



V Encontro de Iniciação Científica e Tecnológica  
V EnICT  
ISSN: 2526-6772  
IFSP – Câmpus Araraquara  
22 e 23 de outubro de 2020



## **Análise de Dados utilizando Ferramentas de *Business Intelligence* aplicadas a *Data Warehouse* sobre Vendas de Varejo de Pet Shop**

NATIELLY CRISTINA FRANCISCO <sup>1</sup>, CRISTIANE AKEMI YAGUINUMA <sup>2</sup>, JOSÉ RODOLFO BELUZO <sup>3</sup>, TALES BOALIM<sup>4</sup>:

<sup>1</sup>Discente do Técnico em Informática Concomitante, Subsequente, Bolsista Fundação, IFSP Câmpus Araraquara, [natielly.c@aluno.ifsp.edu.br](mailto:natielly.c@aluno.ifsp.edu.br).

<sup>2,3</sup>Docente do Instituto Federal de São Paulo – Câmpus Araraquara, <sup>2</sup>[cristiane.yaguinuma@ifsp.edu.br](mailto:cristiane.yaguinuma@ifsp.edu.br), <sup>3</sup>[jrbeluzo@ifsp.edu.br](mailto:jrbeluzo@ifsp.edu.br)

<sup>4</sup>Gerente de Tecnologias e Inovação, JN Moura Informática, [tales@rfidmoura.com.br](mailto:tales@rfidmoura.com.br) <sup>4</sup>

**Área de conhecimento** (Tabela CNPq): Banco de Dados – 1.03.03.03-0

**RESUMO:** Nos dias atuais as informações são um dos principais componentes responsáveis pelo desenvolvimento empresarial. Sendo assim a utilização de ferramentas de Business Intelligence passam a ser essenciais quando utilizados sistemas *Enterprise Resource Planning* (ERP). A partir do momento em que se obtém diversas informações, as mesmas necessitam ser estruturadas e contextualizadas permitindo análises e previsões de tendências em meios corporativos. Tendo em vista tais necessidades, a empresa JN Moura Informática, em parceria com o Instituto Federal, câmpus Araraquara, instituiu um projeto de pesquisa intitulado “Construção de *Data Warehouse* para Integração, Análise e Mineração de Dados” considerando o desenvolvimento de inovações aos softwares já comercializados pela empresa. Sendo assim o objetivo do projeto foi integrar dados transacionais em um *Data Warehouse* (DW) visando a necessidade de integração e estruturação dos dados para geração de relatórios gerenciais. Inserido nesse projeto mais geral, especificamente este projeto de iniciação científica visou investigar técnicas e ferramentas de *Business Intelligence* para visualização gráfica dos dados pertencentes ao *Data Warehouse*, permitindo análises de tendências de mercado.

**PALAVRAS-CHAVE:** Banco de dados, Análise de dados, *Data Warehouse*, *Business Intelligence*.

### **INTRODUÇÃO**

Com o desenvolvimento de novas tecnologias as informações se ampliam aceleradamente, gerando um grande volume de dados. A partir de análises e contextualização dos dados através de ferramentas de *Business Intelligence* (BI), as informações passam a ser visualizadas por diversas perspectivas tornando-se importante para devidas decisões em ambiente empresarial. O termo *Business Intelligence* refere-se livremente à variedade de recursos fornecidos aos usuários de negócio para alavancar a área de apresentação dos dados para tomadas de decisão analítica (KIMBALL; ROSS, 2013).

A partir da integração dos dados em *Data Warehouse* (DW), é necessário o desenvolvimento de uma composição lógica capaz de permitir análises através de diversos pontos de vista. Tal conceito é denominado como *Online Analytical Processing* (OLAP), responsável por estruturar os dados de maneira multidimensional, possibilitando análises de diversas perspectivas apresentadas graficamente através de uma ferramenta de *Business Intelligence* (BI) (KIMBALL; ROSS, 2013).

A *software house* JN Moura Informática, parceira deste projeto, oferece para seus clientes um sistema de gerenciamento e automação comercial (ERP) que resulta uma quantidade expressiva de dados para análises e gerenciamento de relatórios. Entretanto, esses registros armazenados em banco de dados relacionais não estão integrados de modo a atender a demanda por relatórios e análises gerenciais. Tendo em vista tais necessidades, um dos objetivos do projeto de pesquisa interinstitucional intitulado “Construção de *Data Warehouse* para Integração, Análise e Mineração de Dados” foi integrar os dados do sistema ERP em um *Data Warehouse* (DW). Especificamente, este projeto de iniciação científica tem como objetivo investigar técnicas e ferramentas de *Business Intelligence* para visualização gráfica dos dados pertencentes ao *Data Warehouse* desenvolvido no projeto interinstitucional, permitindo análises de tendências de mercado.

## FUNDAMENTAÇÃO TEÓRICA

Na literatura (KIMBALL; ROSS, 2013), a abordagem clássica para integração de dados denominada *Data Warehouse* (DW) é responsável pelo armazenamento de informações relacionadas a uma ou mais empresas, possibilitando análises mais precisas para a geração de relatórios. Em um DW os dados são dispostos a uma série de processos responsáveis pela extração, transformação e carregamento dos registros em um repositório único (Processo ETL- *Extract, Transform, Load*). Dessa forma as consultas podem ser realizadas diretamente ao DW, havendo a otimização dos processos para análises.

A estrutura dimensional de um DW é composta por uma tabela de fatos e algumas dimensões (KIMBALL; ROSS, 2013). Os fatos são responsáveis pelo armazenamento de eventos relacionados à empresa, já as dimensões são tabelas complementares ao fato que armazenam o contexto relacionado ao evento. Para exemplificar, considere o fato como uma venda e as dimensões como o contexto que envolve a venda: os produtos vendidos, a data e horário da venda, o local, entre diversos outros atributos associados ao fato ocorrido.

A partir da definição e estruturação do DW é necessário constituir a composição lógica dos dados denominada como cubos de processamento analítico online (*Online Analytical Processing* - OLAP). O conceito OLAP inclui a noção ou ideia de múltiplas dimensões hierárquicas (THOMSEN, 2002) que possibilita a visão notória de determinados acontecimentos a partir do processamento de dados alocados no DW. Com a definição do cubo os dados são dispostos a diferentes granularidades podendo ser definidas como *drill down* (aumento de detalhamento dos dados) ou *drill up* (aumento da granularidade, ou seja, menor detalhamento) (THOMSEN, 2002). Com desenvolvimento do cubo os dados passam a ser visualizados de maneira multidimensional possibilitando análises de diversas perspectivas quando aplicado a ferramentas de visualização.

Apesar da estruturação de informação e conteúdos capazes de apoiar as tomadas de decisões empresariais, é necessária a visualização por meio de gráficos, tendo em vista a produtividade superior de análises por meio de componentes gráficos ao invés de numéricos. Sendo assim após a estruturação física (DW) e lógica (OLAP) dos dados, os mesmos são dispostos a uma ferramenta de visualização, permitindo que os registros sejam visualizados graficamente.

A partir do DW e do cubo OLAP, pode ser aplicado um conjunto de técnicas e ferramentas para auxiliar na administração de um negócio e nas tomadas de decisões, o que consiste no conceito de *Business Intelligence* (SANTOS, 2009 apud BOTELHO; FILHO, 2014).

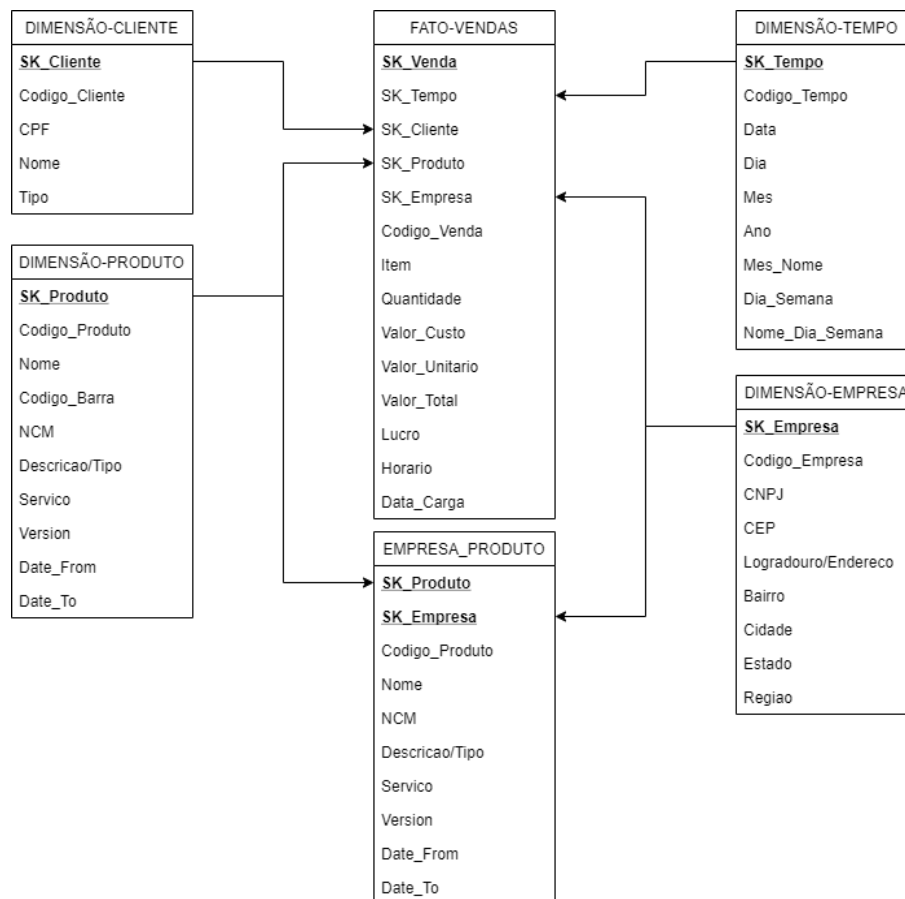
## METODOLOGIA

Um dos resultados do projeto interinstitucional foi o desenvolvimento de um DW, que foi realizado em colaboração com outra bolsista de iniciação científica que tratou especificamente do processo ETL. No DW, os fatos representam as vendas e as dimensões os principais elementos ocorrentes na realização do fato. Para a definição da arquitetura do DW foram aplicadas algumas

análises para designar as principais informações relacionadas ao fato. Essas informações representam as dimensões e métricas que constituem a arquitetura do DW através das seguintes tabelas:

- Tabela de fatos: Contém todas as vendas realizadas e informações relacionadas a mesma, estando interligada com todas as dimensões desenvolvidas.
- Dimensão Cliente: Possui os clientes envolvidos em uma determinada venda.
- Dimensão Empresa: Deve conter as empresas responsáveis pela realização de vendas.
- Dimensão Produto: Deve conter todos os produtos vendidos.
- Empresa Produto: Tabela auxiliar da dimensão produto para manter a rastreabilidade dos dados originais.
- Dimensão Tempo: Armazena as datas relacionada a vendas.

A Figura 1 representa o modelo do DW desenvolvido segundo o esquema floco de neve, onde todas as dimensões estão associadas a tabelas de fatos e a dimensão produto e empresa estão vinculadas também a tabela empresa produto.



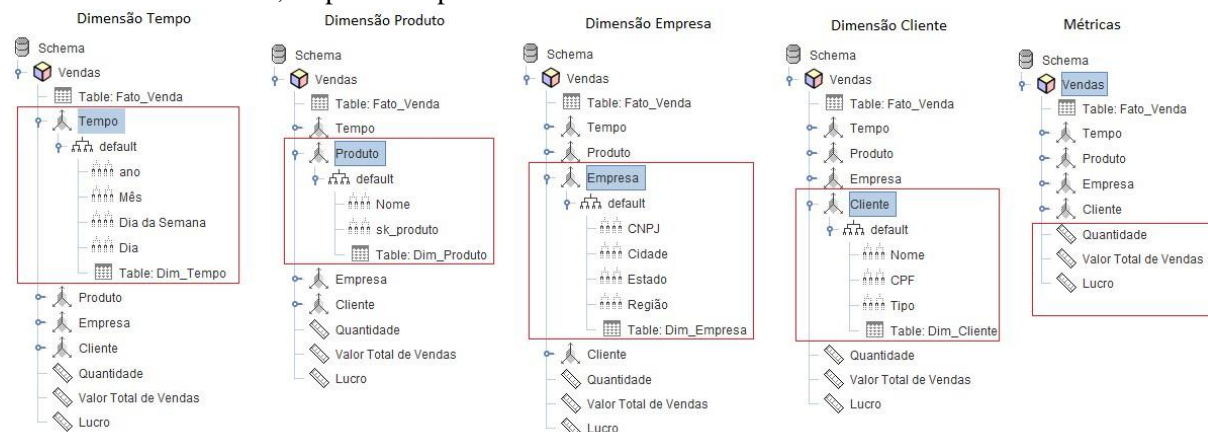
**FIGURA 1. Esquema do DW desenvolvido no projeto interinstitucional. Fonte: próprio autor.**

Por seguinte, após a conclusão do DW, um representante da empresa Moura Informática requisitou algumas consultas para análises, que auxiliam nas tomadas de decisões empresariais. A partir da definição das consultas e requisitos, este projeto de iniciação científica foi responsável por realizar estudos sobre ferramentas de *Business Intelligence* (BI) capazes de suprir as necessidades especificadas.

Com a definição dos requisitos e ferramentas de BI capazes de suprir as necessidades do projeto, foi realizado o desenvolvimento lógico dos dados (OLAP) através de uma ferramenta integrada a suíte *Pentaho BI* chamada *Pentaho Schema Workbench*.<sup>1</sup>

<sup>1</sup> [https://help.pentaho.com/Documentation/8.2/Products/Schema\\_Workbench](https://help.pentaho.com/Documentation/8.2/Products/Schema_Workbench)

A Figura 2 possui a representação lógica (OLAP) dos dados alocados no DW, desenvolvido na ferramenta *Pentaho Schema Workbench*. A definição do cubo OLAP consiste na associação do mesmo com a tabela de Fato e suas respectivas dimensões e métricas a serem analisadas. Na Figura 2, cada componente destacado em vermelho representa atributos a serem analisados de forma hierárquica, conforme a dimensão especificada no topo do quadrante. Todas as dimensões seguem a mesma estrutura, modificando somente os atributos. O último quadrante da Figura 2 representa algumas métricas do fato venda, responsável pelo desenvolvimento de análises.



**FIGURA 2. Estrutura lógica OLAP do DW de vendas.**

**Fonte: próprio autor.**

A partir da composição lógica dos dados, existe a necessidade da visualização gráfica dos registros para melhores análises e tomadas de decisões. Sendo assim o processo de conversão dos dados numéricos para gráficos foi realizado por meio da ferramenta *Pentaho User Console*<sup>2</sup> que, devido alguns *plugins*, disponibiliza um ambiente flexível para o desenvolvimento.

Durante o projeto foi utilizado o *plugin CDE Dashboard*<sup>3</sup> disponibilizado gratuitamente pela comunidade *Pentaho*, que por meio de três componentes principais disponibiliza um ambiente com diversas funcionalidades e métodos para visualização gráfica dos dados. A partir do código SQL relacionado a cada consulta e inserção das mesmas no ambiente *CDE Dashboards*, os dados passaram a ser visualizados de forma numérica para visualização gráfica.

## RESULTADOS E DISCUSSÃO

Os principais resultados obtidos por meio deste projeto são: a construção do DW em colaboração com outro projeto de iniciação científica, a elaboração de cubo OLAP e a geração de relatórios a partir do DW e de ferramentas de BI, contribuindo para análises gerenciais.

Os relatórios desenvolvidos foram requisitados pelo representante da Moura Informática especialista nos dados, considerando alguns resultados esperados, com destaque para:

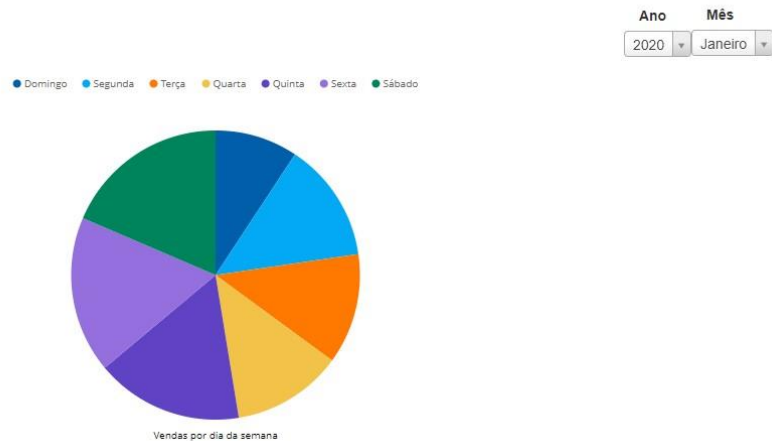
1. Porcentagem de venda por dia da semana nos últimos 30 dias;
2. Preço médio de venda de um produto por cidade, estado e região.

A primeira consulta (Figura 3) tem como finalidade retorna a quantidade de vendas em percentual por dia da semana. Através dessa análise, pode-se tomar decisões como possíveis dias onde existe a necessidade da presença de mais funcionários devido a quantidade de atendimentos. Já em outros, a diminuição dos mesmos tendo em vista o menor fluxo de pessoas ao estabelecimento. Diminuindo, assim, gastos desnecessários e o posicionamento estratégico no mercado visando o melhor atendimento os clientes.

<sup>2</sup> [https://help.pentaho.com/Documentation/8.0/Products/User\\_Console](https://help.pentaho.com/Documentation/8.0/Products/User_Console)

<sup>3</sup> [https://help.pentaho.com/Documentation/7.0/OR0/CTools/CDE\\_Dashboard\\_Overview](https://help.pentaho.com/Documentation/7.0/OR0/CTools/CDE_Dashboard_Overview)

## Porcentagem de vendas por dia da semana



**FIGURA 3. Quantidade de vendas por dia da semana. Fonte: próprio autor.**

Para a segunda consulta, existe a necessidade de visualizar os dados por um determinado tipo de produto, assim, foi criado um filtro responsável por armazenar os produtos mais vendidos, possibilitando a visualização de cada componente separadamente. Na Figura 4, observa-se o preço médio de vendas sobre a perspectiva da operação *drill down*, na qual os dados são apresentados com uma maior granularidade havendo registros sobre região, em seguida por estados e sequencialmente por cidades, que possui um maior detalhamento dos registros. Assim, o analista de dados pode realizar a comparação de preço entre seu estabelecimento e os demais concorrentes, tanto na região da empresa quanto nos demais territórios responsáveis pela venda do produto.



**FIGURA 4 - Preço médio de venda por produto com operação *drill down*. Fonte: próprio autor.**

Com a conclusão do desenvolvimento dos relatórios requisitados pelo representante da empresa Moura Informática, foi aplicado um método de avaliação para os relatórios gerados através da ferramenta *Pentaho User Console* (PUC). Para tanto, foi aplicado um questionário, disponível na plataforma Google Forms, aos especialistas de dados para avaliarem os relatórios e o acesso ao DW. As questões aplicadas consistem em análises para avaliar se os resultados desenvolvidos condizem com o esperado e se as informações geradas ajudam as tomadas de decisões no meio corporativo. As respostas das análises qualitativas sobre todas as questões apresentaram o nível mais alto de satisfação,

com destaque para as respostas que afirmam que as expectativas sobre o projeto interinstitucional foram alcançadas.

## AGRADECIMENTOS

Agradecimentos a empresa JN Moura Informática por financiar esta iniciação científica.

## CONCLUSÕES

Através deste projeto, é possível concluir que a integração dos dados e a análise de informações a partir de diferentes perspectivas é de importância para decisões estratégicas. A partir do estudo de bases de dados relacionais e fontes bibliográficas, foi desenvolvido um *Data Warehouse* e alguns relatórios requisitados pelo especialista de dados a partir de uma ferramenta de *Business Intelligence* e da elaboração de cubo OLAP.

De modo geral, os objetivos do projeto visavam entregar à empresa os dados integrados permitindo o desenvolvimento de relatórios e informações que auxiliassem as tomadas de decisões empresariais. Sendo assim, tais expectativas foram atingidas de forma satisfatória, considerando os resultados obtidos a partir da avaliação realizada pelos especialistas nos dados da empresa parceira do projeto.

Para trabalhos futuros, espera-se realizar análises sobre maiores quantidades de dados, com a integração de mais bases de dados ao DW atual. Além disso, é necessário o desenvolvimento de ferramentas que disponibilizem, para empresas clientes do sistema ERP, um ambiente com diversas análises e previsões para auxílio sobre as tomadas de decisões empresariais.

Os próximos objetivos do projeto estão sendo aplicados no ambiente corporativo da empresa JN Moura Informática, onde a partir da conclusão da iniciação científica foi contratada, com o intuito de dar continuidade ao projeto. Vale salientar que uma nova parceria entre a JN Moura e o Instituto Federal está sendo instituída visando também a continuação do projeto tendo em vista a satisfação e aprendizado entre todos os envolvidos.

## REFERÊNCIAS

BOTELHO, F. R.; FILHO, E. R. **Conceituando o Termo Business Intelligence: Origem e Principais Objetivos**. Universidade Federal do Paraná, UFPR. 2014

KIMBALL, R., ROSS, M. **The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling**. John Wiley & Sons, Inc. Canadá, 2013.

KIMBALL, R., CASERTA, J.: **The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data**. John Wiley & Sons, 2004.

PRIMAK, F.V. **Decisões com BI (Business Intelligence)**. Ciência Moderna. 2008.

SALGADO, A. C.; LÓSCIO, B. F. **Integração de Dados na Web**. Centro de Informática da Universidade Federal de Pernambuco. 2001

THONSEN, E. **OLAP Solutions: Building Multidimensional Information Systems**. 2. ed. John Wiley & Sons, Inc. Canadá, 2002.