

Extração, Transformação e Carga de Dados para Geração de Indicadores Gerenciais do Banco Alimentário de La Plata

Guilherme Pinheiro Latuque, Evaldo de Oliveira da Silva

Curso de Bacharelado em Engenharia de Software
Centro de Ensino Superior de Juiz de Fora (CESJF) – Campus Academia
36016-000 – Juiz de Fora, MG – Brasil

guilherme.latuque@gmail.com, evaldo.oliveira@gmail.com

Abstract. *The objective of this article is to present a case study in a NGO, using theoretical and practical concepts involving ETL (Extract, Transform, Load) which is a fundamental part of the BI process used by many companies today. In addition, an indicator implemented in the generic application process will be used as an example for ease of understanding. The NGO used as a case study, called the Food Bank of La Plata, signed a cooperation agreement with the CESJF for the creation of managerial indicators.*

Resumo. *O objetivo deste artigo é apresentar um estudo de caso em uma ONG, utilizando-se de conceitos teóricos e práticos envolvendo ETL (Extract, Transform, Load) que é uma parte fundamental do processo de BI (Business Intelligence) utilizando por muitas empresas atualmente. Além disso será utilizado como exemplo um indicador implementado no processo da aplicação de forma genérica para facilidade no entendimento. A ONG utilizada como estudo de caso trabalho, chamada de Banco Alimentário de La Plata, assinou um acordo de cooperação com o CESJF para a criação dos indicadores gerenciais.*

1. Introdução

Atualmente, para diferentes tipos de organizações é geralmente importante estar à frente de seus concorrentes, visando a manutenção de seus mercados e até mesmo a melhoria dos processos internos de negócios visando alcançar melhores resultados tanto para seus clientes e usuários. Neste contexto, é necessário que estas organizações se dediquem não somente a questões administrativas, transporte ou marketing, por exemplo, mas também à implementação de tecnologias que sustentem a melhoria dos processos.

O uso de tecnologia visa facilitar a integração de processos e de dados produzidos por estes processos, envolvendo investimento também em infraestrutura e soluções de dados em nuvem. Além desses investimentos é necessário um sistema integrado com informações e dados históricos validados, íntegros e atualizados, para servir de componentes importantes para o banco de informações da empresa.

Tendo em vista essa fonte de dados atualizada e alimentada ao longo tempo a partir de processos da organização, é possível que o CEO (*Chief Executive Officer*) da empresa possa tomar decisões vitais para o futuro de sua organização. Nesse contexto o papel da T.I. (Tecnologia da Informação) traz para as empresas dentre várias outras soluções a necessidade de implantação de BI (*Business Intelligence*) que é capaz de

transformar o registro que é armazenado no banco de dados conhecido como dado, em extensas fontes de informações e conhecimento que servem de visão para o CEO no qual poderá ser capaz de visualizar os dados de sua empresa de forma gráfica por meio de painéis de controle (*dashboards*) de forma dinâmica e atualizada (BRAGHITTONI, 2017).

Portanto, neste trabalho será demonstrado um estudo de caso da implementação de um projeto de BI utilizando-se da base de dados e indicadores gerenciais da ONG (Organização Não Governamental) Banco Alimentário de La Plata, situada na Argentina. O Banco Alimentário de La Plata assinou um acordo de cooperação de fins acadêmicos com o CESJF (Centro de Ensino Superior de Juiz de Fora) e que visa a implementação de indicadores de produtividade. O acordo de cooperação entre as duas instituições, permitiu o registro de um projeto de iniciação científica e registrado no Centro de Extensão e Pesquisa do CESJF com a participação dos autores deste trabalho.

O restante do texto se encontra organizado pelas seções descritas a seguir. A Seção 2 é dedicada ao referencial teórico e definição dos conceitos e técnicas utilizadas. Ao Seção 3 descreve os indicadores solicitados e especificados pela ONG. A Seção 4 se relaciona com a implementação dos indicadores informados utilizando técnicas de extração, transformação e carga de dados. Finalmente, A Seção 5 descreve as considerações finais e os trabalhos futuros.

2. Referencial Teórico

Esta seção apresenta os principais conteúdos que foram utilizadas para elaboração do trabalho. Nele serão abordados conceitos relacionados ao Banco Alimentário de La Plata, BI, ETL e modelagem dimensional de dados.

2.1. O Banco Alimentário de La Plata

O Banco Alimentário de La Plata é uma Organização da Sociedade Civil (OSC), almeja reduzir a fome e a desnutrição através da recuperação de alimentos. Criada no ano 2000 como o primeiro Banco de Alimentos do país. São sócios fundadores da Rede Argentina de Bancos Alimentares, que reúne 16 outros bancos estabelecidos no país e dois em formação. Preza pela defesa do Direito Humano a uma dieta saudável, através dos esforços de seus colaboradores, voluntários e da solidariedade de empresários, produtores e doadores. (BANCO ALIMENTÁRIO DE LA PLATA, 2019).

Tem como missão, reduzir a fome, a desnutrição e as más práticas alimentares na região, através da recuperação de alimentos, para serem distribuídas a organizações comunitárias que prestam serviços de alimentação a setores carentes, desenvolvendo ações conjuntas com a sociedade, baseadas em nossos valores e capacidades. Visão de uma sociedade sem fome, nutrida, com consciência socioambiental, onde o alimento adequado ao consumo não é desperdiçado. Além de valores como: compromisso social, solidariedade, responsabilidade, transparência, educação, respeito pelo próximo e conscientização de parceiros ambientais. (BANCO ALIMENTÁRIO DE LA PLATA, 2019). A Figura 1 exemplifica através de etapas como é feito o encaminhamento dos alimentos.



Figura 1. Processo de distribuição de alimentos. (BANCO ALIMENTÁRIO DE LA PLATA, 2019).

- Etapa 1: Resgate e recuperação de alimentos
- Etapa 2: Chegada da comida ao depósito para pesagem e classificação através de uma balança
- Etapa 3: Distribuição de alimentos recuperados / doados a instituições beneficiárias (cantinas, copos de leite, lares, etc.)

2.2. Business Intelligence

Segundo o autor BRAGHITTONI (p.1, 2017):

“Business Intelligence (ou BI) é um termo cunhado por Howard Dresner do Gartner Group, em 1989, para descrever um conjunto de conceitos e métodos para melhorar o processo de tomada de decisão das empresas, utilizando-se de sistemas fundamentados em fatos e dimensões. O BI baseia-se em agrupar informações de diversas fontes e apresentá-las de forma unificada e sob uma métrica comum, a fim de que indicadores aparentemente distantes possam fazer sentido entre si”.

Segundo o autor Bonel (p.14, 2017) é de extrema importância, para se começar a definir Business Intelligence, saber a diferença entre dado, informação e conhecimento.

- Dado: O dado é forma bruta, a transação, o grão, o nível folha, ou seja, o detalhe. Exemplo: caneta azul ponta média, caneta transparente ponta fina.
- Informação: A informação é a forma inteligente, consolidada e precisa do dado para existir. Exemplo: 3 canetas azul vendidas, 2 canetas de ponta fina vendidas.
- Conhecimento: Já o conhecimento é tudo aquilo que a informação te gerará de insights, ideias e sugestões. Exemplo: Portando foram vendidas mais canetas da cor azul de ponta média em relação a canetas transparentes de ponta fina.

A Tabela 1 exemplifica esses conceitos apresentados.

Tabela 1. Diferenças entre dado, informação e conhecimento. (BONEL, 2017).

Dado	Informação	Conhecimento
Detalhe	Consolidado	Consolidado

Não apoia a decisão	Apoia a decisão	Provê insights e ideias
Operação	Gestão	Gestão
É criado a partir de uma transação	É criado a partir de um conjunto de dados	É criado a partir de um conjunto de informações

A Figura 2 exemplifica a estrutura conceitual de uma plataforma em BI e em seguida, é apresentada suas etapas.

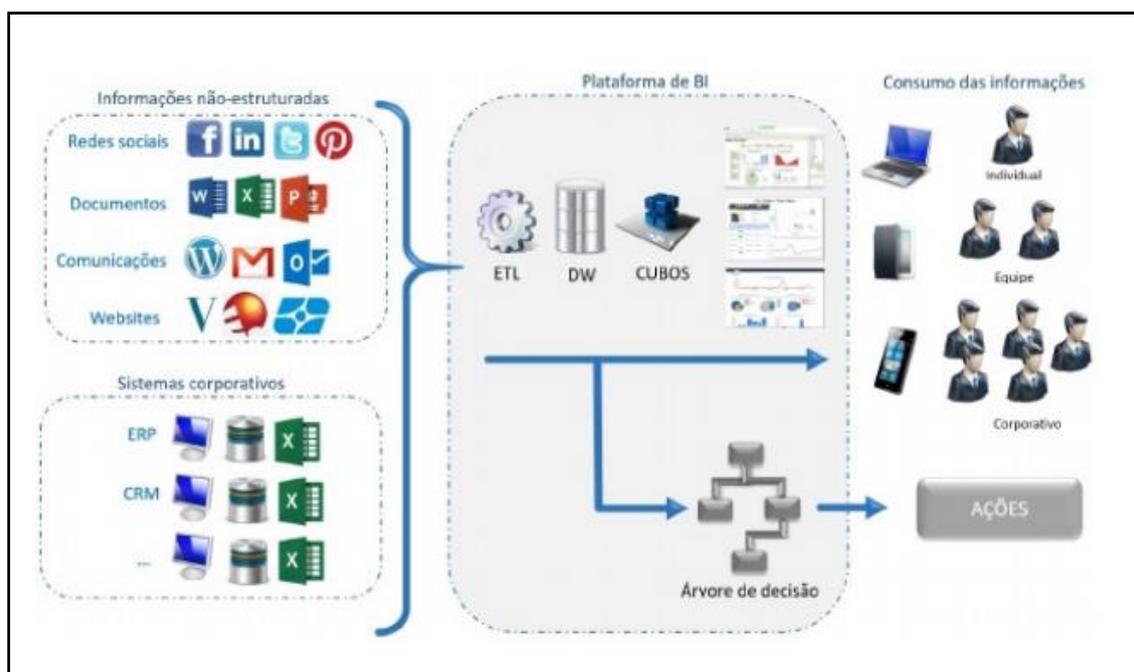


Figura 2. Estrutura da plataforma de BI. (BRAGHITTONI, 2017).

- **Informações não-estruturadas:** são informações não tabuladas. Podem ser e-mails, documentos ou dados espalhados por fontes externas. Por tratar-se de dados não estruturados, ou seja, com informações textuais sem um posicionamento predefinido, o uso dessas fontes requer maior experiência na área.
- **Sistemas corporativos:** são sistemas ERP, CRM, etc. A captura dessas informações pode ser feita por acesso direto ao banco de dados dela, por meio de serviços ou APIs, ou ainda pela extração de dados em planilhas ou arquivos de texto. Ainda assim, são dados estruturados e, geralmente, de obtenção mais simples.
- **Plataforma de BI:** constituída pelo processo de ETL, DW (*Data Warehouse*), Cubos e conseqüentemente geração de painéis de controle e facilitadores para tomadas decisão dentro da organização.
- **Consumo das informações:** usuários finais nos quais podem consumir as informações geradas no processo e tomar medidas necessárias, sendo eles uma pessoa, uma equipe ou uma organização.

2.3. Processo de Extração, Transformação e Carga de Dados

Segundo Bonel (p.24, 2017) todavia, o termo extenso caiu em desuso e, hoje em dia, o ETL permeia o cotidiano das organizações quando o assunto é transportar dados de um ponto a outro. Tem-se uma série de ferramentas específicas para o ETL, como PowerCenter, Pentaho Data Integration, SQL Integration Services, e tantos outros. A Figura 3 exemplifica o processo interno do ETL que ocorre dentro da plataforma de BI.

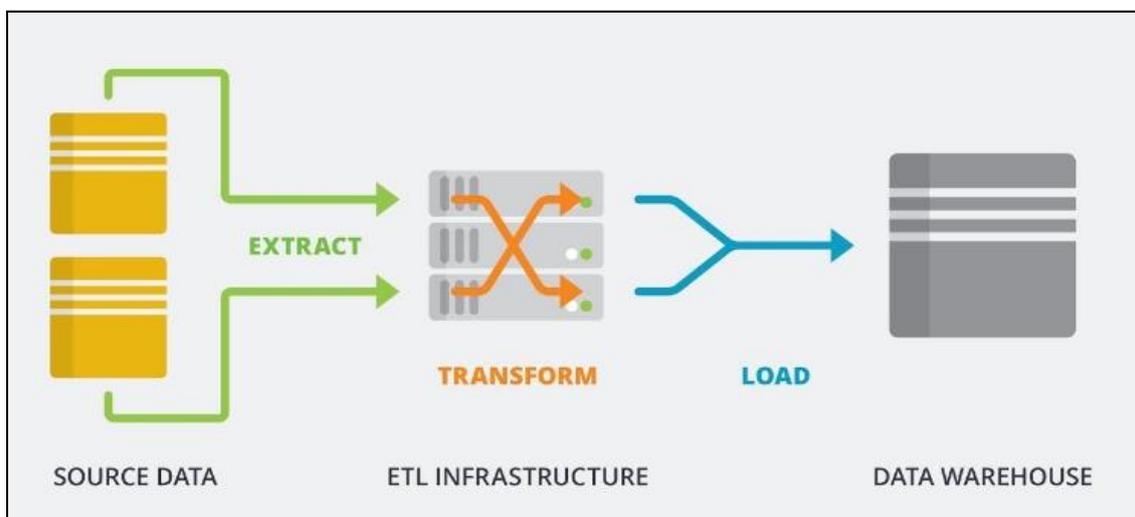


Figura 3. Processo de ETL (MATILLION, 2019).

- **Extração:** A primeira etapa na obtenção de dados para o ambiente do DW é a extração. Essa etapa significa ler e entender as fontes de dados e copiar as partes necessárias para a área de transformação de dados, a fim de serem trabalhadas posteriormente. Na grande maioria dos DW, os dados provêm de várias fontes diferentes e independentes. A extração pode ser conduzida através da construção de programas cujo código é executado sobre um sistema fonte de modo a gerar arquivos com os dados desejados. Outra opção é utilizar ferramentas de extração específicas que geram código próprio, interno à ferramenta, executado sobre o sistema fonte, de forma a obter os dados necessários. SANTANA (p.3, 2015).
- **Transformação dos Dados:** Após a extração dos dados dos sistemas fontes, passa-se à realização de uma série de atividades sobre esses dados, com o intuito de convertê-los em formato adequado para carga no DW e valioso para a organização. Na transformação dos dados estão envolvidos um ou vários processos, dependendo da necessidade e situação. Esses processos são: limpeza, eliminação, combinação, desnormalização, cálculos, derivação e alocação. SANTANA (p.4, 2015).
- **Carga de Dados:** Na última etapa do processo ETL, logo após os dados serem transformados, eles são carregados no DW. A parte de carga dos dados possui uma enorme complexidade no processo, e os seguintes fatores devem ser levados em conta ao realizar essa etapa: integridade dos dados, tipos de carga a serem realizadas, otimização do processo de carga, suporte completo ao processo de carga. SANTANA (p.4, 2015).

2.4. Modelagem Dimensional de Dados

Em relação a modelagem de dados, os envolvidos devem-se concentrar nos aspectos mais importantes que são realidade dentro da empresa, com objetivo de construir um sistema capaz de atender as necessidades informadas.

Ao se retratar as necessidades de informação que uma organização possui, é preciso alcançar os objetivos desta mesma realidade. Ao se coletar e selecionar os fatos relevantes, é necessário identificar os elementos geradores de informação (dados). É preciso também fazer uso de um modelo, ou seja, algo que nos mostre como as informações estão relacionadas (fatos). E, com base no modelo criado, é possível interagir com os usuários validando seus objetivos e metas, permitindo a construção de um sistema de informações próximo da realidade. SANTANA (p.4, 2015).

Como foi citado acima, existem alguns tipos modelos de dados nesse trabalho será comparado o modelo relacional que foi criado para atender os sistemas transacionais OLTP (*On-line Transaction Processing*) e o modelo dimensional criado para atender os sistemas analíticos OLAP (*On-line Analytical Processing*) que no caso é o foco do projeto. A Tabela 2 exemplifica a diferença dos modelos citados.

Tabela 2. Comparação dos sistemas OLAP e OLTP (SANTANA, 2015).

PERSPECTIVA	OLAP	OLPT
Foco	Foco no nível estratégico da organização. Visa a análise empresarial e tomada de decisão.	Foco no nível operacional da organização. Visa a execução operacional do negócio.
Performance	Otimização para a leitura e geração de análises e relatórios gerenciais.	Alta velocidade na manipulação de dados operacionais, porém ineficiente para geração de análises gerenciais.
Estrutura dos dados	Os dados estão estruturados na modelagem dimensional. Os dados normalmente possuem alto nível de sumarização.	Os dados são normalmente estruturados em um modelo relacional normalizado, otimizado para a utilização transacional. Os dados possuem alto nível de detalhes.
Armazenamento	O armazenamento é feito em estruturas de DW com otimização no desempenho em grandes volumes de dados.	O armazenamento é feito em sistemas convencionais de banco de dados através dos sistemas de informações da organização.
Abrangência	É utilizado pelos gestores e analistas para a tomada de decisão.	É utilizado por técnicos e analistas e engloba vários usuários da organização.
Frequência de atualização	A atualização das informações é feita no processo de carga dos dados. Frequência baixa,	A atualização dos dados é feita no momento da transação. Frequência muito alta de atualizações.

	podendo ser diária, semanal, mensal ou anual (ou critério específico).	
Volatilidade	Dados históricos e não voláteis. Os dados não sofrem alterações, salvo necessidades específicas (por motivos de erros ou inconsistências de informações).	Dados voláteis, passíveis de modificação e exclusão.
Tipos de permissões nos dados	É permitido apenas a inserção e leitura. Sendo que para o usuário está apenas disponível e leitura.	Podem ser feito leitura, inserção, modificação e exclusão dos dados.

Segundo Reis (2009), um DW utiliza normalmente dados em formado desnormalizados. Isto aumenta o desempenho das consultas e, como benefício adicional, o processo torna-se mais intuitivo para os usuários finais. O modelo multidimensional é baseado em três tipos de estruturas:

- **Fatos:** Considerada a tabela central do modelo, possuindo os valores levantados e que se deseja analisar. Nessa tabela será encontrado chaves estrangeiras conhecidas no mercado como FK (*Foreign Key*), relacionadas com suas respectivas tabelas de dimensões e seus campos numéricos nos quais receberão medidas obtendo seus valores que serão analisados.
- **Dimensão:** É uma tabela que representa qual o enfoque do negócio ou visão que será trabalhado. Nesse tipo de tabela será armazenado os principais atributos que se deseja obter, além de apresentar uma chave primária conhecida no mercado como PK (*Primary Key*) para manter uma integridade referencial com a tabela Fato.
- **Medidas:** Representam o desempenho de um indicador através de atributos que caracterizam a tabela Fato. Vale destacar que cada medida é determinada em função das dimensões envolvidas com a tabela Fato.

Outro ponto a se destacar é que o DW pode seguir alguns esquemas como o Estrela ou Floco de Neve. Segundo Braghittoni (p.63, 2017), no esquema Estrela (*Star Schema*), as tabelas de Dimensões são diretamente relacionadas com a tabela Fato. Não existe normalização das informações e as consultas são respondidas da maneira mais rápida possível. A Figura 4 exemplifica esse esquema, onde as tabelas: Clientes, Canal, Produtos, Local e Data são consideradas as dimensões do esquema e Vendas a tabela Fato do esquema.

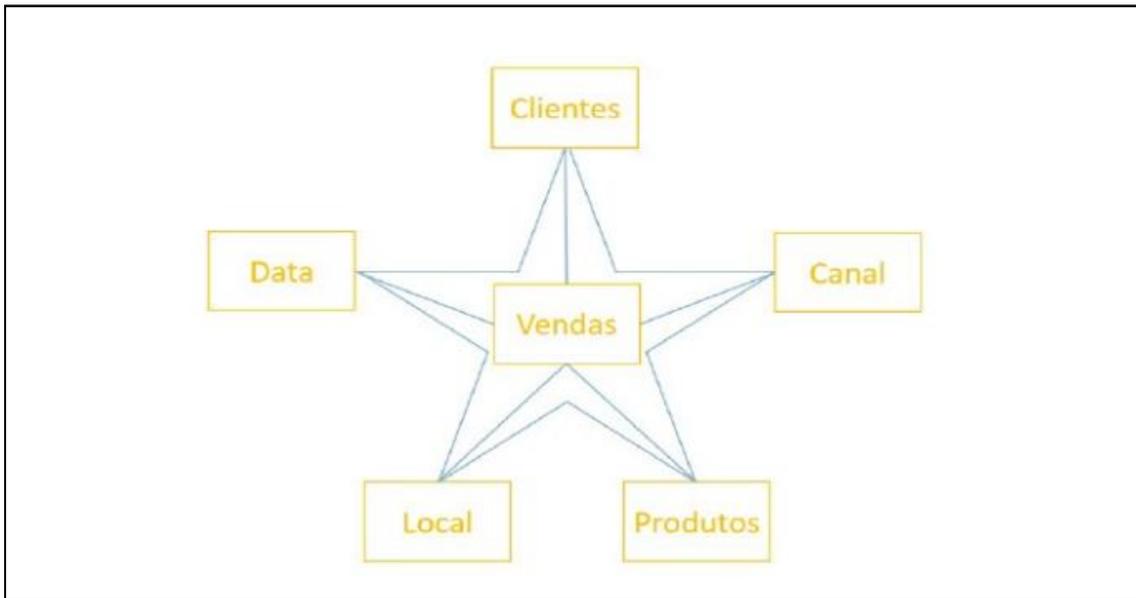


Figura 4. Esquema Estrela. (BRAGHITTONI, 2017).

Ainda na citação de Braghittoni, no esquema Floco de Neve (*Snow Flake Schema*) as dimensões podem possuir algum nível de normalização, gerando relacionamentos entre as tabelas das dimensões. A Figura 5 exemplifica esse esquema onde, as tabelas: Grupo e Clientes ou País, Estado e Cidade ou Produto e Detalhes ou Canal estão se relacionando para geração de dimensões e Vendas como a tabela central do esquema.

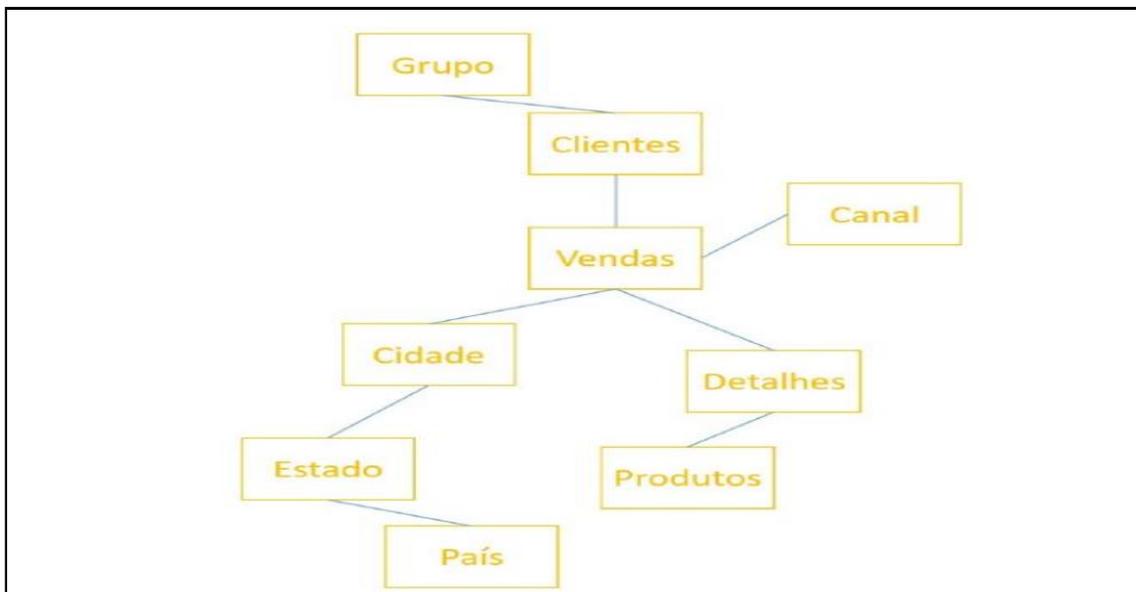


Figura 5. Esquema Floco de Neve. (BRAGHITTONI, 2017).

3. Indicadores de Gestão do Banco Alimentário de La Plata

Esta seção descreve os principais indicadores de performance para melhoria do processo de gestão do Banco Alimentário de La Plata, em um trabalho conjunto entre os diretores

do banco e o grupo de projeto de extensão da faculdade CESJF, além de apresentar como foi realizado o processo de extração, transformação e carregamento de dados através de um pseudocódigo.

3.1. Indicadores de Performance do Banco Alimentário

Para a execução deste projeto foram especificados um total de 20 indicadores de performance, relacionados e categorizados em 4 perspectivas. Cada indicador e perspectiva foi analisado cuidadosamente com os diretores do Banco Alimentário de La Plata, nos quais após várias reuniões foram extraídos as principais informações relacionadas aos problemas informados, até chegar na conclusão de transformação dos problemas em indicadores e consequentemente as perspectivas listados na Tabela 3 apresentada a seguir.

Tabela 3. Indicadores de performance do Banco Alimentário. (DO AUTOR).

Perspectivas	Indicador
DONANTES	KILOS INGRESADOS/DONANTE
	KILOS INGRESADOS/CATEGORIA
	VALOR NUTRICIONAL/DONACION
	VALUACIÓN/DONACION
	MERMA/DONANTE
TRANSPORTE	GASTO EN \$ DE NAFTA/KILOS DEPOSITO
	GASTO EN \$ DE NAFTA/KILOS MERCADO REGIONAL
	CANTIDAD DE TRANSPORTES DONADOS POR MES
	GASTO DE FLETES
STOCK	KILOGRAMOS DE ENTRADA/KILOGRAMOS DE ENTRADA OBJETIVOS
	KILOGRAMOS DE SALIDA/KILOGRAMOS DE SALIDA OBJETIVOS
	KILOGRAMOS VENCIDOS/KILOGRAMOS ENTRADA
	KILOGRAMOS AJUSTES/KILOGRAMOS ENTRADA
	KILOGRAMOS ROTURA/KILOGRAMOS ENTRADA
	KILOGRAMOS MERMA/KILOGRAMOS ENTRADA
	VALOR NUTRICIONAL ENTREGADO
	KILOS DE ENTRADA/CATEGORIA
	KILOS DE SALIDA/CATEGORIA
	CANTIDAD DE PEDIDOS ARMADOS POR MES
SALIDAS	TRAZABILIDAD DE CADA DONACION

4. Implementação dos Indicadores de Gestão do Banco Alimentário de La Plata

Essa seção descreve as tecnologias envolvidas na geração dos indicadores de gestão, como foi realizado a criação e desenvolvimento da modelagem utilizada no processo de BI para o Banco Alimentário. Além de demonstrar a execução da implementação do processo de ETL através de pseudocódigo.

4.1. Tecnologias envolvidas na Geração de Indicadores de Gestão

Ao longo do projeto foi utilizado uma série de tecnologias, metodologias e softwares. Inicialmente é válido ressaltar a utilização do conceito de sistemas colaborativos, que é vital no projeto pelo motivo de estar sendo envolvido um grupo de pessoas de trabalham em regiões distintas em uma cidade ou até mesmo em outro país.

Em outras palavras, “Sistemas Colaborativos são ferramentas de software utilizadas em redes de computadores para facilitar a execução de trabalhos em grupos. Essas ferramentas devem ser especializadas o bastante, a fim de oferecer aos seus usuários formas de interação, facilitando o controle, a coordenação, a colaboração e a comunicação entre as partes envolvidas que compõe o grupo, tanto no mesmo local, como em locais geograficamente diferentes e que as formas de interação aconteçam tanto ao mesmo tempo ou em tempos diferentes. Percebe-se com isso que o objetivo dos Sistemas Colaborativos é diminuir as barreiras impostas pelo espaço físico e o tempo”. (CAMARGO, KHOURI, GIAROLA, 2005).

No projeto foi utilizado a plataforma Jira¹, disponibilizado pela empresa Atlassian, no qual permite o monitoramento das tarefas e acompanhamento de projetos garantindo o gerenciamento de todas as suas tarefas em um único lugar.

Para armazenamento das informações utilizada pela aplicação foi utilizado alguns bancos de dados. O primeiro é o SQL Server 2005², disponibilizado pela empresa Microsoft Corporation. No caso do SQL Server sua utilização foi originada por motivos do software antigo utilizado pela ONG e que no caso mesmo apesar da idade do software foi de ótima integração com a aplicação. Já no segundo banco foi utilizado o PostgreSQL³, disponibilizado pela empresa *PostgreSQL Global Development Group*. No qual foi escolhido principalmente por questões de experiência dos integrantes do grupo de extensão no projeto além de ser bem avaliado e utilizado no mercado.

Para geração dos painéis de controle que serão formas gráficas de visualização da aplicação como resultante das informações foi escolhido o software Metabase⁴, herdando o nome da própria empresa e que no caso optado por ser uma ferramenta gratuita e de fácil utilização e integração no projeto.

Outro ponto importante a se destacar é a utilização da tecnologia da Microsoft conhecida como Agendador de Tarefas do S.O (Sistema Operacional) Windows, no qual irá ser muito útil para agendamento das rotinas diárias executadas pelo interpretador de comandos do Windows.

Como linguagem da aplicação principal foi utilizado a sintaxe Java⁵, fornecido pela Oracle Corporation e utilizado pelo motivo consagrada no mercado além de forte influência pela faculdade ao adotar como linguagem de programação nas disciplinas cursadas pelos alunos envolvidos no projeto.

¹ Pode ser acessada em: <https://br.atlassian.com/software/jira>

² Pode ser acessada em: <https://www.microsoft.com/pt-br/download/details.aspx?id=21844>

³ Pode ser acessada em: <https://www.postgresql.org/>

⁴ Pode ser acessada em: <https://metabase.com/>

⁵ Pode ser acessada em: <https://www.oracle.com/technetwork/java/javase/downloads/java-archive-downloads-javase7-521261.html>

Consequentemente pelo mesmo motivo citado anteriormente a IDE (*Integrated Development Environment*) que será utilizada como ambiente de desenvolvimento para essa linguagem foi optado a escolha da plataforma do NetBeans IDE 8.0.1⁶, desenvolvido pela empresa Apache Software Foundation.

Por fim tendo em vista todas as tecnologias envolvidas nos projetos e a organização a receber os serviços do grupo de alunos ser uma ONG é de extrema relevância a utilização de tecnologias gratuitas nas quais são de grande aceitação pela comunidade de tecnologia da informação.

4.2. Modelagem de BI para o Banco Alimentário de La Plata

Após toda abordagem realizada sobre conceitos de ETL, e a partir de reuniões feitas com os diretores do Banco Alimentário de La Plata foi definido o seguinte modelo de acordo com a Figura 6, que será descrito por etapas em seguida.

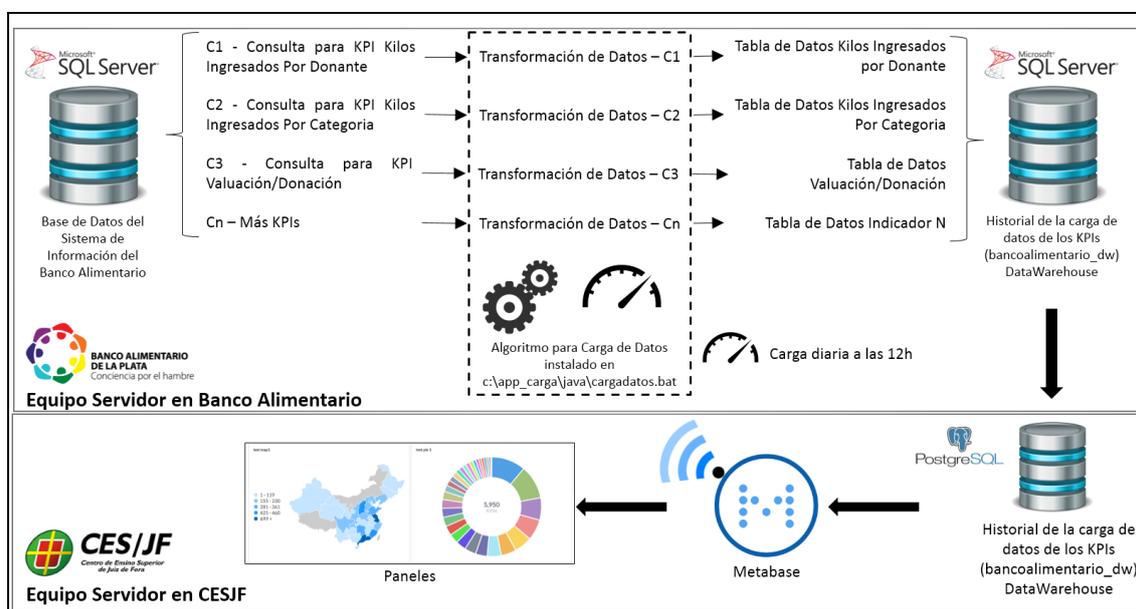


Figura 6. Modelagem de um processo de BI para o Banco Alimentário de La Plata. (DO AUTOR).

- Etapa 1: inicialmente todos os dados e informações que serão coletadas pelo processo de Extração do ETL, são retiradas exclusivamente do banco de dados do Banco Alimentário de La Plata, administrado pelo SQL Server 2005. Ainda nessa etapa é então utilizado portanto algumas consultas específicas através de scripts SQL (instruções de código para o banco de dados) extraíndo então as principais informações futuras para os indicadores.
- Etapa 2: nessa etapa seguindo o fluxo de ETL que seria de Transformação dos dados, é realizado o tratamento das informações obtidas no fluxo anterior, nas quais serão aplicadas em regras e padrões, em alguns casos multiplicidade das mesmas, mas sempre priorizando a integridade dos dados do Banco Alimentário de La Plata até então a execução da próxima etapa.

⁶ Pode ser acessada em: https://netbeans.org/downloads/8.0.1/?pagelang=pt_BR

- Etapa 3: após então a Transformação dos dados, é necessário executar o último passo do processo de ETL que seria o Carregamento dos dados. Nessa etapa consiste em transpor os dados tratados para o banco de DW. No modelo da Figura 6 acima, pode ser observado a existência de dois bancos DWs, em que no contexto do Banco Alimentário de La Plata foram criados por motivos segurança e back-up, nos quais são:
 - DW SQL Server 2005, está situado no servidor local no Banco Alimentário de La Plata.
 - DW PostgreSQL, está situado no servidor local na Faculdade CESJF.

Vale ressaltar que foi negociado com os diretores do Banco Alimentário de La Plata que o carregamento dos dados para os bancos citados acima é realizado às 12h com o auxílio do Agendador de Tarefas do S.O Windows.

- Etapa 4: Após esse todo fluxo os dados são manipulados por um outro sistema conhecido como Metabase utilizando-se das informações armazenadas nos DWs por meio das tabelas fatos para geração dos painéis de controle que irão retratar os indicadores solicitados e finalizando o ciclo do fluxo.

4.3. Implementação do processo de ETL

Para execução do processo de ETL demonstrado na seção 4, foi utilizado pelo grupo de extensão como ambiente de desenvolvimento o software NetBeans citado na seção 4.1 deste trabalho, por motivos de maior experiência em outros projetos. A linguagem escolhida foi o Java, além disso para o funcionamento correto da linguagem e o ambiente de desenvolvimento de acordo com a infraestrutura do Banco Alimentário de La Plata, foi utilizado o Java Development Kit, versão 7u60.

O projeto foi dividido em dividido em pacotes para melhoramento de questão de documentação e visualização, da seguinte forma:

- **com.cargadimensiones:** Pacote onde está localizado todos os arquivos java utilizados para criação das classes modelos das dimensões para criação das tabelas fatos.
- **com.cargafatos:** Pacote onde está localizado todos os arquivos java utilizados para criação das classes modelos dos indicadores e suas respectivas execuções.
- **com.dao:** Pacote onde está localizado todos os arquivos java utilizados para criação das conexões com os bancos de dados utilizados no projeto (SQL Server e PostgreSQL).
- **com.firebase:** Pacote onde está localizado todos os arquivos java utilizados para execução de carregamento de dados para o banco de dados Firebase.
- **com.sistema:** Pacote onde está localizado o arquivo java utilizado para carregamento das informações contidas no arquivo de propriedade do projeto.
- **sqljdbc4.jar:** Biblioteca jar de conexão com o banco de dados SQL Server.
- **postgresql-42.2.5.jre7.jar:** Biblioteca jar de conexão com o banco de dados PostgreSQL.

- **jdk1.7.0_60:** Pacote utilitário do Java Development Kit, versão 7u60 necessário para execução específica do projeto.
- **JDK 1.7:** Pacote utilitário padrão do software NetBeans IDE 8.0.1.

Antes de iniciar propriamente o processo de extração é válido citar quais são as tabelas que serão extraídas as informações da fonte de dados fornecida pelo Banco Alimentário e que também irão servir de base para as classes modelos dos indicadores. Na Tabela 4 abaixo é possível visualizar qual o nome da tabela e sua respectiva descrição:

Tabela 4. Tabelas utilizadas para geração dos indicadores. (DO AUTOR).

Tabela	Descrição
ENTPRD	Tabela de movimento de entrada de produto.
MAEDON	Tabela de registro de doador
MAEPRO	Tabela de registro do produto
MAECAT	Tabela de registro de categorias de produtos
MAELIP	Tabela de registro de linhas de produtos por categorias
MAEPRE	Tabela de registro do histórico de preços do produto
STOAJU	Tabela de ajustes de ações
SALPRD	Tabela de registro de produtos de saídas
MAEREC	Tabela de registro de receptores
PEDPRD	Tabela de registro de solicitação de produto

Portanto inicialmente para execução da etapa de extração do ETL é necessário executar uma query que consiga buscar as informações necessárias para o indicador informado, para exemplificar será utilizado o indicador “KILOS INGRESADOS/DONANTE”. De forma genérica é utilizada a seguinte query (Figura 7) dentro do sistema em Java para filtrar essa busca, conforme a sequência a seguir:

- SELECIONAR TODAS IDs e DESCRICOES
- DAS TABELAS REQUISITADAS PELO INDICADOR
- ONDE BUSQUEM SEUS IDs REFERENCIADOS
- E BUSQUEM PELA DATA INFORMADA DIARIA CASO PRECISSE DE REPROCESSAMENTO
- AGRUPANDO PELAS IDs e DESCRICOES

```

String reprocessamiento = prop.getProp().getProperty("prop.reprocessing");
String initialDate =prop.getProp().getProperty("prop.initialdate");
String finalDate =prop.getProp().getProperty("prop.finaldate");

SQL = "SELECT ENT.ID_NUMENT ";
SQL = SQL + ", DON.ID_MAEDON ";
SQL = SQL + ", DON.DE_RAZDON ";
SQL = SQL + ", PRO.ID_MAEPRO ";
SQL = SQL + ", PRO.DE_PRODUC ";
SQL = SQL + ", PRO.DE_LINPRO ";
SQL = SQL + ", CAT.DE_CATEGO ";
SQL = SQL + ", ent.FH_INGRES FH_INGRES ";
SQL = SQL + ", SUM(ENT.CA_TOTPE) TO_TOTPE ";
SQL = SQL + "FROM ENTPRD ENT ";
SQL = SQL + ", MAEDON DON ";
SQL = SQL + ", MAEPRO PRO ";
SQL = SQL + ", MAECAT CAT ";
SQL = SQL + ", MAELIP LIP ";
SQL = SQL + "WHERE ENT.ID_MAEDON = DON.ID_MAEDON ";
SQL = SQL + "AND ENT.ID_MAEPRO = PRO.ID_MAEPRO ";
SQL = SQL + "AND PRO.ID_MAEDON = DON.ID_MAEDON ";
SQL = SQL + "AND PRO.DE_LINPRO = LIP.DE_LINPRO ";
SQL = SQL + "AND LIP.DE_CATEGO = CAT.DE_CATEGO ";
if (reprocessamiento.contains("N")){
    SQL = SQL + "AND CONVERT(DATETIME,CONVERT(VARCHAR, ent.FH_INGRES ,112)
    SQL = SQL + "AND CONVERT(DATETIME,CONVERT(VARCHAR, ent.FH_INGRES ,112)
} else{
    SQL = SQL + "AND CONVERT(DATETIME,CONVERT(VARCHAR, ent.FH_INGRES ,112)
    SQL = SQL + "AND CONVERT(DATETIME,CONVERT(VARCHAR, ent.FH_INGRES ,112)
}
SQL = SQL + "GROUP BY ENT.ID_NUMENT ";
SQL = SQL + ", DON.ID_MAEDON ";
SQL = SQL + ", DON.DE_RAZDON ";
SQL = SQL + ", PRO.ID_MAEPRO ";
SQL = SQL + ", PRO.DE_PRODUC ";
SQL = SQL + ", PRO.DE_LINPRO ";
SQL = SQL + ", CAT.DE_CATEGO ";
SQL = SQL + ", ent.FH_INGRES; ";

```

Figura 7. Código em java utilizado para extração dos dados do indicador informado. (DO AUTOR).

Em seguida os dados brutos extraídos são armazenados em uma tabela temporária, nessa tabela os dados das colunas são comparados com os valores vindos da tabela de dimensão para manter a integridade dos dados, através então dessa comparação de dados armazenados, os mesmos que estão também em uma tabela virtual, serão portanto carregados no futuro na tabela fato, após todo esse tratamento dos dados. A Figura 8 e Figura 9 mostra esse processo em uma sequência de instruções:

- INSERIR NA TABELA TEMPORARIA
- OS VALORES IDs e DESCRICOES EXTRAIDOS
- CASO A COMPARACAO SEJA SUCESSO SELECIONAR A TABELA VIRTUAL
- PARA UMA FUTURA INSERCAO NA TABELA FATO
- CASO CONTRARIO REFAZER O PROCESSO

```

SQL="INSERT INTO T_T_KILO_INGRESADOS_DONANTE VALUES (";
SQL = SQL +"" + rec.getString("ID_MAEDON") +",";
SQL = SQL +""+ rec.getString("DE_RAZDON").replace("","") +",";
SQL = SQL +"" + rec.getString("ID_MAEPRO").replace("","") +",";
SQL = SQL +""+ rec.getString("DE_PRODUC").replace("","") +",";
SQL = SQL +""+ rec.getString("DE_LINPRO").replace("","") +",";
SQL = SQL +""+ rec.getString("DE_CATEGO").replace("","") +",";
SQL = SQL +"" substring(''+ rec.getString("FH_INGRES") +"" , 1,4)+'-' +s
SQL = SQL +"" + rec.getString("ID_NUMENT").replace("","") +"" ,"";
SQL = SQL +"" + rec.getString("TO_TOTPES").replace("","") +""");

```

Figura 8. Inserção dos valores extraídos na tabela temporária. (DO AUTOR).

```

stDW = connDW.getConnDW().createStatement();
stDW.executeUpdate(SQL);
stDW.close();

SQL = "SELECT ID_D_CATEGO, ID_D_DONANTE, ID_D_PRODUCTO, ID_D_TIEMPO, ID_NUMENT, TO_TOTPES FROM VW_T_KILO_INGRESADOS_DONANTE";
recDWhecho = connDW.getConnDW().createStatement().executeQuery(SQL);

```

Figura 9. Seleção na tabela virtual. (DO AUTOR).

Em seguida é realizado a parte de preparação da tabela fato dos novos dados que serão armazenados, onde é analisado principalmente o fator da dimensão tempo (Figura 10).

```

SQL = "DELETE FROM T_H_KILO_INGRESADOS_DONANTE ";
SQL = SQL + "WHERE ID_D_H_KILO_INGRESADOS_DONANTE IN ";
SQL = SQL + "(";
SQL = SQL + "SELECT HKILODONANTE.ID_D_H_KILO_INGRESADOS_DONANTE ";
SQL = SQL + "FROM T_H_KILO_INGRESADOS_DONANTE HKILODONANTE ";
SQL = SQL + ", T_D_TIEMPO TIEMPO ";
SQL = SQL + "WHERE HKILODONANTE.ID_D_TIEMPO = TIEMPO.ID_D_TIEMPO ";
if (reprocesamiento.contains("N")){
    SQL = SQL + "AND CONVERT(DATETIME,CONVERT(VARCHAR, TIEMPO.FECHA_TIEMPO
    SQL = SQL + "AND CONVERT(DATETIME,CONVERT(VARCHAR, TIEMPO.FECHA_TIEMPO
}else{
    SQL = SQL + "AND CONVERT(DATETIME,CONVERT(VARCHAR, TIEMPO.FECHA_TIEMPO
    SQL = SQL + "AND CONVERT(DATETIME,CONVERT(VARCHAR, TIEMPO.FECHA_TIEMPO
}
SQL = SQL + ")";

stDW = connDW.getConnDW().createStatement();
stDW.executeUpdate(SQL);
stDW.close();

SQL = "SELECT ID_D_CATEGO, ID_D_DONANTE, ID_D_PRODUCTO, ID_D_TIEMPO, ID_NUMENT, TO_TOTPES FROM VW_T_KILO_INGRESADOS_DONANTE";
recDWhecho = connDW.getConnDW().createStatement().executeQuery(SQL);

```

Figura 10. Preparação da tabela fato. (DO AUTOR).

Por fim é executado então o processo de carregamento dos dados para a tabela fato e exportação para um arquivo em CSV (*Comma Separated Values*), conforme a sequência e figura a seguir (Figura 11):

- INSERIR NA TABELA FATO
- OS VALORES IDs e DESCRICOES EXTRAIDOS

```
SQL="INSERT INTO T_H_KILO_INGRESADOS_DONANTE (ID_D_H_KILO_INGRESADOS
SQL = SQL + ""+ID_Carga+"";
SQL = SQL + ""'+ recDWhecho.getString("ID_D_CATEGO") +"";
SQL = SQL + ""'+ recDWhecho.getString("ID_D_DONANTE") +"";
SQL = SQL + ""'+ recDWhecho.getString("ID_D_PRODUCTO") +"";
SQL = SQL + ""'+ recDWhecho.getString("ID_D_TIEMPO") +"";
SQL = SQL + ""'+ recDWhecho.getString("ID_NUMENT") +"";
SQL = SQL + ""'+ recDWhecho.getString("TO_TOTPES") +"";

connDW.getConnection().createStatement().executeUpdate(SQL);

fileCSV = fileCSV + ""+ID_Carga+"";
fileCSV = fileCSV + ""'+ recDWhecho.getString("ID_D_CATEGO") +"";
fileCSV = fileCSV + ""'+ recDWhecho.getString("ID_D_DONANTE") +"";
fileCSV = fileCSV + ""'+ recDWhecho.getString("ID_D_PRODUCTO") +"";
fileCSV = fileCSV + ""'+ recDWhecho.getString("ID_D_TIEMPO") +"";
fileCSV = fileCSV + ""'+ recDWhecho.getString("ID_NUMENT") +"";
fileCSV = fileCSV + ""'+ recDWhecho.getString("TO_TOTPES") +"";

ID_Carga++;
```

Figura 11. Inserção dos valores extraídos na tabela fato. (DO AUTOR).

Após isso o sistema informa alguns tipos de mensagens caso a aplicação execute corretamente ou quando haja alguma falha, e representado pelo trecho de código descrição na Figura 12:

- de sucesso,
- de não pode encontrar registros,
- de que não foi possível realizar a carga dos dados.

```
if (achouReg==true) {
    ConnectionLog log = new ConnectionLog();
    log.ExecuteLog("Carga Diaria de la tabla hecho - Kilos Ingresados Por Donantes", "HECHO");
} else{
    System.err.println("Carga Diaria de la tabla hecho - Kilos Ingresados Por Donantes No encontró registros y no fue procesada!");
}

System.err.println("Carga del DataWarehouse de la tabela hecho Kilos Ingressado por Donante Cerrada!");
```

Figura 12. Resultados das mensagens ao término de execução da aplicação. (DO AUTOR).

Portanto ao término da aplicação é possível visualizar o modelo dimensional criado na Figura 13, onde demonstra portanto a criação das tabelas de dimensões ao redor da tabela fato-central do indicador.

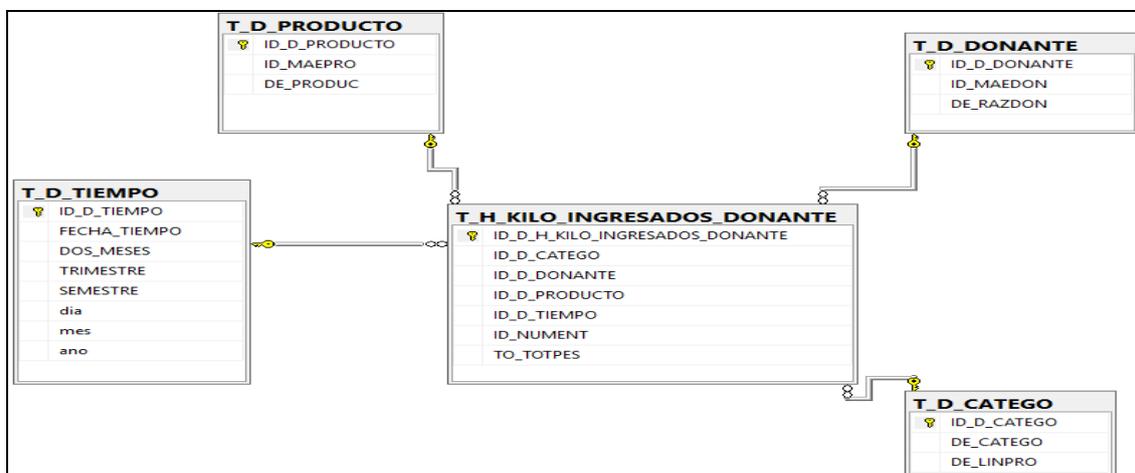


Figura 13. Modelo dimensional representativo ao indicador “KILOS INGRESADOS/DONANTE”. (DO AUTOR).

Por fim, pode ser acessada via internet a ferramenta do Metabase para visualizar através dos painéis de controles, o indicador que foi criado. No exemplo tratado nesse trabalho o indicador “KILOS INGRESADOS/DONANTE”.

A Figura 14, exhibe alguns dos painéis de controle por padrão criados pela ferramenta do Metabase, tendo em vista a quantidade de quilos admitidos por cada doador.

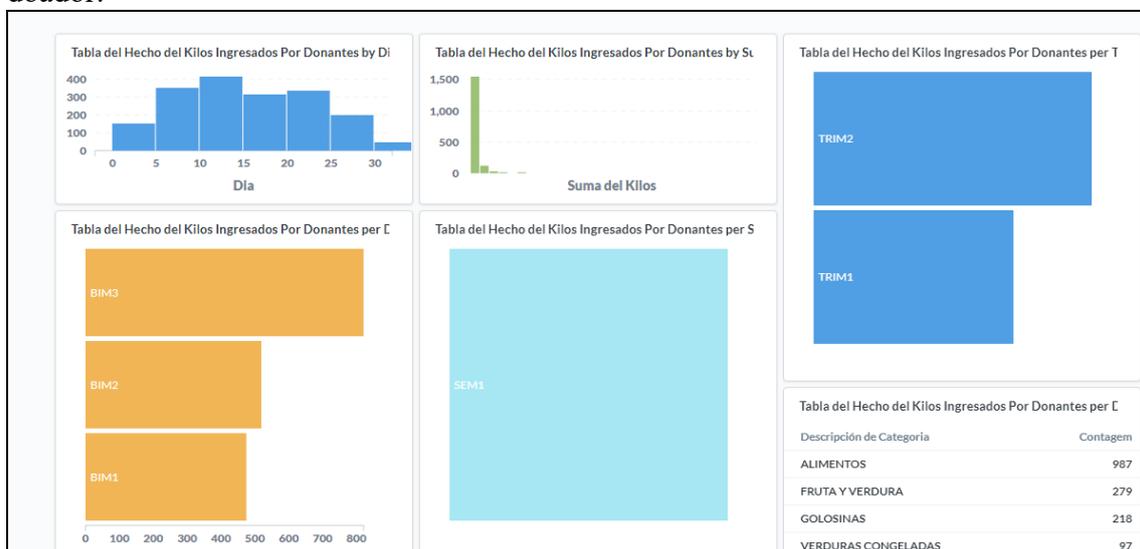


Figura 14. Painéis de Controle oferecidos inicialmente pela ferramenta Metabase. (DO AUTOR).

Já na Figura 15, exhibe um gráfico em forma de pizza a quantidade de quilos admitidos por doadores ao longo de 30 dias, mais especificamente dentre os dias 10 e 15 com cerca de 22.8% foram coletados a maior quantidade de quilos do mês totalizando 414 quilos.

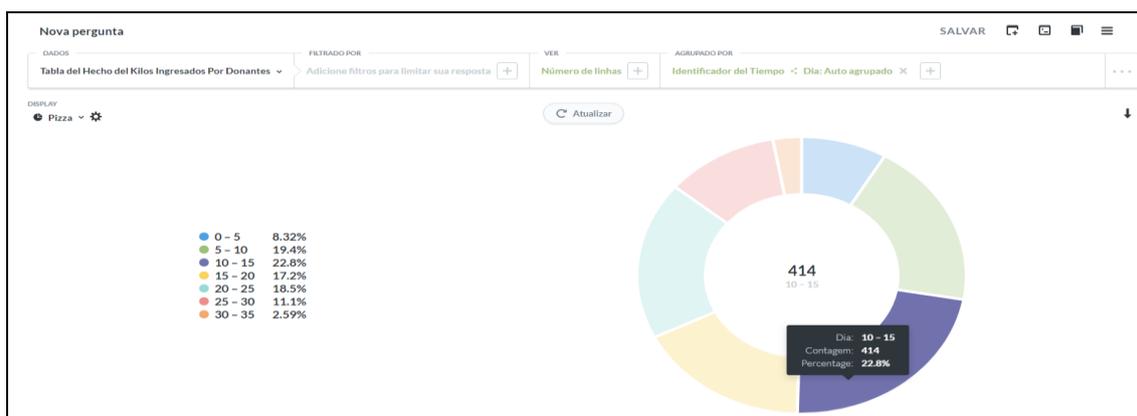


Figura 15. Painel de Controle em formato Pizza, caso o usuário queira uma interação dinâmica com o painel, ele pode alterar filtros ou outros painéis ou agrupamentos. (DO AUTOR).

5. Considerações Finais

Este trabalho mostrou, portanto, o estudo de caso realizado em parceria com a ONG do Banco Alimentário de La Plata, através do curso de Engenharia de Software a criação de uma aplicação capaz de atender as necessidades atuais de uma organização em questões de tecnologia da informação por meio do processo de BI abordando conceitos teóricos e práticos.

Vale a pena ressaltar que o grande foco desse trabalho foi em priorizar a parte de ETL da aplicação e que a parceria da faculdade CESJF e a ONG ainda está acontecendo até o final de 2019, portanto está em desenvolvimento alguns indicadores restantes de modo que ficará uma abordagem aprofundada para os próximos trabalhos relacionados a este projeto e também a outras áreas como por exemplo a elaboração de redocumentação dos sistemas da ONG envolvidos com esse projeto.

Logo espera-se que a solução desenvolvida por esse trabalho possa servir como um caminho de auxílio para tomada de decisões estratégicas do Banco Alimentário de La Plata através de um DW consistente e em crescimento.

Agradecimentos

Agradeço à minha família, principalmente minha mãe Patrícia Pinheiro e minha irmã Letícia Pinheiro Latuque por todo apoio e incentivo. Aos amigos, em especial aqueles cuja amizade foi criada e se fortaleceu ao longo dos anos de curso, Arllen Demétrio Mota de Souza, Eduardo Couto Ciscoto, Maicon Alcântara de Oliveira e Raphael Soares Moreira.

Minha gratidão ao Centro de Ensino Superior de Juiz de Fora, com menção especial ao meu orientador Evaldo de Oliveira da Silva pela paciência e revisão deste trabalho, ao professor Carlos Alberto Ribeiro por ter apresentado a área que mais me identifiquei que é Banco de Dados e ao professor/coordenador Marco Antônio Pereira Araújo pela excelência e qualidade em sala e administração do curso de Engenharia de Software. A parceria concretizada com o Banco Alimentário de La Plata, com grande respeito e profissionalismo a Gastón Zappalá e Dolores Puig pelo reconhecimento do trabalho realizado.

Agradeço à Deus, pela força e coragem durante toda esta longa caminhada.

Por fim, meu mais sincero obrigado a quem, de forma direta ou indireta, ajudou com este trabalho e com minha formação profissional.

Referências

- BANCO ALIMENTÁRIO DE LA PLATA. Disponível em: <<http://bancoalimentario.org.ar/quienes-somos/>>. Acesso em: 1 de Maio de 2019.
- BONEL, Claudio. "Afinal, o que é Business Intelligence". Ed. 1º Edição. São Paulo: 2017.
- BRAGHITTONI, Ronaldo. "Business Intelligence: Implementar do jeito certo e a custo zero". Ed. Casa do Código. São Paulo: 2017.
- CAMARGO, A. A. B. de, KHOURI, L. H. El & GIAROLA, P. C., (2005). "O Uso de Sistemas Colaborativos na Gestão de Projetos: Fatores Relevantes para o Sucesso". Trabalho de Conclusão de Curso. Fundação Instituto de Administração-FIA. Disponível em: <https://www.researchgate.net/publication/321984612_OS_SISTEMAS_DE_INFORMACAO_COMO_SUPORTE_A_TOMADA_DE_DECISAO ESTRATEGICA>. Acesso em 1 de Maio de 2019.
- JAVA DEVELOPMENT KIT 7u60. Disponível em: <<https://www.oracle.com/technetwork/java/javase/downloads/java-archive-downloads-javase7-521261.html>>. Acesso em 1 de Maio de 2019.
- JIRA. Disponível em: <<https://br.atlassian.com/software/jira>>. Acesso em 1 de Maio de 2019.
- MATILLION. "ETL vs. ELT – What’s the Big Difference?" Disponível em: <<https://www.matillion.com/events/etl-vs-elt-whats-big-difference/>>. Acesso em: 1 de Maio de 2019.
- NETBEANS IDE 8.0.1. Disponível em: <https://netbeans.org/downloads/8.0.1/?pagelang=pt_BR>. Acesso em 1 de Maio de 2019.
- POSTGRESQL. Disponível em: <<https://www.postgresql.org/>>. Acesso em 1 de Maio de 2019.
- REIS, E., TEIXEIRA, F. e ARAÚJO, M. A. (2009). "Implementando uma solução de Business Intelligence com o Microsoft SQL Server 2005 – Parte 1". SQL Magazine, Rio de Janeiro, n. 59, p. 52-66.
- SANTANA, Leandro. "Análise de modelos dimensionais: um estudo de caso com dados do sistema de processo seletivo do IFET". Trabalho de conclusão de Curso. Centro de Ensino Superior de Juiz de Fora. Disponível em: <<https://seer.cesjf.br/index.php/cesi/article/view/518>>. Acesso em 1 de Maio de 2019.
- SQL SERVER 2005. Disponível em: <<https://www.microsoft.com/pt-br/download/details.aspx?id=21844>>. Acesso em 1 de Maio de 2019.