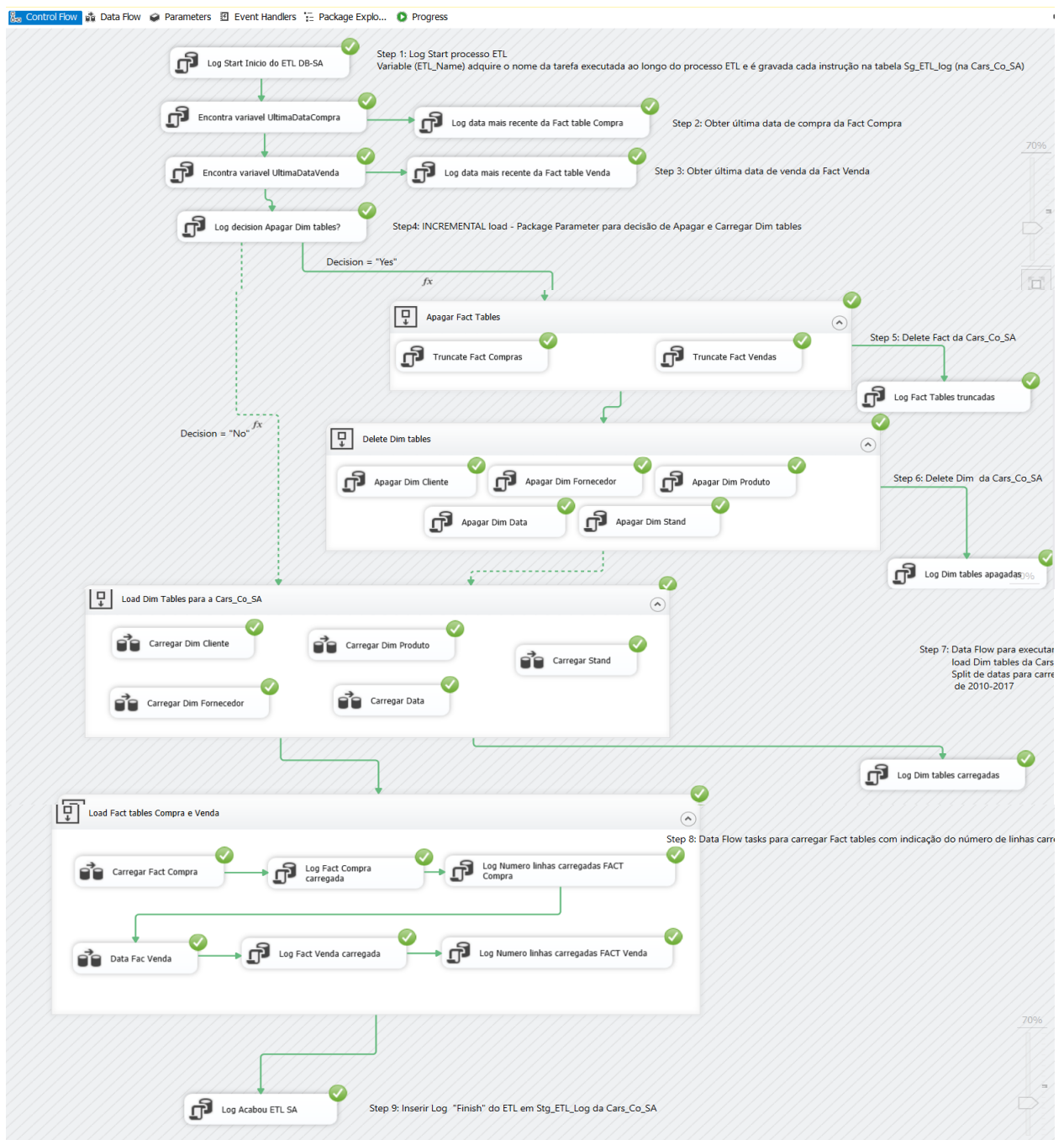


Fig. 15 – Imagem do processo ETL com parâmetro ParamClearDimensionTables="Yes", que pode ser obtido a partir do ficheiro SSIS\_group11.sln.

Foram consideradas duas SCD tipo 2, aplicadas a Stg\_Dim\_Fornecedores e a Stg\_Dim\_Produto, como se exemplifica para a tabela dimensional Stg\_Dim\_Fornecedores na Fig. 16.

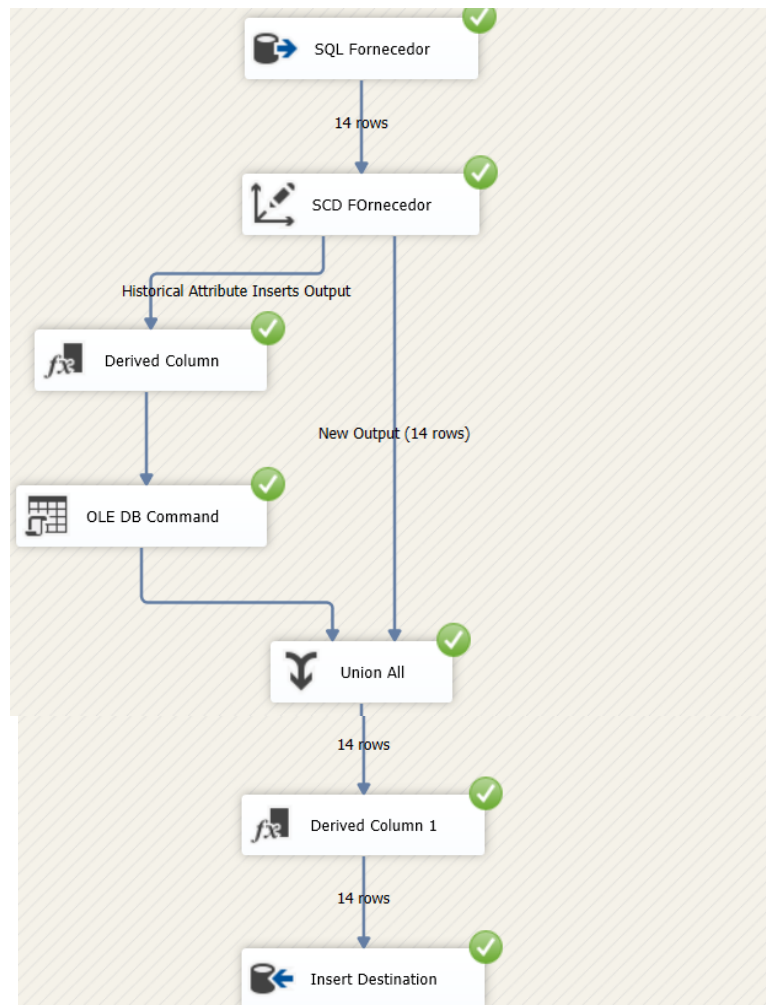


Fig. 16 – Exemplo de *Data flow* com SCD tipo 2, aplicado a Stg\_Dim\_Fornecedor, na Cars\_Co\_SA.

Foram utilizados dois *flat files* tipo \*.csv, referentes ao detalhe das datas e às taxas de câmbio. Foram efectuadas conversões de dados, *sort* e *merge join* de dados de forma a poderem ser usados os registos desses ficheiros. Na Fig. 17 exemplifica-se a utilização de câmbios.csv para a tabela Stg\_Fact\_compra.

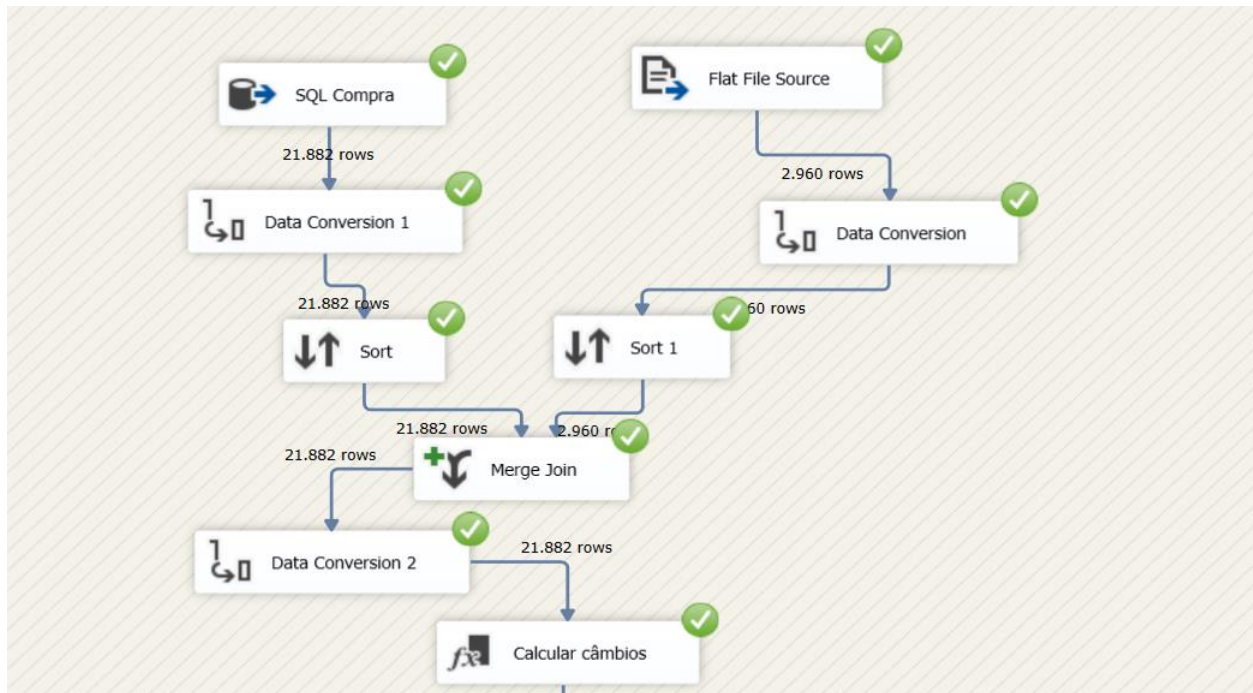


Fig. 17 – Exemplo de *Data flow* com *flat file* câmbios.csv aplicado a Stg\_Fact\_Compra, na Cars\_Co\_SA.

## 5.2. DATA WAREHOUSE

Uma vez estabilizada a Staging Area a DW é carregada com controlo apenas das SCD referentes às tabelas Dim e Fact, conforme exposto anteriormente. O processo de ETL pode ser visualizado na árvore de tarefas de *Control Flow* representada da Fig. 18.

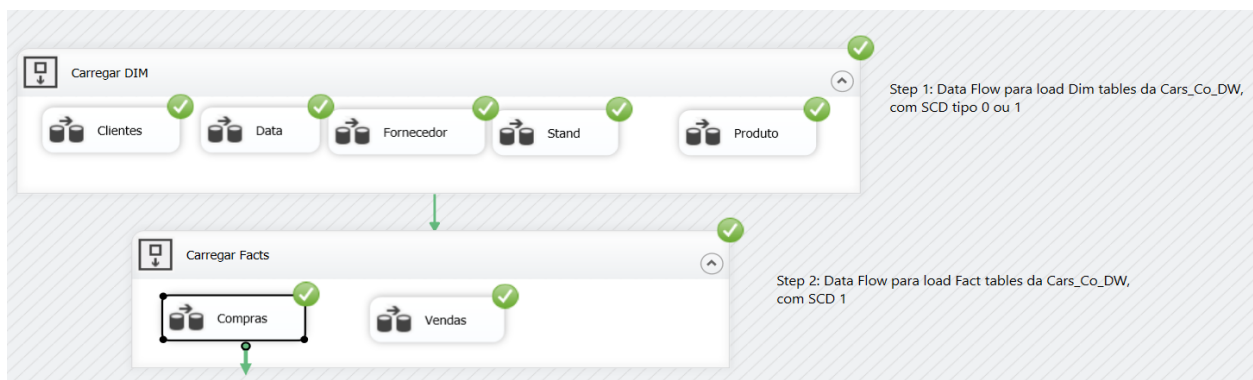


Fig. 18 – Imagem do processo ETL de carregamento da Cars\_Co\_DW.

Na Fig. 19 apresenta-se um exemplo de SCD tipo 1, aplicada à tabela DW\_Fact\_Compra.

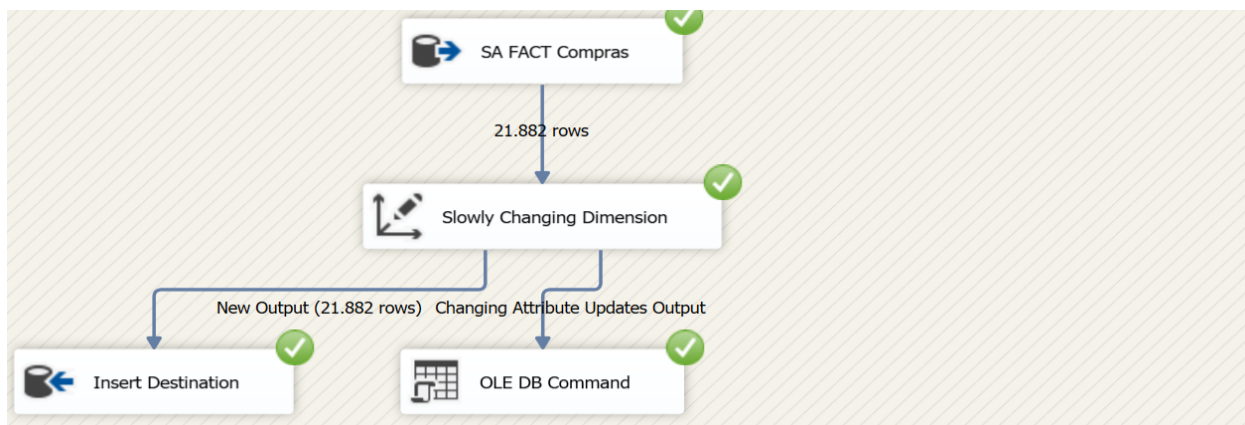


Fig. 19 – Imagem de SCD tipo 1 no processo ETL de carregamento da Cars\_Co\_DW, sobre a DW\_Fact\_Compra.

Na Fig. 20 apresenta-se a utilização de variáveis, particularizando o caso de UltimaDataCompra.

Query utilizado directamente no SQL	SQL statement utilizado em Visual Studio													
<pre>select case when max(data) is null then cast('2001-01-01' as Date) else max(data) end as UltimaDataCompra from Stg_Fact_Compra INNER JOIN Stg_Dim_Data ON Stg_Fact_Compra.FK_data=Stg_Dim_Data.SK_data</pre> <div> <div>Results</div> <div>Messages</div> <table border="1"> <thead> <tr> <th></th> <th>UltimaDataCompra</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>2017-09-29</td> </tr> </tbody> </table> </div>		UltimaDataCompra	1	2017-09-29	<div> <div>General</div> <div>Parameter Mapping</div> <div>Result Set</div> <div>Expressions</div> </div> <div> <div>General</div> <div>Enter SQL Query</div> <pre>select case when max(data) is null then cast('2001-01-01' as Date) else max(data) end as UltimaDataCompra from Stg_Fact_Compra INNER JOIN Stg_Dim_Data ON Stg_Fact_Compra.FK_data=Stg_Dim_Data.SK_data</pre> </div> <div> <div>Results</div> <div>Messages</div> <table border="1"> <thead> <tr> <th>Log_id</th> <th>ETL_Name</th> <th>ETL_Desc</th> </tr> </thead> <tbody> <tr> <td>203</td> <td>203</td> <td>ETL ID:07-01-2018 06:54:14</td> </tr> <tr> <td>204</td> <td>204</td> <td>ETL ID:07-01-2018 06:54:14</td> </tr> </tbody> </table> </div>	Log_id	ETL_Name	ETL_Desc	203	203	ETL ID:07-01-2018 06:54:14	204	204	ETL ID:07-01-2018 06:54:14
	UltimaDataCompra													
1	2017-09-29													
Log_id	ETL_Name	ETL_Desc												
203	203	ETL ID:07-01-2018 06:54:14												
204	204	ETL ID:07-01-2018 06:54:14												

Fig. 20 – Imagem de SCD tipo 1 no processo ETL de carregamento da Cars\_Co\_DW, sobre a DW\_Fact\_Compra.