



Pontifícia Universidade Católica do Rio de Janeiro
Coordenação Central de Extensão

Artur Genelhoud Bastos

**Desenvolvimento de Data Warehouse para auxiliar nas tomadas de
decisão de empresas varejistas.**

Trabalho de Conclusão de Curso de Pós-Graduação

Departamento de Engenharia Elétrica

Rio de Janeiro
dezembro de 2018

Artur Genelhoud Bastos

**Desenvolvimento de Data Warehouse para auxiliar nas tomadas de decisão
de empresas varejistas**

Trabalho de Conclusão de Curso de Pós-Graduação

Trabalho de Conclusão de Curso apresentado ao Departamento de Engenharia Elétrica da PUC/RIO, como parte dos requisitos para obtenção do título de Especialização em Business Intelligence.

Professor Orientador: Anderson Silva do Nascimento

Rio de Janeiro
dezembro de 2018

Agradecimentos

Gostaria de agradecer a todos professores do curso de BI Master da PUC que transferiram os conhecimentos necessários para elaboração deste trabalho e especialmente ao meu orientador.

Resumo

O objetivo deste trabalho é facilitar o armazenamento dos dados, harmonizar as informações internas com o mercado e auxiliar na tomada de decisão.

As ferramentas utilizadas neste projeto são todas Open Source, com exceção da ferramenta utilizada para a elaboração do DashBoard, porém a mesma tem uma versão gratuita (neste trabalho foi utilizado a versão gratuita).

Para organizar e armazenar todos os dados, tanto internos quanto externos, será criado um Data Warehouse (DW). A fim de harmonizar os dados internos e os dados informados pelo mercado, será utilizada a classificação utilizada nos dados de mercado, com isso os dados internos da empresa deverão ser classificados da mesma maneira que os dados de mercado, assim todos os dados serão armazenados e classificados da mesma forma, facilitando o acesso e a comparação da informação.

Ao final será elaborado um dashboard, onde será possível comparar as vendas da empresa em relação ao mercado, ajudando a empresa a entender se as suas vendas estão alinhadas com o mercado, também tem como objetivo mostrar novas tendências do mercado em alguma categoria ou marca, auxiliando os gerentes a tomarem melhores decisões de acordo com o comportamento do mercado.

Sumário

1. Introdução	7
1.1 <i>Motivação</i>	7
1.2 <i>Objetivos</i>	7
1.3 <i>Descrição</i>	8
1.4 <i>Organização da monografia</i>	8
2. Conceitos e modelo de dados.....	10
2.1 <i>Modelo de dados transacional</i>	11
2.2 <i>Modelo de dados Dimensional</i>	12
3. Tratamento e carga dos dados	14
3.1 <i>Carga modelo transacional</i>	14
3.2 <i>Carga modelo dimensional</i>	16
4. Dashboard operacional	22
5. Resultados Obtidos.....	24
6. Conclusões	27
6.1. <i>Trabalhos futuros</i>	27
Referências Bibliográficas	29
Anexos	30

1. Introdução

O varejo nasce a partir das necessidades emergentes por produtos e serviços baseados em um processo de troca entre indivíduos, onde revelam-se as figuras de produtores e consumidores. As trocas eram feitas através de um processo de escambo, ou seja, um indivíduo que possuía certa quantidade de determinado produto negociava seu produto com outro indivíduo que possuía um bem diferente. Caso ambas as necessidades dos negociantes fossem atendidas, efetuava-se a troca (Diego Amorim, 2016).

Com a expansão dos mercados produtor e consumidor em meados do século XIX, o varejo estabelece-se como um setor de grande importância. Atualmente, a tecnologia da informação tem apoiado substancialmente a evolução do varejo (Diego Amorim, 2016).

Com o avanço da tecnologia, hoje em dia é possível ter uma visão mais ampla do mercado varejista, com isso entender melhor as tendências do mercado e identificar novas oportunidades, também com o auxílio de algumas técnicas de aprendizagem de máquina se torna possível melhorar a distribuição e estoque de produtos.

1.1 Motivação

Este projeto visa automatizar o tratamento dos dados do mercado e criar um DW onde facilite a consulta de dados, melhorando a visualização dos mesmos, e que possibilite a comparação dos dados internos com o mercado. Com o objetivo de identificar possíveis *gaps* e oportunidades de venda, auxiliando na tomada de decisão da empresa.

1.2 Objetivos

Uma das maiores preocupações da empresa é saber se os produtos que estão vendendo estão alinhados com o mercado, ter a certeza de apostar em uma marca que os clientes irão consumir, para isso é preciso entender quais são as tendências

e necessidades que existe no mercado, pensando nestes pontos esse trabalho tem dois grandes objetivos.

- Criar uma estrutura de dados consistente, que seja atualizada de forma automática mensalmente.
- Desenvolver um dashboard para melhor acompanhamento dos dados e auxiliar nas tomadas de decisões da empresa.

1.3 Descrição

O desenvolvimento da monografia se faz em algumas partes.

- Elaboração dos modelos transacional e multidimensional.
- Desenvolvimento do DW (Data Warehouse).
- Tratamento e transformação dos dados (ETL).
- Criação do DM (Data Mart).
- Análise de resultados (Dashboard).

A coleta de dados é realizada em duas fontes, sendo a primeira os dados internos de uma empresa varejista e a segunda fonte uma empresa especializada na divulgação de dados do mercado.

Serão utilizados dados internos de uma empresa de varejo, que não será divulgada o nome para fins de preservar a segurança da informação, para comparação com o mercado.

Para auxiliar na visão de como o mercado de varejo se comportar são recebidas mensalmente informações sobre vendas em todo mercado brasileiro através da instituição GFK.

1.4 Organização da monografia

Esta monografia está dividida em 6 capítulos, descritos a seguir:

- O capítulo 1 faz a introdução do tema, assim como traz a motivação e o objetivo desta monografia e também uma breve descrição e a organização deste trabalho.
- O capítulo 2 traz o conceito de um Data Warehouse e os modelos de estruturas utilizados.
- O capítulo 3 tem como foco o processo de ETL.
- O capítulo 4 apresenta a estrutura e metodologia de carga do Data Mart e um Dashboard operacional.
- O capítulo 5 traz alguns resultados que podem ser analisados através do dashboard.
- O capítulo 6 apresenta as conclusões obtidas e identifica possíveis trabalhos futuros.

2. Conceitos e modelo de dados

Antes de começar a apresentar a estrutura e modelo utilizado e valido lembrar o porque da utilização de um Data Warehouse (DW) para este projeto, o DW possui uma estrutura que permite uma analise de grandes volumes de dados, o mesmo e a base do armazenamento das informações necessárias para a utilização dos diretores e gestores na tomada de decisão (Diego elias, 2014).

A arquitetura do DW e dividida em diversas partes, conforme pode ser visto na Figura 1 - Arquitetura Data Warehouse. Como fonte de dados temos um arquivo com os dados internos e um arquivo contendo os dados de mercados, o processo de ETL (extração, transformação e carga dos dados) e responsável por todo o processo de extração, condução, tratamento e limpeza dos dados, neste projeto toda o processo de ETL será realizado com uma feramente Open Sourcer (para o processo de ETL pode ser utilizado a feramente que mais convém), os dados que agregam valor a corporação serão reunidos no DW.

Os Data Mart apresentam estruturas muito similar ao DW, porém com uma menor quantidade de informação e que podem ser identificados por assunto ou gerencias especificas.

Na Figura 1 pode ser visto no final da cadeia uma saída OLAP e Data Mining, a saída OLAP (Online analytical Processing) se refere a utilização de ferramentas de análise em múltiplas perspectivas, entrando no conceito do Cubo de dados uma estrutura que permite uma rápida analise de dados de acordo com múltiplas dimensões. Já o Data Minning (mineração de dados) e o processo de retirar conhecimento dos dados utilizando métodos como aprendizado de maquinas. Como o foco deste projeto está na criação do DW não iremos abordar os processos OLAP e de Data Minning (Daniel Lemire, 2010).

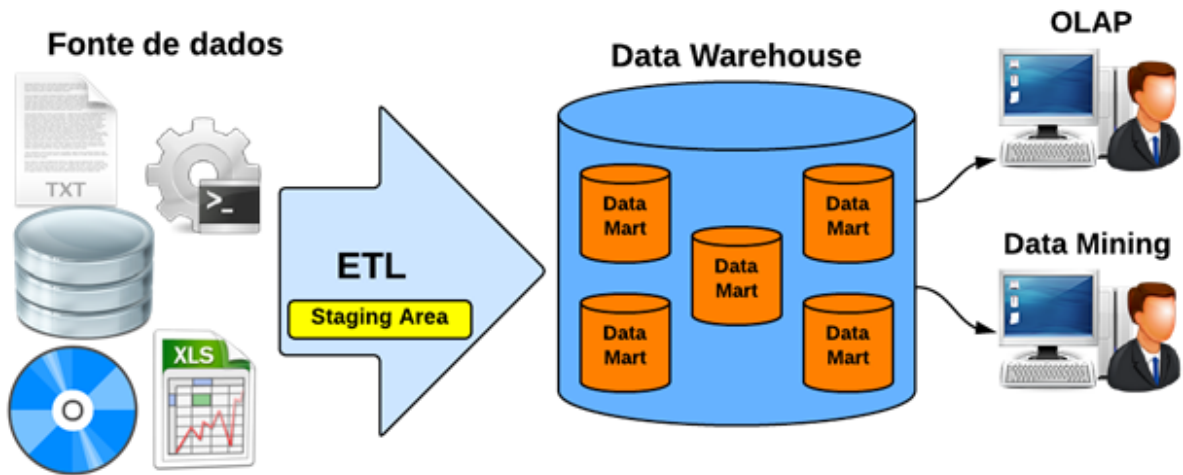


Figura 1 - Arquitetura Data Warehouse

Fonte: Diego Elias 2014

2.1 Modelo de dados transacional

O modelo de dados transacional é um modelo que é projetado para sistemas transacionais, ou seja, transação por transação. É projetado unicamente para o processamento de informações e ainda pode ser usado para alguns tipos de análise imediatas, porém não consegue atender a necessidade do cliente quanto a informações úteis para apoio nas tomadas de decisões (Allyson Alves, 2011).

Como os dados coletados são enviados mensalmente foi montado um modelo de dados transacional, a estrutura das tabelas pode ser visto no Anexo 1, para trabalhar e visualizar melhor os dados, na Figura 2 **Erro! Fonte de referência não encontrada.** pode ser visto o modelo de dados enviados pela empresa especializada em dados de mercado. Enquanto que na Figura 3 pode ser visto o modelo dos dados internos, esse modelo está muito simplificado pois será feita a classificação dos dados internos conforme no mesmo modelo dos dados do mercado.

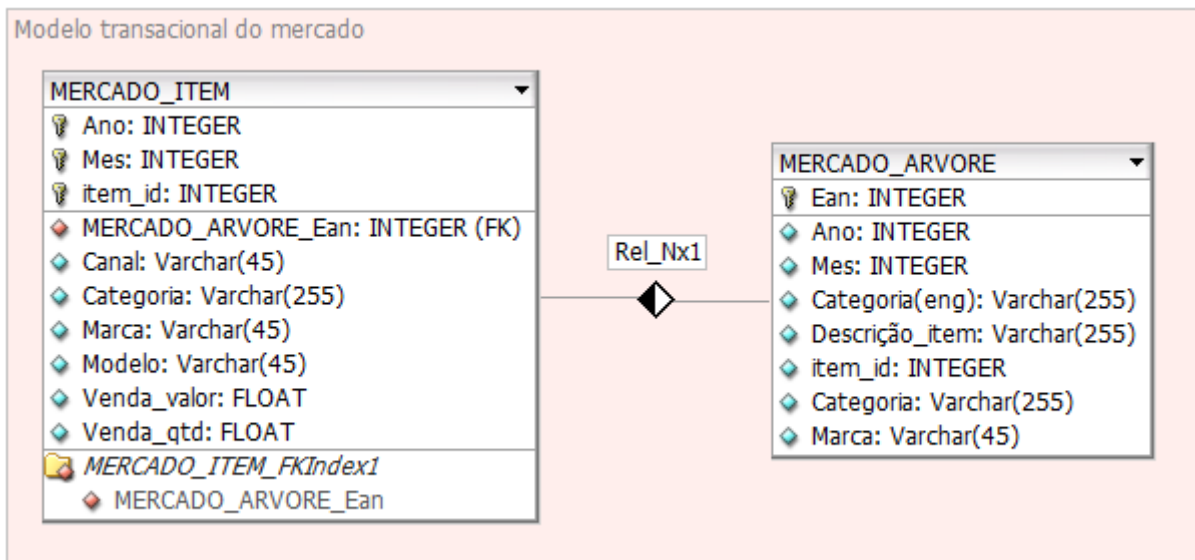


Figura 2- Modelo de dados transacional do mercado

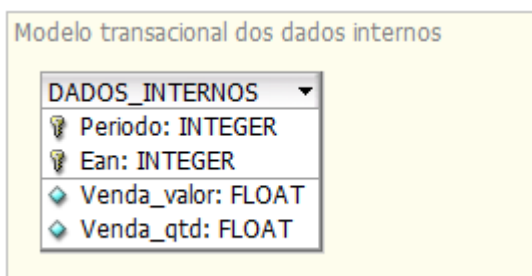


Figura 3- Modelo de dados transacional dos dados internos

2.2 Modelo de dados Dimensional

O modelo de dados dimensional é exatamente o oposto do modelo transacional. É um modelo que visa armazenar uma imensa quantidade de dados. Um modelo que vai permitir uma análise histórica dos dados, desempenho nas consultas e facilidade no desenvolvimento das mesmas. Essas são algumas das finalidades de um modelo de dados dimensional e para isso existem alguns tipos de modelos, que são: Star Schema e Snowflake (Allyson Alves, 2011).

Neste projeto será utilizado o modelo Star Schema (modelo estrela), que consiste em uma tabela fato e várias dimensões. Na Figura 4 pode ser observado o modelo estrela que foi estruturado para receber os dados internos e os dados do mercado.

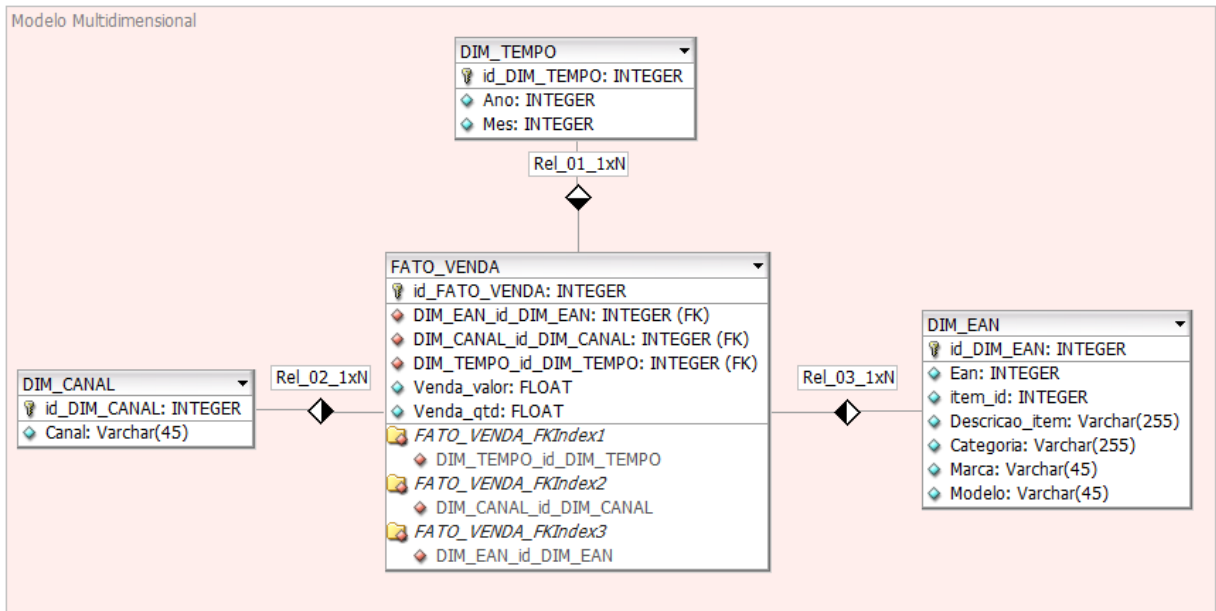


Figura 4- Modelo de dados Dimensional

O modelo consiste em três tabelas de dimensões e uma tabela fato.

- **DIM_TEMPO:** Como os dados recebidos pelo mercado são disponibilizados mensalmente e não trazem a informação do dia de venda de cada item, logo não é possível aumentar a granularidade da dimensão, fazendo com que a dimensão tempo fique limitada com somente dois campos (Ano e Mês).
- **DIM_CANAL:** Esta dimensão foi criada para distinguir os dados internos dos dados do mercado, a dimensão canal vai ficar limitar a dois valores “dados de mercado” e “dados internos”. Com isso será fácil a identificação dos dados na tabela FATO.
- **DIM_EAN:** Nesta dimensão vão ser encontrados todas as características dos itens, pois cada item tem um EAN (código de barras) único, logo cada item vai ter as suas próprias características.
- **FATO_VENDA:** Todos os dados de mercado e internos serão armazenados com o mesmo padrão nesta tabela, simplificando a consulta e análise de dados.

3. Tratamento e carga dos dados

Como os dados internos e os dados de mercados são disponibilizados em arquivos mensais dos tipos .csv e .xlsx, e não foi disponibilizado nenhum API para coleta dos dados de forma automática, não será necessário a criação de um script ou scrap para obter os dados, com isso passamos direto para o tratamento dos dados recebidos.

3.1 Carga modelo transacional

Para realizar toda a parte de transformação de dados e carga dos dados no banco de dados, tanto no modelo transacional quanto no modelo dimensional, foi utilizado o software Talend Open Studio (TOS). Todos os esquemas que serão mostrados mais a frente foram feitos neste software.

Talend Open Studio é uma ferramenta *Open Source* de integração de dados baseada em Eclipse RCP que suporta principalmente implementações orientadas a ETL e é fornecido para implementação local, bem como em um modelo de entrega de software como serviço (SaaS). O Talend Open Studio é usado principalmente para integração entre sistemas operacionais, bem como para ETL (Extrair, Transformar, Carregar) para Business Intelligence e Data Warehouse. (Talend).

Os dados de mercado são disponibilizados em dois arquivos distintos (“árvore” e “Dados mercado”). No arquivo “árvore” (a estrutura do arquivo pode ser vista no Anexo 1) foi preciso somente transformar a data, separando o mês e ano da data original, os mesmos também precisaram ser convertidos para inteiros para realizar esta transformação foi utilizado o componente tMap do TOS (conforme pode ser visto na Figura 6).

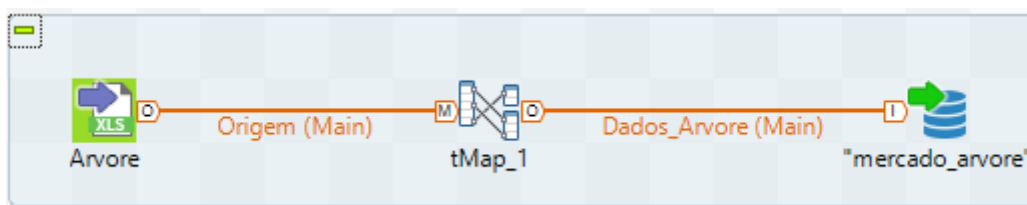


Figura 5- TOS, Pacote de carga de dados do arquivo “árvore” para o modelo transacional

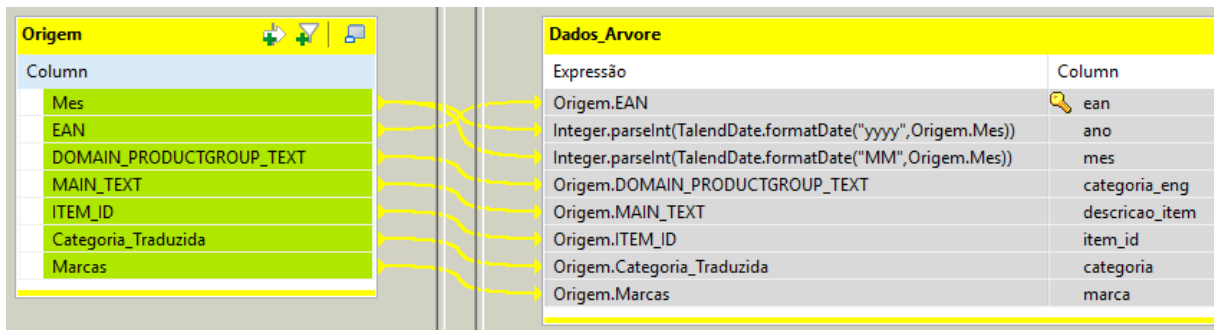


Figura 6- TOS, tMap para transformação e carga dos dados árvore para o banco transacional

Para o arquivo “Dados mercado” (a estrutura do arquivo pode ser vista no Anexo 1) primeiro foi preciso converter o campo “Mês”, que inicialmente vem com o nome do mês em formato de *string* e em inglês, para um inteiro que represente o mês dos dados. No campo “canal” os dados originais apresentam três opções “Físico”, “Online” e “Total mercado”. Alguns itens apresentam as opções “Físico” e “Online”, os mesmos representam as vendas destes itens nos mercados físicos e online, já no caso dos itens que só tem a opção “Total mercado”, segundo o fornecedor dos dados, estes itens não apresentam a abertura físico e online para preservar a identidade dos fornecedores. Para este estudo será realizado um filtro no campo “canal” na opção “Online” afim de focar na venda de itens em lojas físicas. Para buscar e preencher o campo EAN no banco de dados, foi preciso realizar um *join* com a tabela árvore como pode ser visto na Figura 8, também foi preciso realizar uma conversão do campo mês para inteiro.

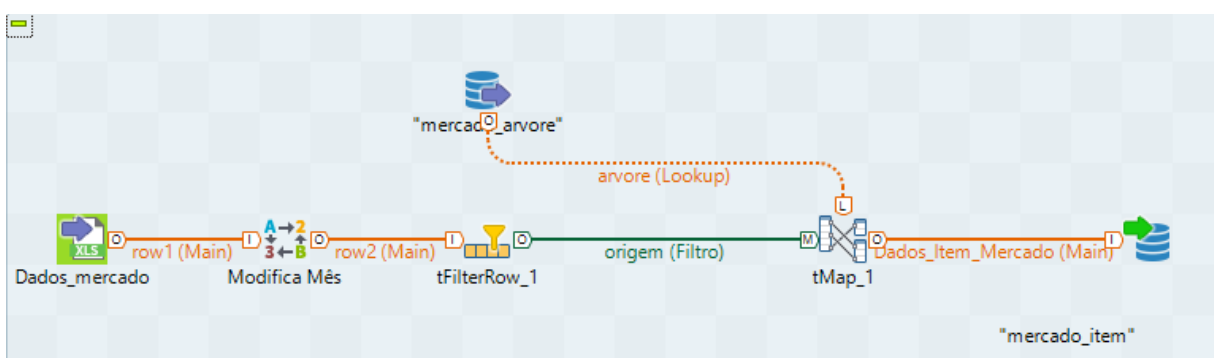


Figura 7 - TOS, Pacote de carga de dados do arquivo “Dados mercado” para o modelo transacional

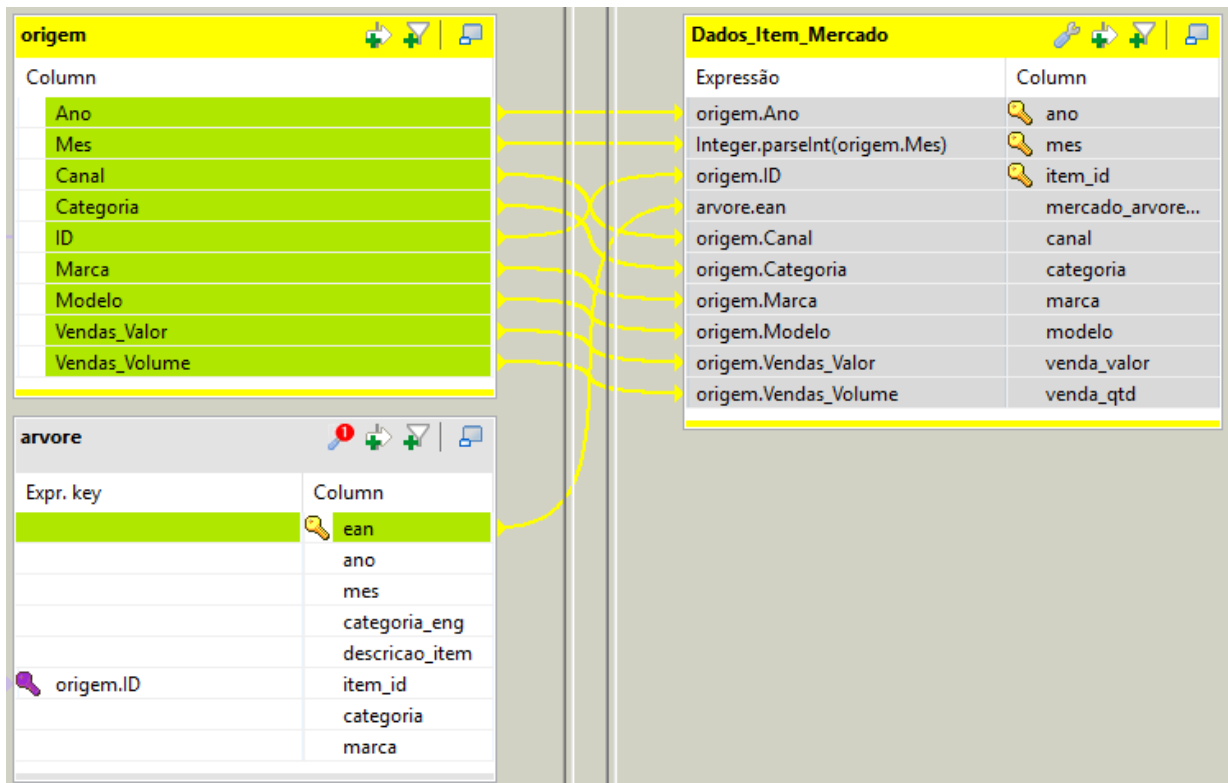


Figura 8- TOS, tMap join com a tabela árvore para buscar os dados do EAN

Para os dados internos não foi preciso realizar nenhum tipo de transformação nos dados, como a classificação vai ser feita conforme os dados de mercado os dados internos só precisam do EAN para se encaixar no modelo do mercado, também foi utilizado um tMap para carga simplesmente para facilitar no direcionamento correto dos dados.

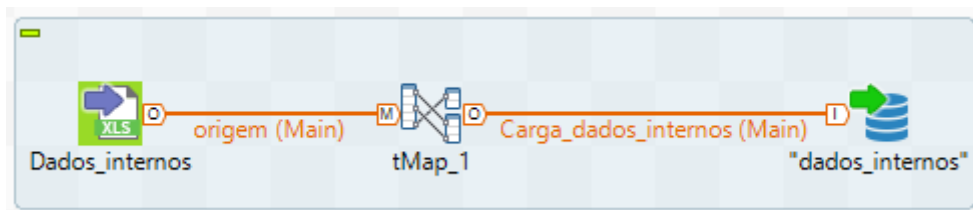


Figura 9- TOS, pacote de carga dos dados internos para o modelo transacional

3.2 Carga modelo dimensional

Após a carga dos dados para o modelo transacional será realizada a carga das dimensões e da tabela fato do modelo dimensional. Para este estudo temos três dimensões “canal”, “tempo” e “ean”.

Dimensão “canal” será utilizada para identificar a origem dos dados, esta dimensão apresentará somente duas opções “dados internos” e “dados mercado”. Como esta dimensão representa estritamente a origem dos dados logo não será preciso criar uma rotina de atualização desta dimensão, pois a mesma só será atualizada no caso de uma nova fonte de dados.

Para as dimensões “tempo” e “ean”, foi preciso a criação de uma rotina de atualização e incremento de dados (conforme pode ser visto na Figura 10).

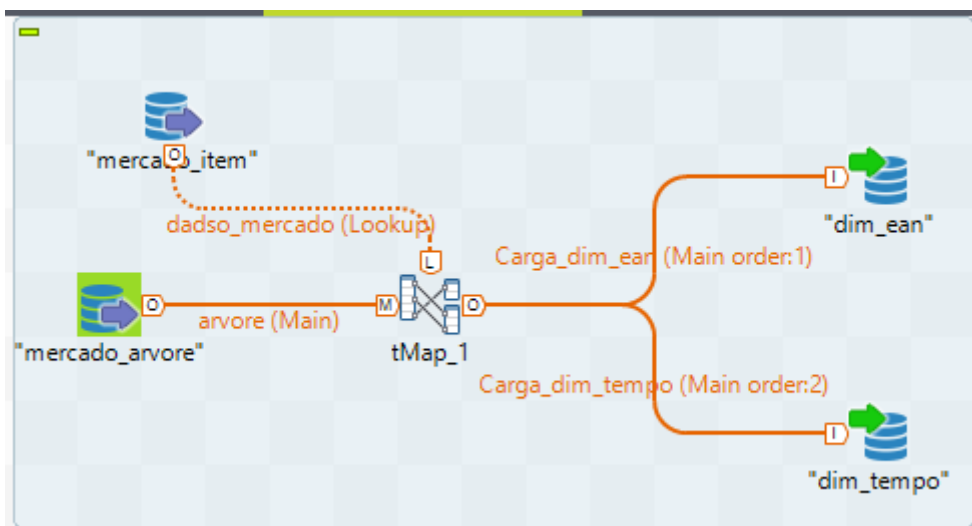


Figura 10- TOS, pacote de carga das dimensões “tempo” e “ean”

Na dimensão “tempo” são inseridos os dados dos campos “ano” e “mês” da tabela “mercado_arvore”, afim de evitar duplicidade nos dados os campos “ano” e “mês” da tabela “dim_tempo” foram tratados como chaves da tabela (conforme pode ser visto na Figura 11). Para a dimensão “ean” foram utilizados os demais campos da tabela “mercado_arvore” e foi criado um *join* com a tabela “mercado_item” afim de obter a informação do modelo do item (conforme Figura 11).

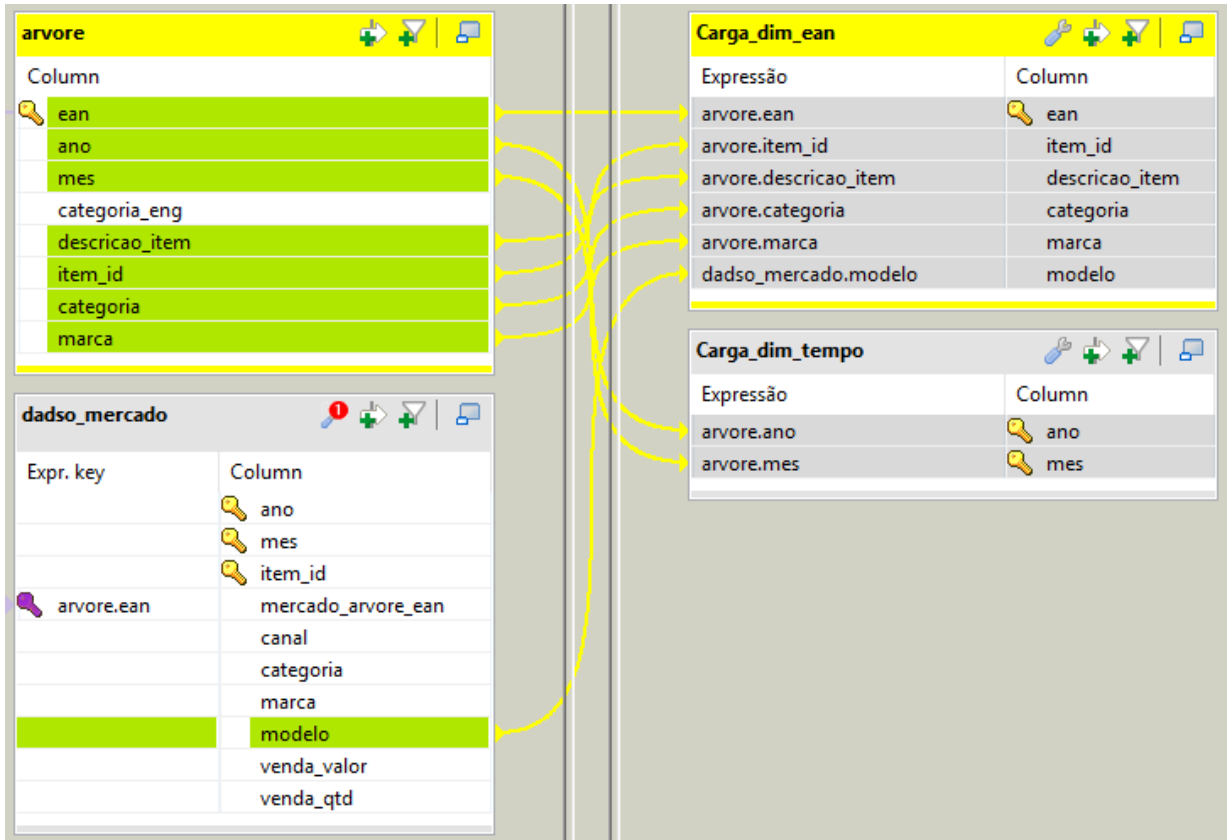


Figura 11- TOS, tMap carga das dimensões “tempo” e “ean”

Após a carga das dimensões do modelo, a tabela fato pode ser carregada. A carga da tabela fato será feita em dois momentos, o primeiro com a carga dos dados do mercado e no segundo com a carga dos dados internos.

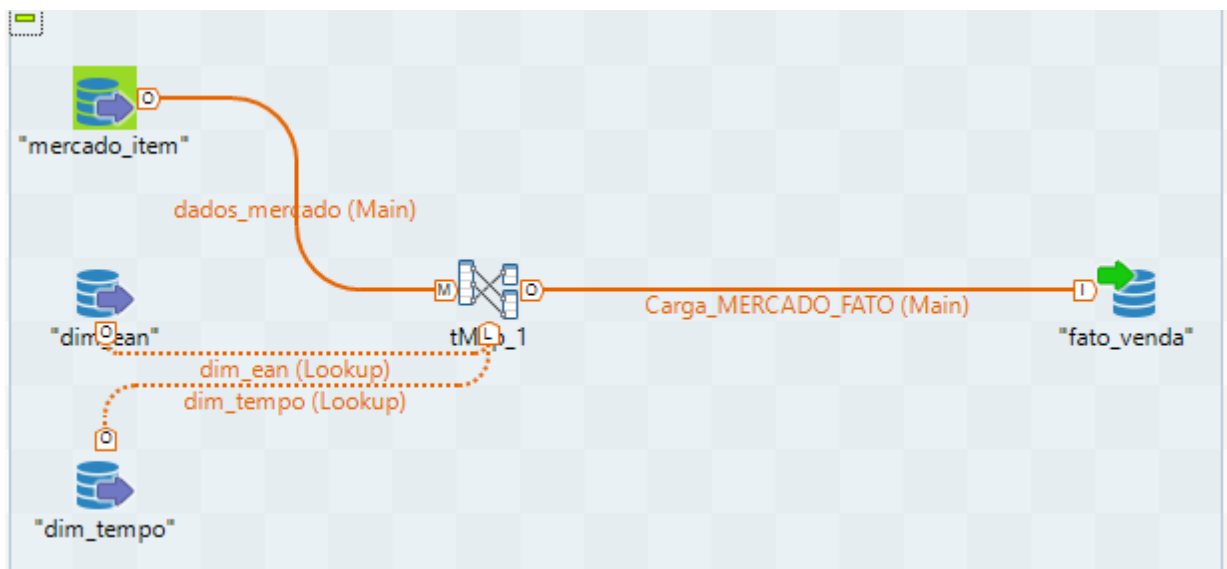


Figura 12- TOS, pacote de carga dos dados do mercado para a tabela FATO do modelo dimensional

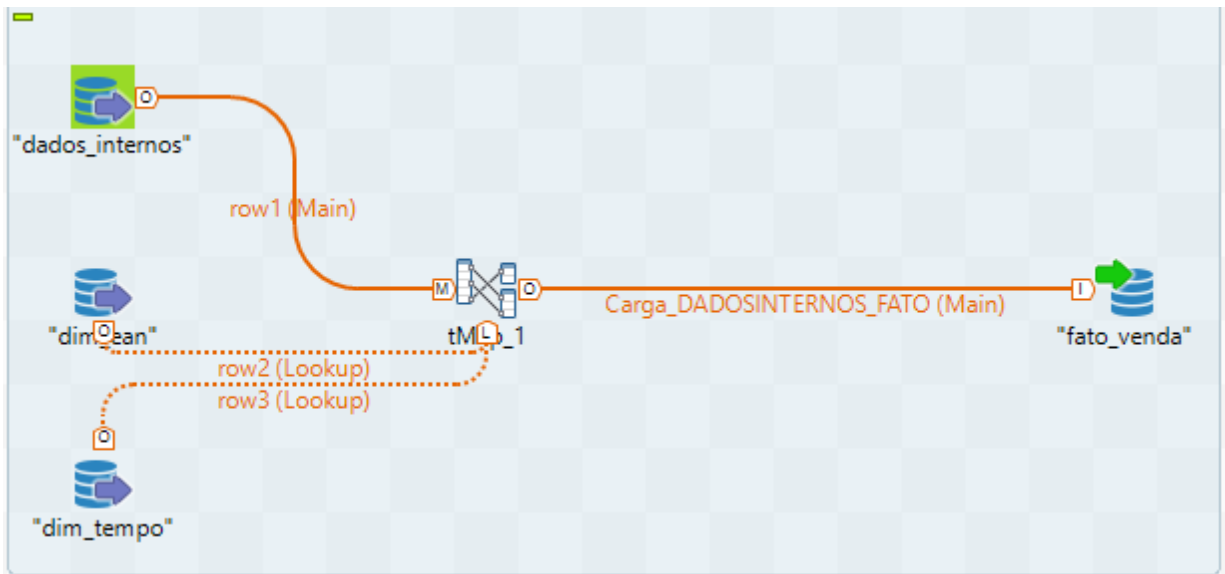


Figura 13- TOS, pacote de carga de dados internos para a tabela FATO do modelo dimensional

Para a carga dos dados de mercado foi realizado um *join* entre as tabelas “mercado_item” e “dim_tempo” utilizando os campos “ano” e “mes” para buscar o id da dimensão “tempo”, também foi realizado um *join* entre as tabelas “mercado_item” e “dim_ean” utilizando o campo “item_id” como chave para buscar o id da dimensão “ean”. Para a carga dos dados de mercado foi colocado como default o valor “1” para a dimensão “canal”, que representa a opção “dados mercado”, todas as conexões podem ser vistas na Figura 14.

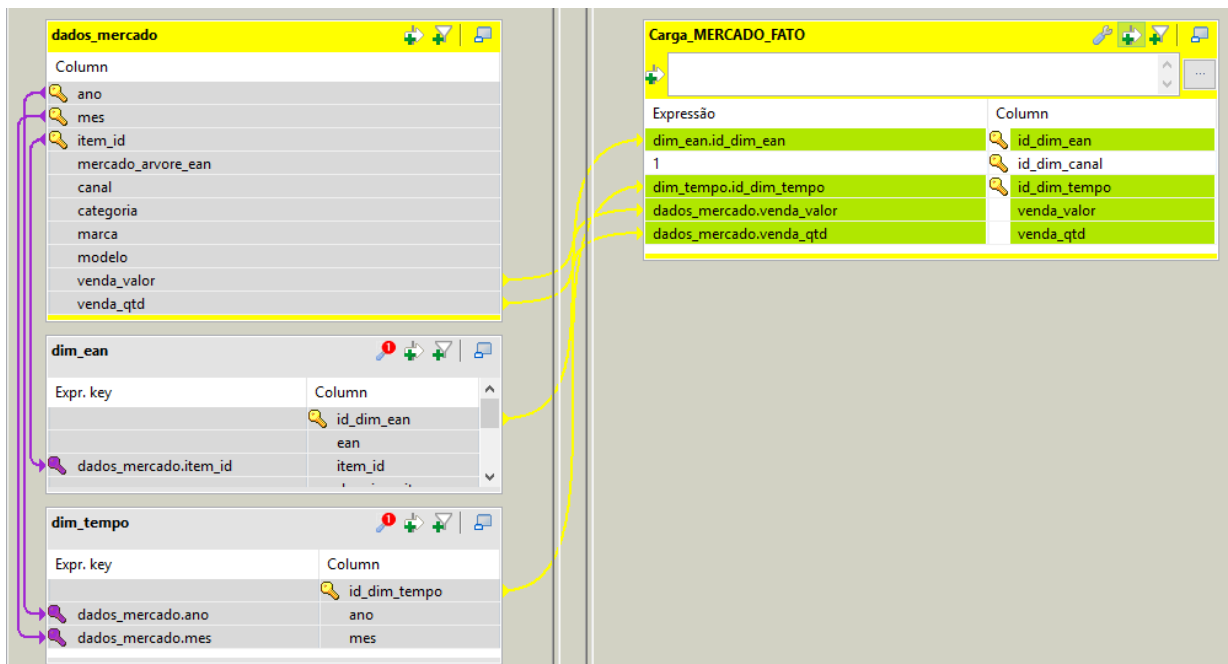


Figura 14 - TOS, tMap carga dos dados de mercado na tabela Fato

Para a carga dos dados internos foi realizado um *join* entre as tabelas “dados_internos” e “dim_tempo”, utilizando o campo “mes” e colocando como default o ano vigente, para buscar o id da dimensão “tempo”, também foi realizado um *join* entre as tabelas “dados_internos” e “dim_ean” utilizando o campo “ean” como chave para buscar o id da dimensão “ean”. Para a carga dos dados internos foi colocado como default o valor “2” para a dimensão “canal”, que representa a opção “dados internos”, todas as conexões podem ser vistas na Figura 15.

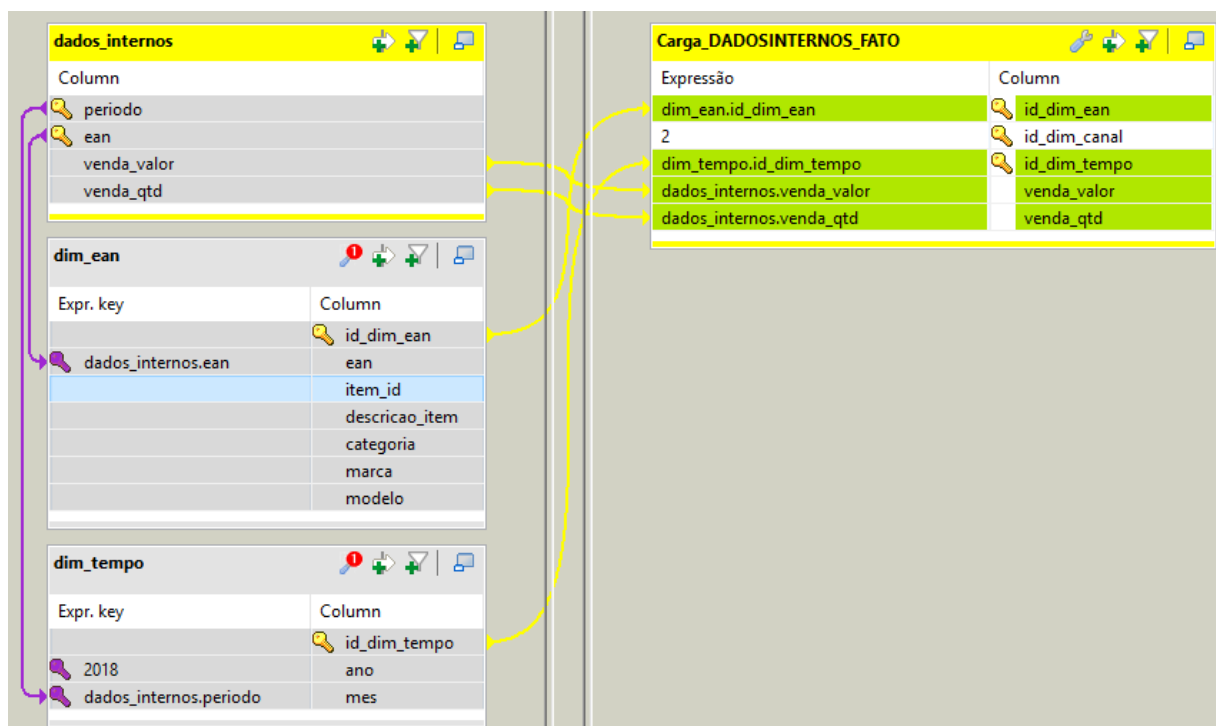


Figura 15 - TOS, tMap carga dos dados internos na tabela FATO

Afim de facilitar todo o processo de carga dos dados foi elaborado um pacote, conforme Figura 16, para realizar a chamada dos pacotes de carga de forma sequencial e automática.

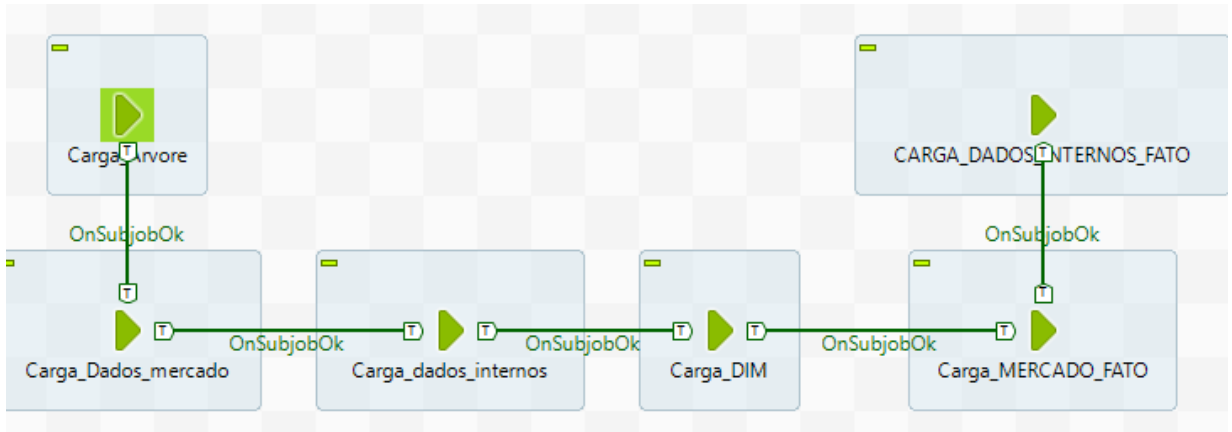


Figura 16 - TOS, pacote ETL

4. Dashboard operacional

Para facilitar a consulta e melhorar a visualização dos dados foi construído um dashboard, para agilizar a consulta dos dados também foi criado um data mart (DM) onde ficará armazenado todos os dados que alimentam o dashboard.

Como os dados vão ser analisados pela equipe de inteligência de mercado da empresa, foi criado um data mart com o nome “DM_IM” que contém alguns campos conforme pode ser visto na Figura 17.

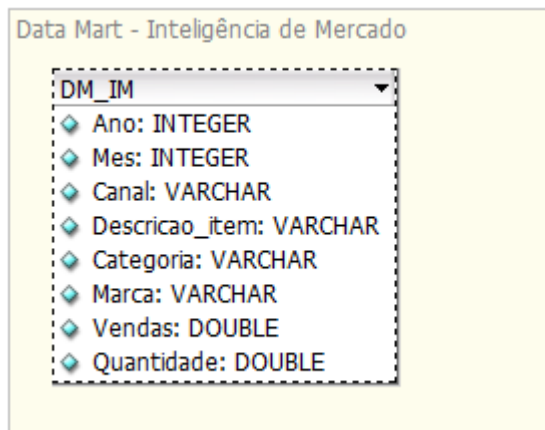


Figura 17 - Data Mart da equipe de inteligência de mercado

Para realizar a atualização dos dados no DM de forma automática, foi elaborado um pacote na ferramenta TOS conforme pode ser visto na Figura 18.

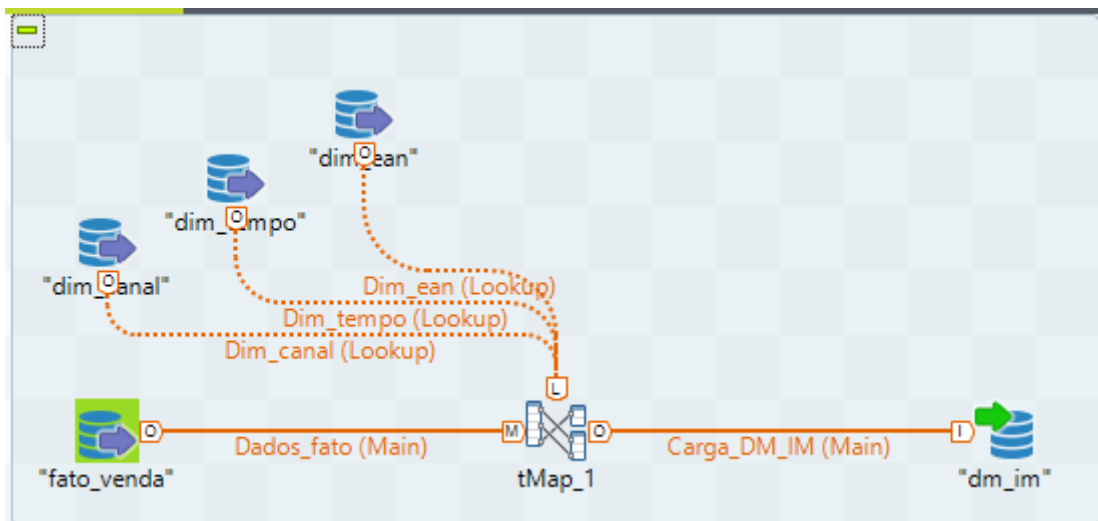


Figura 18 – TOS, pacote para atualização dos dados no DM.

A partir do “DM_IM” foi possível a criação de um Dashboard que contém alguns gráficos e tabelas para auxiliar nos processos de tomada de decisão da empresa. Para a elaboração do dashboard foi utilizado a ferramenta Qlikview.

O QlikView é uma plataforma que fornece o *self-service* BI, o que possibilita todos os usuários da organização acessar e trabalhar com os dados, levando os usuários a descoberta e *insights* de dados e relacionamento entre várias fontes. O QlikView compacta os dados e os mantém na memória, onde estão disponíveis para exploração imediata por vários usuários. Para conjunto de dados muito grande o QlikView se conecta diretamente com a fonte de dados (Qlik).

Para fazer a conexão do Qlikview com o SGBD (sistema de gerenciamento de banco de dados) foi necessário instalar o driver do banco de dados desejado no ODBC, que é o administrador de fontes de dados do sistema operacional.

5. Resultados Obtidos

Para auxiliar os diretores e gerentes a entender melhor todo o comportamento da empresa foi desenvolvido um dashboard operacional onde os dados internos podem ser facilmente comparados com a informação passada pelo mercado.

Na Figura 19 pode ser visto o dashboard que foi elaborado para utilizar os dados do “DM_IM”, o principal objetivo deste dashboard é facilitar a visualização das vendas internas por categorias em relação ao mercado, para isso no campo superior esquerdo foi colocado uma lista com todas as categorias que a empresa tem venda, ao selecionar qualquer item desta lista os dois gráficos, a tabela e os *cards* são atualizadas.

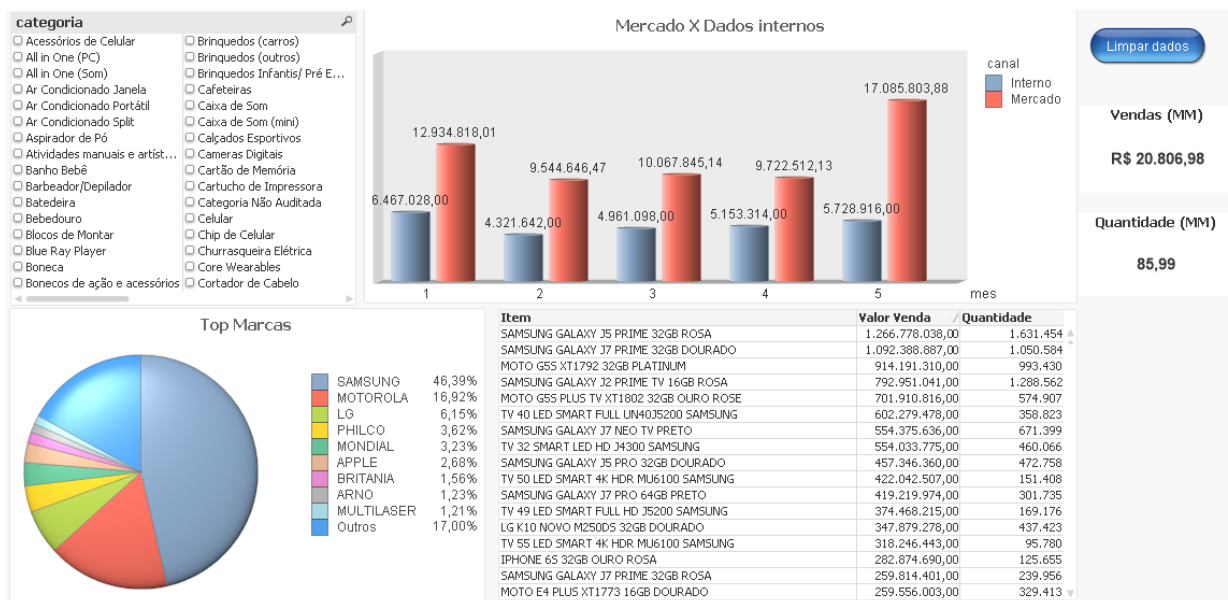


Figura 19 - Dashboard IM

No gráfico em barras (Figura 20) “Mercado X Dados internos” pode ser visto o total de vendas por mês da empresa e do mercado, o que facilita a identificação de grandes *gaps*, onde o mercado teria uma venda significativa e a empresa não.

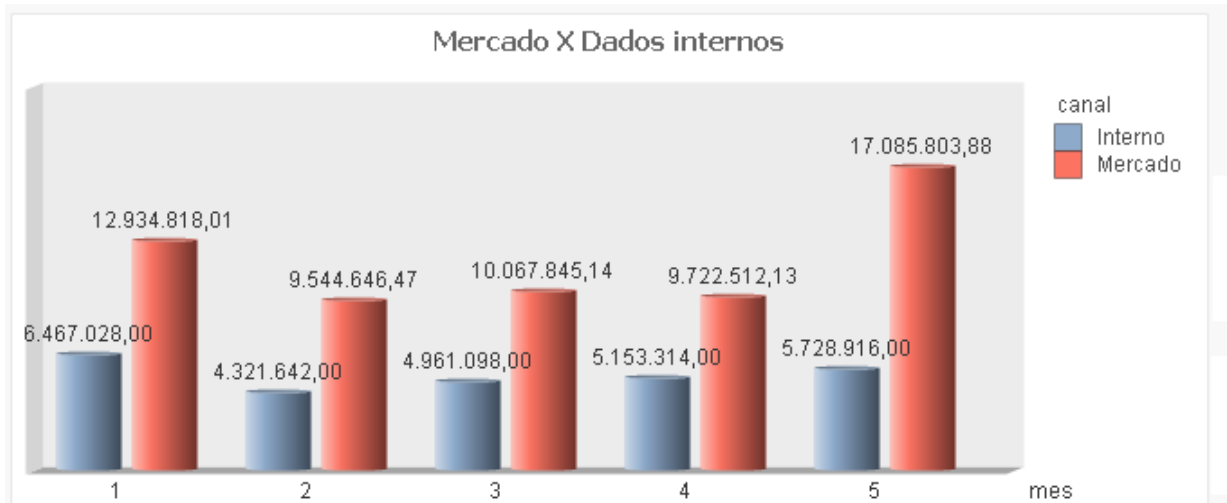


Figura 20 - Gráfico "Mercado X Dados internos"

No canto superior direito pode ser observado um botão para limpar todos os filtros realizado, voltando o dashboard para o formato original mostrando todos os dados. Também tem dois *cards* onde apresentam o valor e quantidade total dos itens vendidos no mercado.

O gráfico de pizza (Figura 21) "Top Marcas" apresenta a participação das marcas que são mais vendidas no mercado, possibilitando a empresa buscar novos contratos com diferentes marcas que apresentam bom desempenho de venda no mercado.

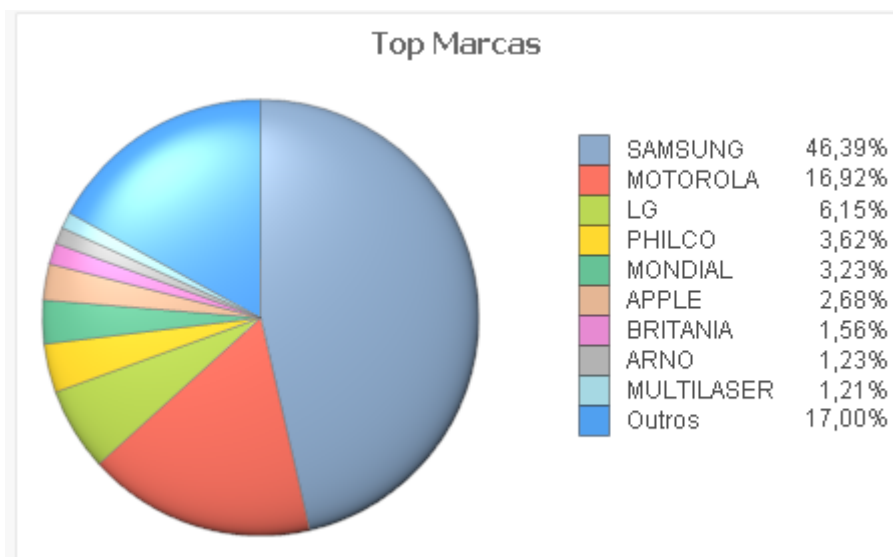


Figura 21- Gráfico "Top Marcas"

A tabela do canto inferior direito vai contra as boas praticas da elaboração de um dashboard pois a mesma traz uma informação com muito texto, poluindo visualmente o dashboard, porém nesta tabela temos uma abertura por item vendido

no mercado que é um dado muito importante para a empresa, pois possibilita acompanhar e entender quais itens realmente são expressivos em vendas e quais não são, interferindo diretamente em novas compras.

Ao realizar uma análise na categoria de Smartphone (Figura 22), pode ser notado alguns pontos.

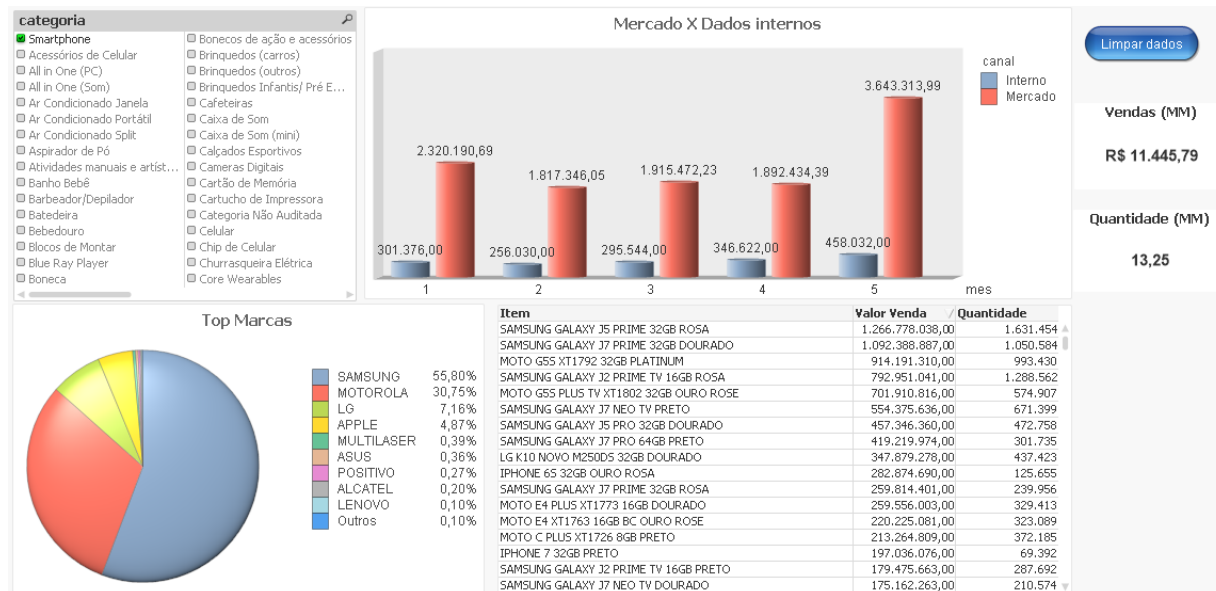


Figura 22 - Dashboard, Categoria Smartphone

- A venda de Smartphones da empresa é muito baixa em relação ao mercado, podendo ser um bom ponto de investimento já que o mercado tem uma venda bem expressiva.

- As marcas Samsung e Motorola tem mais de 80% das vendas em todo o mercado, sendo os smartphones mais vendidos (conforme a tabela de itens da Figura 22)

Com a informação apresentada no dashboard da Figura 22 os gerentes de departamentos responsáveis pela venda de smartphone podem ver os pontos onde a empresa não está alinhada com o mercado e também obter informação dos itens que estão sendo mais comprados. Assim como foi feito para a categoria Smartphone também é possível realizar uma mesma análise para quaisquer outras categorias de forma simples.

6. Conclusões

Neste projeto foram utilizadas, em grande maioria, ferramentas Open Source pela facilidade de trabalhar com ferramentas que não precisam de licença, porém todo o projeto pode ser facilmente adaptado a outras ferramentas, cabe a empresa definir quais ferramentas irá utilizar e realizar a adaptação.

É comum empresas e organizações pensarem no BI como a ferramenta que vai resolver e otimizar todos os problemas, e cada vez mais querem implementar aplicações de Machine Learning, previsão de séries temporais, entre outros tópicos que estão auxiliando nas tomadas de decisão e ganhando espaço no mundo corporativo. Porém geralmente as empresas trabalham com dados relacionados sem nenhuma estrutura ou organização dos mesmos, o que dificulta a consulta e análise dos dados, tornando complexo a extração de informações relevantes, por este motivo uma empresa ao pensar em BI primeiro deveria investir em estruturar os dados para facilitar a consulta e análise.

O objetivo deste projeto foi desenvolver toda a metodologia necessária para a empresa elaborar o seu próprio DW e a construção do dashboard, um grande facilitador para este projeto foi a obtenção dos dados tanto internos quanto os dados de mercado, através do dashboard foi possível obter bons *insight*, porém este projeto se limita principalmente na transformação e carga dos dados deixando a parte de aprendizagem de máquinas para próximos trabalhos.

Com toda a estrutura das tabelas e o processo de atualização dos dados do Data Warehouse e o Data Mart montados, realizar a atualização do dashboard se torna muito simples. Com isso a empresa pode focar mais os seus recursos nos processos de análise de dados.

6.1. Trabalhos futuros

Para melhorar a informação recebida pelo mercado pode-se fazer novas parcerias com empresas que divulgam informações do mercado, neste estudo focamos nos dados somente de uma empresa que disponibiliza informação do departamento de eletrodomésticos, o mesmo estudo pode ser ampliado buscando novos dados do mercado e também para outros departamentos afim de buscar um melhor entendimento do mercado como um todo.

Conforme mencionado no capítulo 2, é possível criar um processo de análise de dados online (OLAP), como o DW é uma estrutura multidimensional pode ser montado um Cubo de dados facilitando ainda mais a interação dos usuários com os dados.

A mineração de dados pode acrescentar diversos indicadores importantes para os gestores, por exemplo com a utilização de algoritmos de mineração de dados talvez seja possível encontrar padrões entre itens e até entre categorias distintas.

Com o acúmulo de dados poderá ser observado séries temporais a nível de produtos, o que possibilita utilizar modelos para realizar um trabalho de previsão, auxiliando as empresas varejistas na tomada de decisão em relação a compra e estoque de produtos para determinadas épocas do ano.

Referências Bibliográficas

FELIPE BORGES DE AMORIM, D. Um pouco de história sobre o mercado. 2015. Artigo publicado em:

<http://www.administradores.com.br/mobile/artigos/academico/um-pouco-de-historia-sobre-o-mercado-varejista/91238/>

ALVES, A. Modelo de dados: Transacional x Dimensional. 2011. Artigo publicado em: <http://all-carvalho.blogspot.com/2011/04/modelo-de-dados-transacional-.html>

ELIAS, D. Conhecendo a arquitetura de Data Warehouse. 2014. Artigo publicado em: <https://canaltech.com.br/business-intelligence/conhecendo-a-arquitetura-de-data-warehouse-19266/>

LEMIRE, D. Data Warehousing and OLAP. 2010. Paper publicado em: <http://www.daniel-lemire.com/OLAP/>

TALEND, Acesso em 3 de novembro de 2018, <https://www.talend.com/>

QLIK, Acesso em 3 de novembro de 2018, <https://help.qlik.com/en-US/qlikview/November2017/Content/what-is.htm>

Anexos

Anexo 1 – Estrutura das tabelas de dados

MERCAOD_ITEM	Campo	Descrição	Tipo
	Ano	Ano de referencia dos dados enviados pelo mercado	int
	Mês	Mês de referencia dos dados enviados pelo mercado	int
	Canal	Canal de venda do item (alguns itens não tem abertura do físico e online)	var(100)
	Categoria	Informação sobre qual grupo o item pertence	var(100)
	ID	Código de identificação do fornecedor	int
	Marca	Marca do fabricante do item	var(100)
	Modelo	Modelo do item	var(100)
	Vendas Valor	Valor total das vendas do item	double
Vendas Volume	Quantidade de itens vendidos	double	

ARVORE	Campo	Descrição	Tipo
	Mês	Mês de referencia dos dados enviados pelo mercado	data
	EAN	Código de barra do item	int
	DOMAIN_PRODUCTGROUP_TEXT	Informação sobre qual grupo o item pertence (em inglês)	var(100)
	MAIN_TEXT	Descrição do item	var(255)
	ITEM_ID	Código de identificação do fornecedor	int
	Categoria Traduzida	Informação sobre qual grupo o item pertence	var(100)
	Marcas	Marca do fabricante do item	var(100)

Dados internos	Campo	Descrição	Tipo
	Periodo	Mês de referencia dos dados internos	int
	EAN	Código de barra do item	int
	Valor venda	Valor total das vendas do item	double
Quantidade Venda	Quantidade de itens vendidos	double	