

UNIVERSIDADE FEDERAL DE GOIÁS  
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E  
COMPUTAÇÃO

ENGENHARIA DE COMPUTAÇÃO

LARISSA RIBEIRO DE MELO PEREIRA

**Business Intelligence e Data Warehouse  
Aplicados na Análise de Consumo e  
Geração de Energia Elétrica na UFG**

Goiânia  
2019

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR  
VERSÕES ELETRÔNICAS DE TRABALHO DE CONCLUSÃO DE CURSO DE  
GRADUAÇÃO NO REPOSITÓRIO INSTITUCIONAL DA UFG**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio do Repositório Institucional (RI/UFG), regulamentado pela Resolução CEPEC nº 1204/2014, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

**1. Identificação do Trabalho de Conclusão de Curso de Graduação (TCCG):**

Nome completo do autor: Larissa Ribeiro de Melo Pereira

Título do trabalho: Business Intelligence e Data Warehouse Aplicados na Análise de Consumo e Geração de Energia Elétrica na UFG

**2. Informações de acesso ao documento:**

Concorda com a liberação total do documento  SIM  NÃO<sup>1</sup>

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF do TCCG.

Larissa Ribeiro de Melo Pereira  
(Larissa Ribeiro de Melo Pereira)<sup>2</sup>

Ciente e de acordo:

Marcelo Stehling de Castro  
(Marcelo Stehling de Castro)<sup>2</sup>

Data: 13/12/2019

<sup>1</sup> Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Versão abril de 2018

<sup>2</sup> As assinaturas devem ser originais sendo assinadas no próprio documento, imagens coladas não serão aceitas.

LARISSA RIBEIRO DE MELO PEREIRA

# **Business Intelligence e Data Warehouse Aplicados na Análise de Consumo e Geração de Energia Elétrica na UFG**

Monografia de projeto final de curso apresentado à Coordenação de Projeto Final do Curso de Engenharia de Computação da Escola de Engenharia Elétrica, Mecânica e Computação da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Bacharel em Computação.

**Orientador:** Prof. Dr. Marcelo Stehling de Castro

Goiânia  
2019

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Ribeiro de Melo Pereira, Larissa  
Business Intelligence e Data Warehouse Aplicados na Análise de Consumo e Geração de Energia Elétrica na UF [manuscrito] / Larissa Ribeiro de Melo Pereira. - 2019.  
40, f.

Orientador: Prof. (Marcelo Stehling de Castro  
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Goiás, Escola de Engenharia Elétrica, Mecânica e de Computação (EMC), Engenharia da Computação, Cidade de Goiás, 2019.

Bibliografia.

Inclui siglas, lista de figuras, lista de tabelas.

1. Business Intelligence. 2. Data Warehouse. 3. ETL. I. Stehling de Castro, (Marcelo , orient. II. Título.

CDU 519.25



### ATA DE AVALIAÇÃO DE PROJETO FINAL

#### CURSO

( ) Eng. Elétrica    ( ) Eng. Mecânica    (X) Eng. de Computação  
( ) Projeto Final 1    (X) Projeto Final II

#### AVALIAÇÃO DE PROJETO FINAL

Título do projeto: Business Intelligence e Data Warehouse Aplicados na  
Análise de Consumo e Geração de Energia Elétrica na UFG

#### BANCA AVALIADORA

Membro 1: Marcelo Stehling de Castro

Membro 2: Sandrerley Ramos Pires

Membro 3: Gustavo Dias de Oliveira

#### ESTUDANTES

Matrícula	Nome
20140369	Larissa Ribeiro de Melo Pereira

#### NOTAS


Matrícula	Membro 1: Marcelo Stehling de Castro				Membro 2: Sandrerley Ramos Pires				Membro 3: Gustavo Dias de Oliveira				Média
	NPT	NTE	NAA	NF	NPT	NTE	NAA	NF	NPT	NTE	NAA	NF	
20140369	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5

NPT – Nota plano de trabalho; NTE – Nota do trabalho escrito; NAA – Nota de apresentação e arguição

Para Eng. Elétrica, Mecânica e PFC2 da Eng. Da Computação:  $NF = 0,1 \times NPT + 0,45 \times NTE + 0,45 \times NAA$

Para PFC1 da Eng. Da Computação:  $NF = 0,3 \times NPT + 0,7 \times NAA$

Goiânia, 09 de dezembro de 2019.

  
\_\_\_\_\_  
Marcelo Stehling de Castro

  
\_\_\_\_\_  
Sandrerley Ramos Pires

  
\_\_\_\_\_  
Gustavo Dias de Oliveira



Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador(a).

**Larissa Ribeiro de Melo Pereira**

Graduando em Engenharia de Computação na UFG - Universidade Federal de Goiás. Durante sua graduação, realizou atividades de estágio pela Centro de Distribuição de Energia de Goiás(CELG) por um ano, foi estagiária na Escola de Engenharia Elétrica, Mecânica e Computação (EMC) por quatro meses e também na Umobi Soluções Digitais por três meses. Além disso, fez parte do projeto de melhoria de eficiência energética através da implantação de usinas fotovoltaicas.

Dedico ao meus pais.

## **Agradecimentos**

Primeiramente, eu agradeço a Deus pela graça de ajudar-me a vencer nos momentos de dificuldade, por me oferecer o dom de aprender e de persistir diante os desafios, por me mostrar o caminho que eu devo seguir.

Agradeço a minha família, principalmente aos meus pais Rômulo e Rosimeire, que forneceram durante a minha vida todo apoio, dedicação, amor e orações para que eu trilhasse meu caminho.

Agradeço ao meu companheiro Murillo por sempre acreditar em mim, pelo amor, carinho, compreensão em todos os momentos, por me socorrer nos momento de desespero, por me apoiar nos momentos de insegurança.

Agradeço a minha irmã Letícia, que além de irmã sempre foi minha amiga, por todo apoio e conselhos durante minha vida. As minhas tias Maria, Gerusa pelas demonstrações de afeto e pela minha tia Lúcia que mesmo não estando mais entre nos sempre demonstrou carinho.

Agradeço ao meu orientador Marcelo Stehling de Castro, por todo o suporte, apoio para a realização deste trabalho, por ter sido um excelente orientador. Sou grata por todas as dicas e por ter acreditado em mim. Também agradeço ao Gustavo Dias por seus ensinamentos e sugestões para a pesquisa do projeto.

Agradeço ao Guilherme Marcorio, por me ajudar a solucionar uns problemas no processo prático do meu projeto.

“Feliz o homem que acha sabedoria, e o homem que adquire conhecimento. [2]”

**BÍBLIA**,  
*Provérbio 3:13.*

## Siglas

API - Application Programming Interface.

DB - Banco de Dados.

BI - Business Intelligence.

CSV - comma-separated-values.

DW - Data Warehouse.

DWH - Data Warehousing.

ETL - Extração Transformação e Carregamento.

FK - Foreign Key.

HDF5 - Hierarchical Data Format.

IP - Internet Protocol.

KDD - Knowledge-Discovery in Databases.

OMR - Object Relational Mapper.

ORDBMS - Sistema de Gerenciamento de Banco de Dados.

PIP - Package Manager.

PK - Primary Key.

SGBD -Sistema Gerenciador de Banco de Dados.

SESu - Secretaria de Ensino Superior.

SQL - Structured Query Language.

SaaS - Software as a Service.

SCD - Dimensão em Mudança Lenta.

TCP - Transmission Control Protocol.

UFG - Universidade Federal de Goiás.

UNIX - UNiplexed Information and Computing System.

## **Resumo**

Este trabalho concentrou-se na construção de um *Data Warehouse* mais adequado, para o desenvolvimento deste compreendeu-se das etapas de análise e avaliação dos dados, assim como utilizar ferramentas auxiliares para a implementação. Foram analisadas curvas de cargas dos clientes e através dessa análise, observou-se o perfil de consumo dos mesmos, embasados na análise foram aplicados algoritmos para a realização do ETL. Além disso, foi necessário estudos nas ferramentas PostgreSQL, Power BI, linguagem de programação para realizar alguns processos de ETL. Os resultados obtidos validam o modelo desenvolvido e implementado permitindo sua utilização na UFG, sendo utilizado como mais uma ferramenta de Gestão para auxiliar e somar-se as ações já existentes na UFG.

### **Palavras-chave**

Data Warehouse, Extração Transformação e Limpeza, Business Intelligence, Energia Elétrica

## **Abstract**

This work focused on the construction of a more adequate Data Warehouse, for its development it comprised the steps of data analysis and evaluation, as well as using auxiliary tools for implementation. Customers' load curves were analyzed and through this analysis, their consumption profile was observed, based on the analysis were applied algorithms to perform the ETL. In addition, studies on the PostgreSQL, Power BI, programming language tools were required to perform some ETL processes. The obtained results validate the model developed and implemented allowing its use in UFG, being used as another management tool to assist and add to the existing actions in UFG.

## **Keywords**

Data Warehouse , Business Intelligence, Extract Transform and Load, Electricity

# Sumário

Lista de Figuras	10
Lista de Tabelas	11
1 Introdução	12
1.1 Objetivos	13
1.2 Motivação da pesquisa	13
1.3 Metodologia	14
1.4 Estrutura	14
2 Fundamentos Teóricos	15
2.1 Business intelligence	15
2.2 Extração de conhecimento utilizando o processo de KDD	16
2.3 Data Warehouse	17
2.3.1 Tabelas de dimensões	17
2.3.2 Granularidade	19
2.3.3 Tabela Fato	19
2.3.4 Modelo estrela	21
2.3.5 Processo de ETL	21
Extração	22
Transformação	22
Carregamento	22
2.4 Python	23
2.4.1 Pycpg2	23
2.4.2 SQLAlchemy	24
2.4.3 Petl	24
2.4.4 Panda	25
2.5 PostgresSQL	25
2.6 Power BI	26
3 Implementação do estudo de caso	28
3.1 Pré-processamento	29
3.2 Modelagem do <i>Data Warehouse</i>	29
3.3 Utilização do esquema estrela	30
3.4 Script em python para realizar o ETL	33
4 Resultados experimentais	34
5 Considerações Finais	37
5.1 Trabalhos Futuros	38
Referências Bibliográficas	39

## Lista de Figuras

2.1	<i>Framework do Business Intelligence.</i>	15
2.2	Processo do KDD.	17
2.3	Modelo genérico em estrela.	21
2.4	Ciclo do processo de ETL.	21
2.5	Portal Phyton.org.	23
2.6	Portal PowerBI <i>service</i> .	27
3.1	Ferramenta utilizada para a modelagem do esquema estrela.	30
3.2	Tabela fato.	30
3.3	Detalhamento das tabelas no <i>PostgreSQL</i>	31
3.4	Desenvolvimento da modelagem dimensional no SGBD <i>PostgreSQL</i>	31
3.5	Esquema estrela do <i>Data Warehouse</i> .	32
4.1	Conexão do PowerBI com o DW.	34
4.2	Detalhe da figura 4.1 de conexão do PowerBI com o DW.	35
4.3	Relatório referente a temperatura por irradiância.	36
4.4	Gráfico referente ao valor da fatura pelo consumo no tempo.	36

## Lista de Tabelas

2.1	Alteração lenta dos tipos de dimensões [12]	18
2.2	Tipo de tabela fato [12]	20

## Introdução

Com o crescente aumento das tarifas nas faturas de energia elétrica, com a abertura do mercado de energia elétrica e com o aumento da competitividade no setor de distribuição de energia elétrica brasileiro, os clientes passaram buscar alternativas para redução do gasto com energia elétrica. Visando solucionar este problema foi criado um *Data Warehouse* (DW) que permita a detecção de desvio de comportamento no uso de energia elétrica, bem como estudos de comportamento de consumo. Pois quanto menos perde-se, menos precisa ser gerado, e menos se desperdiça recursos naturais.

Entre as Instituições Públicas de Educação Superior, encontram-se, atualmente, 63 (sessenta e três) Universidade Federais. Essas instituições desempenham um importante papel no desenvolvimento científico e tecnológico do país, respondendo por uma parcela significativa da produção científica brasileira e pela formação de profissionais nas mais diversas áreas do conhecimento. Contatou-se que, em grande parte destas instituições públicas, a conta de energia elétrica é uma de suas maiores despesas, quando elencados todos seus itens de pagamento mensal.

De acordo com a Secretaria de Ensino Superior (SESu) do Ministério da Educação, o valor total pago em 2015 apenas pelas Universidades Federais foi de cerca de R\$ 430.000.000,00 (quatrocentos e trinta milhões de reais). Segundo essa mesma fonte, as despesas com energia elétrica dessas instituições despontam como o terceiro maior grupo, representando cerca de 9% dos gastos apurados em 2015.

O Brasil é cortado ao sul pelo trópico de Capricórnio e apresenta 92% do território na zona intertropical (entre os trópicos de Câncer e de Capricórnio), com um clima tropical que exige a utilização massiva de aparelhos de ar condicionado, que se trata de um grande componente das contas de energia.

Por outro lado, essa mesma localização favorece a incidência de raios solares o que garante ao Brasil um grande potencial energético solar. Tal potencial, em geral, é aproveitado através de geração distribuída, onde cada consumidor instala em seu próprio terreno (em geral nos telhados) uma usina solar que pode ou não atender às suas necessidades integrais de energia.

Sabe-se, porém, que uma parte considerável desses gastos se deve ao uso de equipamentos ineficientes e a práticas inadequadas de instalação, uso e manutenção dos

aparelhos, entre outros inconvenientes da falta de uma cultura de uso eficiente e racional de energia no país.

Por outro lado, quando se trata do pagamento de ações de eficiência energética, principalmente quando envolve a compra e a substituição de equipamentos, vários obstáculos são verificados. Entre os principais entraves, destacam-se a falta de capital próprio para investimento e as dificuldades para obtenção de financiamento ou obtenção de recursos em outra(s) fonte(s).

Além da redução do valor da fatura de energia dessas instituições por meio da melhoria da eficiência no uso final de energia, há, também, a possibilidade de redução dos gastos com energia elétrica por meio da geração própria de energia (micro e mini geração).

No contexto, a redução dos custos de energia elétrica na UFG pode significar mais investimento em pesquisa com melhoria das instalações dos laboratórios e aumento do número de bolsas ofertadas.

Dada a importância econômica de tal modelo de fornecimento e as diversas nuances que permeiam a sua aplicabilidade, torna-se importante um estudo aprofundado do tema.

## 1.1 Objetivos

O estudo proposto por esse projeto pretende auxiliar em futuras análises mais aprofundadas referente aos dados relacionados a energia elétrica, envolvendo principalmente a parte gerencial. Para a construção desse projeto, foram empregadas ferramentas consideradas apropriadas utilizada no estudo de caso.

Os objetivos específicos deste projeto são:

- Avaliar os bancos de dados, verificando a necessidade de implantar alterações no *Data Warehouse* criado em um banco PostgreSQL.
- Integrar os esquemas das fontes de dados coletadas.
- Realizar o processo de extração, transformação e carregamento (ETL).
- Utilizar ferramentas auxiliares para a implementação.

## 1.2 Motivação da pesquisa

A motivação para o presente trabalho surge no interesse em melhorar a qualidade dos dados, a fim de serem feitos relatórios mais detalhados e com informações gerenciais que auxiliem na tomada de decisões. A implantação de usinas fotovoltaicas na UFG, cujo objetivo é a redução de gastos com energia elétrica, demanda a apresentação de soluções

mais aprimoradas para a utilização das informações abstraídas, resultando na construção de um *Data Warehouse* aplicado especificamente às necessidades da Universidade Federal de Goiás.

## 1.3 Metodologia

O desenvolvimento dessa investigação iniciou com o levantamento bibliográfico a cerca do tema, com objetivo de compor a base teórica para a pesquisa. A técnica empregada para a melhoria dos dados foi apresentada no caso de estudo, o qual teve como cenário toda a composição de dados referentes a energia elétrica da UFG, por possuir um ambiente considerado favorável para os propósitos do projeto.

Seguindo as etapas do DW, os esquemas das bases de dados passaram por um processo de integração. Uma vez integrados, a *Data Staging Area* e o DW puderam ser modelados e terem seus esquemas desenhados. Para a implementação, utilizou-se como o Sistema Gerenciador de Banco de Dados (SGBD), o PostgreSQL.

Para a extração, transformação e carregamento foi desenvolvido algoritmos. Além disso, alguns nomes a serem utilizados no decorrer da monografia será o seu nome em inglês nativo, como eles são mais conhecidos.

## 1.4 Estrutura

O presente trabalho está organizado em 5 capítulos, incluindo este introdutório. O Capítulo 2 caracteriza o objetivo de apresentar os principais conceitos envolvidos com o tema do trabalho, sob forma de uma revisão bibliográfica. No Capítulo 3 serão discutidos as tecnologias de banco de dados aplicados ao setor de energia elétrica na UFG, além de descrever o processo de criação e implementação do *Data Warehouse*. No Capítulo 4 será apresentado a validação do projeto, os resultados do estudo de caso. O Capítulo 5 caracterizada pela retomada as discussões gerais do trabalho com intuito conclusivo, finalizando o projeto com os resultados e contribuições relevantes, dificuldades encontradas e também indicações para trabalhos futuros. Por último será exposto todas as referências bibliográficas consultadas.

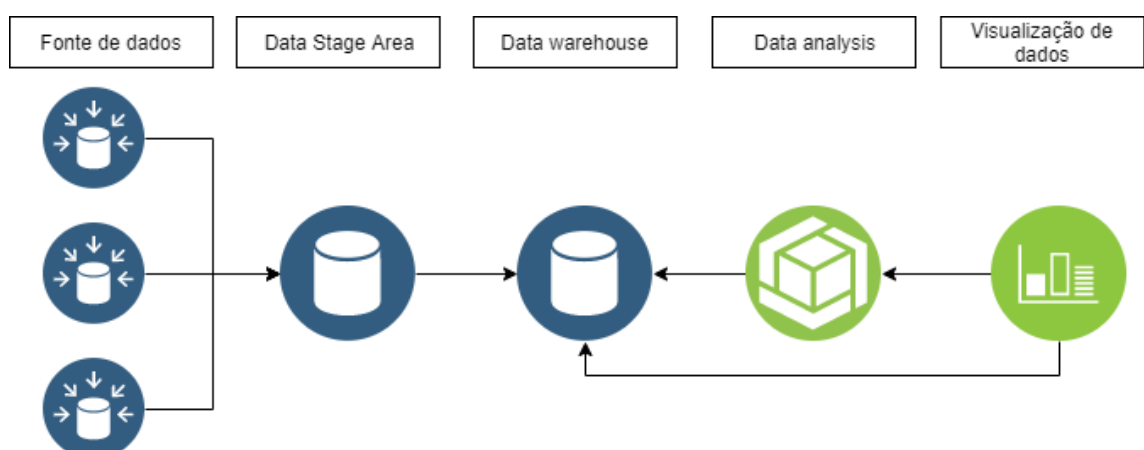
## Fundamentos Teóricos

Neste capítulo, será apresentado a fundamentação teórica para a realização da implementação de um Data Warehouse através do processo de ETL no PostgreSQL. Então, para a obtenção de informações relevantes para os usuários por meio de dados é essencial tê-los estruturados e limpos. O capítulo foi dividido em quatro seções: a primeira Seção 2.1 sobre *Business Intelligence* (BI), a segunda Seção 2.3 foi dedicada a descrever o *Data Warehouse*, a terceira Seção 2.4 para falar sobre as bibliotecas utilizadas do Python e a última Seção 2.6 desse capítulo são sobre a ferramenta de visualização dos relatórios.

### 2.1 Business intelligence

O *Business Intelligence*, segundo Solomon NegAsh e Paul Gray, são sistemas que combinam a coleta, armazenamento de dados e a gestão do conhecimento, com intuito de avaliar e extrair informações complexas. Esse, objetiva expor dados dos negócios, com finalidade de utilizarem-no para a obtenção de informações importantes e auxiliar nas decisões [7]. Observa-se que a modelagem do BI pode ser descrito conforme a figura 2.1.

**Figura 2.1:** *Framework do Business Intelligence.*



A explosão de dados, as instituições carecem de informações úteis para serem aplicadas, com intenção de gerar melhorias. Para isso, é essencial abranger "um conjunto de metodologias, processos, arquiteturas e tecnologias que transformam dados brutos

em informações significativas para permitir *insights* e tomadas de decisões estratégicas, táticas e operacionais mais eficazes", segundo disse a Forrester Research. Porém, mesmo com o número crescente de dados, a capacidade de gerar conhecimento através dela, pouco aumentou [20].

"Nossas redes estão inundadas de dados. Um pouco disso é informação. Uma pequena parte disso aparece como conhecimento. Combinado com ideias, algumas delas são realmente úteis. Misture experiência, contexto, compaixão, disciplina, humor, tolerância e humildade, e talvez o conhecimento se torne sabedoria"[13].

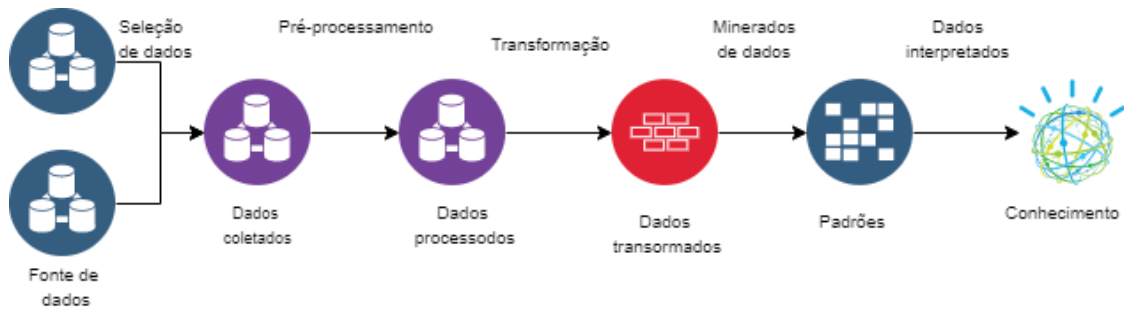
Portanto, para saciar a escassez de gerar conhecimento por meio de dados massivos obtidos, torna-se importante a implantação de ferramentas de BI nas instituições. Além disso, o BI carrega consigo diversos benefícios para encontrar as causas de certos problemas e também identificar e analisar os principais fatores de sucesso [10], como os descritos abaixo:

- Remoção de estimativas e adivinhações no momento de tomar decisões.
- Relatórios, análises e planejamento mais rápidos e precisos.
- Obtenção de relatório importantes de métricas da instituição quando e onde precisar.
- Identifique oportunidades de *up-selling* e de *cross-selling*.
- Obter formas de simplificar as operações.
- Eficiência no desenvolvimento de melhorias na instituição.
- Exposição de gastos desnecessários e otimização de custos.
- Realizar inventários melhores.
- Compreender melhor o passado, presente e futuro da instituição.
- Aumento da qualidade de dados.

## 2.2 Extração de conhecimento utilizando o processo de KDD

O processo *knowledge-discovery in databases* (KDD) formalizado em 1989, devido a necessidade de extrair conhecimento a partir das bases de dados. Essa descoberta de conhecimentos envolve os passos ilustrados na figura 2.2. Este, trata-se de um processo iterativo, já que pode ser recommençado a quantidade de vezes que for julgado necessário para que seja obtido resultados esperados, também tem este nome por haver a necessidade de interação com outras equipes. Embora cada etapa do processo seja tratada independentemente, ainda existe grande dependência entre elas. Visto que, para ter uma extração de conhecimento adequada precisa-se que todos os processos anteriores sejam realizados.

**Figura 2.2: Processo do KDD.**



## 2.3 Data Warehouse

Data Warehouse pode ser traduzido como Armazém de Dados, ou seja, são considerados um tipo diferente de banco de dados. De acordo com o INMOM, o termo é definido como 'um depósito de dados orientado por assunto, integrado, não volátil, variável com o tempo, para apoiar as decisões"[11].

Segundo Ponniah, o *Data warehousing* não é simplesmente um produto de hardware ou software adquirido para o fornecimento de informações estratégicas. Mas, um ambiente computacional na qual, os usuários são inseridos diretamente com os dados necessário na tomada de decisão [16].

A estrutura de armazenamento de informações decisivas é denominado de *Data Warehouse*. Na qual, tem como objetivo, solucionar problemas referente a análise de dados. Percebe-se, a necessidade de integrar dados de fontes heterogêneas, para suportar relatórios analíticos, consultas estruturadas, elaboração de *dashboards*, otimização dos dados.

O DW tem como design lógico, a composição do modelo em estrela, tabelas fato e tabelas de dimensões, a granularidade. A aplicação de processos de ETL e *Data Warehouse* possui a capacidade de aumentar a eficiência na integração dos dados, aprimoramento na consistência dos dados, na consulta e performance do sistema permissão de acesso oportuno aos dados.

### 2.3.1 Tabelas de dimensões

Dentro de um *Data Warehouse* precisa construir uma modelagem dimensional. Duas características importantes nesta estruturação são as tabelas Fatos e as tabelas Dimensões [6]. As tabelas Dimensões são os bits principais da estrutura, na qual se produz as medidas importantes. Essa possui característica qualitativa da informação e relacionamento, a fim de ser medidos. Um aspecto destas tabelas é a multiplicidade dos atributos, ou seja colunas filtradas, resumidas ou agregadas [22]. Além disso, contém uma

chave baseada na chave comercial, possui função de vinculação da tabela dimensão com a(s) tabela(s) fato(s).

Segundo Ralph Kimball cada tipo de técnica de dimensão implica em uma pequena mudança nas tabelas chamada de dimensão em mudança lenta (SCD) e as técnicas devem ser escolhidas com base no objetivo final da análise, conforme mostrado na tabela 2.1.

**Tabela 2.1:** Alteração lenta dos tipos de dimensões [12]

Tipo SCD	Ação da Tabela de Dimensões	Impacto na Análise de Fatos
Tipo 0	Os valores dos atributos não mudam	Fatos associados com valores de atributos originais
Tipo 1	Valores de atributos sobrescritos	Fatos associados com valores de atributos atuais
Tipo 2	Adiciona nova linha de dimensão ao perfil com novos valores de atributos	Fatos associados a valores de atributos em vigor quando fatos ocorrer
Tipo 3	Adiciona nova coluna para preservar os valores atuais e anteriores dos atributos	Fatos associados ao valor atual e anterior dos atributos
Tipo 4	Adiciona uma tabela mini-dimensão contendo atributos que mudam rapidamente	Fatos associados a atributos que mudam rapidamente em vigor quando o fato ocorreu
Tipo 5	Adicione a mini-dimensão do tipo 4, juntamente com as chaves da mini-dimensão do tipo 1 sobrescritas na dimensão base	Fatos associados a atributos que mudam rapidamente em vigor quando o fato ocorreu, mais atributos atuais que mudam rapidamente.
Tipo 6	Adiciona atributos de sobrescrita do tipo 1 para linhas dimensionais do tipo 2 e sobrescreve todas as linhas das dimensões anteriores	Fatos associados a valores de atributos em vigor quando o fato ocorreu, mais os valores anteriores
Tipo 7	Adicionar linha de dimensões do tipo 2 com novos valores de atributos, mais visualizar limites de linhas anteriores e/ou valores de atributos	Fatos associados a valores de atributos em vigor quando o fato ocorreu, mais valores anteriores

Fonte:[12]

A qualidade e a profundidade do *primer* para modelagem dimensional pode empoderar o *Data Warehouse* na mesma proporção [12]. Ou seja, o tempo dedicado no fornecimento de atributos detalhados, garante a bons valores na coluna de atributos, melhora a qualidade do DW. Diversas vezes, na coleta dos dados, não temos conhecimento se os dados de produção é um fato ou dimensão atributo.

### 2.3.2 Granularidade

Um dos aspectos importantes dentro do *Data Warehouse* é a análise da granularidade dentro das tabelas fatos. Esse refere-se ao nível de detalhamento dos dados armazenados. Segundo PONNIAH, a granularidade está ligada diretamente ao volume de informações armazenadas e o tipo de consultas que podem ser pelo usuário de um DW.

O grão é o menor nível da informação e é definido de acordo com as necessidades descritas no início do projeto. Ele é determinado para cada tabela Fato, já que normalmente as Fatos possuem informações e granularidades distintas.

O relacionamento entre o detalhamento e a granularidade é importante para ser compreendido. Quando falamos de menor granularidade (fina), significa maior detalhamento (menor sumarização) dos dados. Maior granularidade (grossa), significa menor detalhamento (maior sumarização). Assim podemos notar que a granularidade e o detalhamento são inversamente proporcionais.

Deve ser avaliado o equilíbrio entre detalhamento e sumarização para que a granularidade seja modelada com a melhor eficiência e eficácia para as consultas dos usuários, sempre levando em consideração as necessidades levantadas no começo do projeto.

### 2.3.3 Tabela Fato

Primeiramente, foi iniciado a definição do sistema de data warehousing (DWH), obteve toda a descrição das tabelas dimensão, em acordo com os requisitos pré-determinados. Agora, precisamos definir a tabela fato, estas são as medidas dos dados [21]. Responsável por agregar os dados em dimensões combinadas. Os dados numéricos nela contidos, corresponde a um evento. Somente serão permitidos, fatos consistentes com o grão declarado [13].

A tabela fato proporciona que os usuários acessem um repositório único na obtenção de informações. A medição corresponde a cada linha da tabela fato e estão a um nível específico de detalhamento chamado de granularidade [13]. De acordo com o Kimball uma das medidas mais importantes na tabela fato são os valores aditivos, como os valores financeiros. Algumas vezes, os fatos podem ser aditivos ou semi-aditivos. A primeira, não pode ser relacionado com o fator tempo para somar e a segunda nunca poderá ser somado, como valores únicos. Determinados valores, serão usados médias, medianas, desvio dentre outras medidas.

Ao projetar a tabela fato deve ter atenção para inserir os dados textuais sempre na tabela dimensões e os dados medidos na tabela fato. Redundância de dados pode ocasionar dificuldades durante a realização dos relatórios. Essas tabelas, em geral, consiste numa

divisão de atributos em dois grupos, são eles as chaves estrangeiras apontando para as tabelas dimensionais e as medidas [21].

**Tabela 2.2:** *Tipo de tabela fato [12]*

Tipos de tabela fatos	Vantagens	Desvantagem
Tabelas fatos de Transação	Detalhamento por meio da granularidade permite melhor monitoramento de atividades comerciais	Dificuldade de interpretação dos comportamentos e tendências em pontos fixos no tempo, isso gera problemas no desempenho na consulta
Tabelas periódicas de fatos instantâneo	Aumento de desempenho, visão longitudinal, comparação física	Granularidade mais baixa
Acumulando tabelas de fatos instantâneo	Permite análise complexa e de fluxo de trabalho	Alta complexidade de ETL
Tabelas fatos sem fatos	Simplicidade	Utilidade limitada

Fonte: [12]

Além disso, pode ter as colunas denominadas técnicas utilizadas nas auditorias e manutenção, ou seja, marcar a adição de novas atualizações ocorridas. Durante a projeção da tabela fato é importante considerar a dispersão na tabela, isto é, a relação entre a quantidade de linhas preenchidas e vazias. Pode-se usar a cardinalidade para calcular o valor de dispersão. Uma regra que pode ser usada para medir, de acordo com Aldo Zelen Arquiteto de Dados, é considerar o valor abaixo de 2% muito escasso e o oposto podem ser considerados relevantes. A equação  $1 - (f)$ , onde  $(f)$  é a razão do preenchimento.

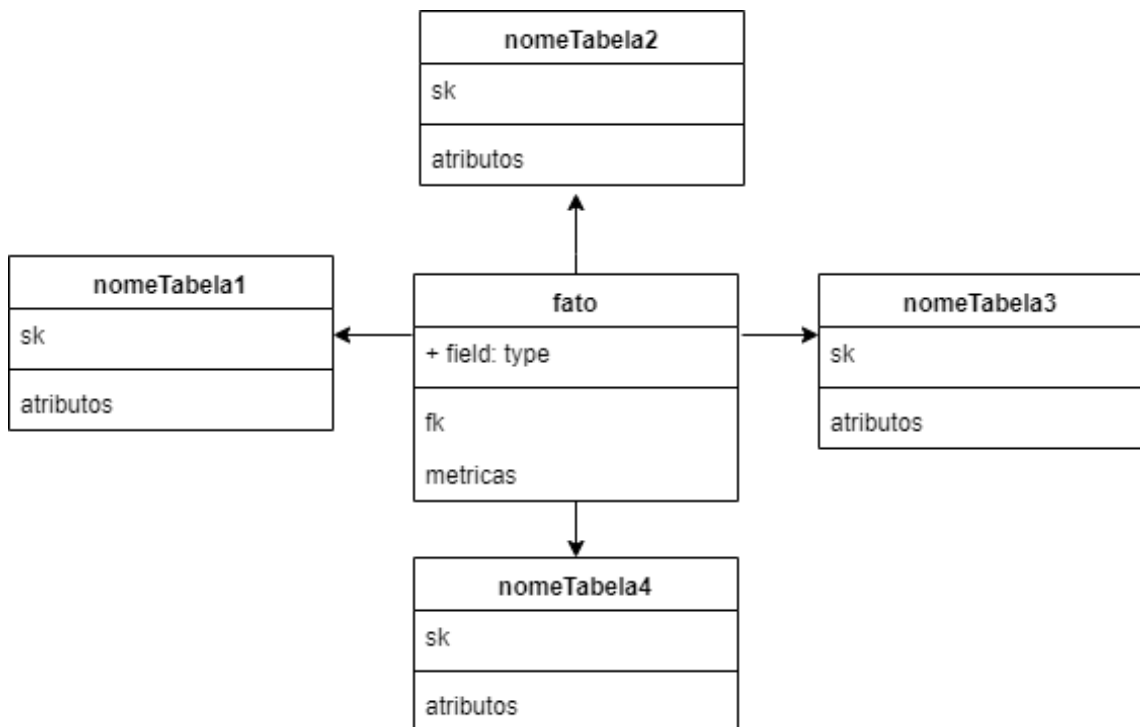
As vantagens e desvantagens e o tipo de tabela fato que podem ser escolhidas serão descritas na tabela 2.2 [21].

### 2.3.4 Modelo estrela

O modelo estrela consiste em pelo menos uma tabela fato, na qual faz referência ao número de tabelas de dimensão [7]. A tabela fato, possui pelo menos duas colunas contendo, respectivamente, as chaves estrangeiras e as medidas (fatos).

Normalmente, as medidas são caracterizadas pelas métricas relativas ao negócio. Algumas chaves na tabela fato não representa nenhuma chave em outra tabela. Além disso, as dimensões do modelo estrela são consideradas normalizadas.

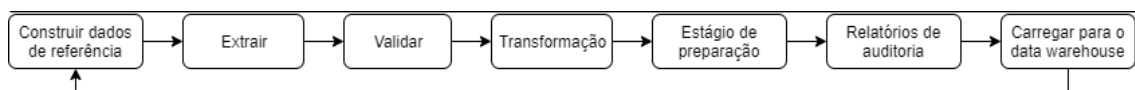
**Figura 2.3:** Modelo genérico em estrela.



### 2.3.5 Processo de ETL

No processo de gerenciamento do *Data Warehouse* é realizado o processo de extração, transformação e carregamento (ETL), antes de serem armazenados definitivamente[14]. Esta é uma fase importante no *Business Intelligence*, para a obtenção de relatórios consistentes e coerentes para análise dos dados. Na figura 2.4 indica todas as etapas que ocorre no ETL.

**Figura 2.4:** Ciclo do processo de ETL.



## Extração

A primeira etapa do processo é de extrair dados da fonte de origem até uma área intermediária, a fim de utilizá-los posteriormente em um ambiente de DW. Dentre todas as operações do ETL, a extração é uma das tarefas mais longas. Existem dois métodos de extração, extração completa e a extração incremental. Na completa não há o acompanhamento das alterações feitas nas fontes de origem e na incremental somente os dados alterados serão extraídos para o *Data Warehouse*. Além disso, existem mecanismos para que os dados sejam transportados até o DW, como o transporte usando arquivos simples, transporte através de operações distribuídas e também usando as *tablespaces* transportáveis [14].

## Transformação

A etapa final do ETL é o carregamento dos dados na estrutura de destino, essa pode se realizada de duas maneiras distintas. Sendo elas o carregamento completo que consiste na primeira carga de dados no DW em que são inseridos todos os elementos como as sub dimensões, tabelas bridge, tabelas fato, dimensões reduzidas, fatos agregados. Além disso, há o carregamento incremental na qual os dados são inseridos em intervalos regulares.

Nesta etapa, os dados extraídos e transformados são inseridos na estruturas para serem acessadas pelos usuários finais e aplicações [5]. O Carregamento é considerado uma parte simples do processo. De acordo com Doug Vucevic, a consideração final no processo de carga é atingir uma velocidade apropriada desse carregamento [20].

## Carregamento

A etapa final do ETL é o carregamento dos dados na estrutura de destino, essa pode se realizada de duas maneiras distintas. Sendo elas o carregamento completo que consiste na primeira carga de dados no DW em que são inseridos todos os elementos como as sub dimensões, tabelas bridge, tabelas fato, dimensões reduzidas, fatos agregadas. Além disso, há o carregamento incremental na qual os dados são inseridos em intervalos regulares.

Nesta etapa, os dados extraídos e transformados são inseridos na estruturas para serem acessadas pelos usuários finais e aplicações [5]. O Carregamento é considerado uma parte simples do processo. De acordo com Doug Vucevic, a consideração final no processo de carga é atingir uma velocidade apropriada desse carregamento [20].

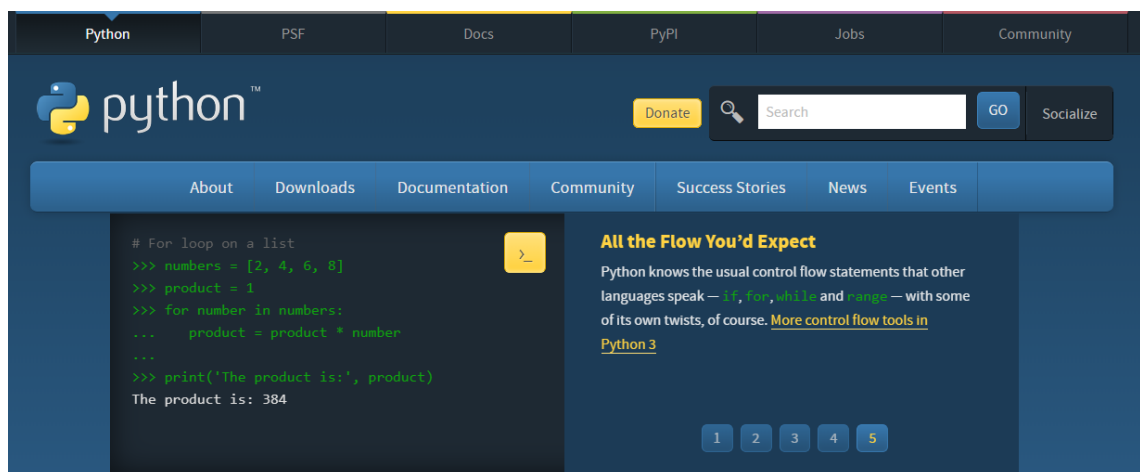
## 2.4 Python

Na análise de dados, uma das principais linguagens usadas é o Python, pois ele fornece uma diversidade de benefícios como a facilidade de uso e flexibilidade [18], sintaxe concisa, coerente.

Esse é usado para simplificar complexos conjuntos de dados, a fim de conceder sentido aos dados massivos. As ferramentas de análise de dados auxiliam nas avaliações de desempenho nos negócios, através da abstração de informações relevantes. A linguagem é capaz de compreender os padrões, a correlação de dados massivos [18].

Além disso, com a popularidade do Python também alimenta um amplo e crescente ecossistema de bibliotecas. Essa podem ser importadas através do instalador de pacotes PIP [8]. A *frameworks* como o pandas é muito utilizado para a renderização e manipulação de dados.

Figura 2.5: Portal Phyton.org.



### 2.4.1 Pycpg2

O Pycpg é o adaptador mais popular para banco de dados PostgreSQL utilizado na linguagem de programação Python. As características primordiais para a implementação completa da especificação Python DB API 2.0 e a segurança do *thread* (vários *thread* podem compartilhar a mesma conexão) [4].

Além disso, foi planejado para aplicativos demasiadamente *multithread* com possibilidade de criação e destruição de muitos cursores e também são capazes de produzir muitos "INSERT" ou "UPDATE" simultâneos. Sabe-se que toda conexão Python aberta conserva um grupo de conexões reais (UNIX ou TCP / IP) no banco de dados [4]. Quando um novo cursor é empregado no código, não é necessário abrir uma nova conexão em vez disso, utiliza-se uma das conexões ignoradas no pool.

Isso possibilita que o Psycopg seja muito ágil em aplicativos cliente-servidor. Toda vez que um pedido do cliente chegar, cria-se um encadeamento de serviço. O Psycopg 2 é uma melhoria do Psycopy, esse foi implementado principalmente em linguagem C como um revestimento do libpq, isso resulta em mais rendimento e segurança.

Foi inserido cursores do lado do cliente e do servidor, a comunicação assíncrona e as notificações, o suporte "COPY TO / COPY FROM"[4]. Além disso, O sistema de adaptação de objetos flexíveis permite uma acomodação ampliada e personalizada aos tipos de dados PostgreSQL na linguagem. O Psycopg 2 também é compatível com Unicode e Python 3 [4].

### 2.4.2 SQLAlchemy

SQLAlchemy é uma biblioteca de mapeamento objeto-relacional SQL em código aberto desenvolvido para a linguagem de programação Python e disponibilizado sobre a licença MIT.

O SQLAlchemy fornece "um conjunto completo e bem conhecido de modelos de persistência de nível corporativo, projetado para eficiência e alta performance de acesso a banco de dados, adaptado em uma linguagem de domínio simples e Pytônica". Sua filosofia é que bancos de dados SQL se comportem cada vez menos como coleções de objetos, em que mais tamanho e performance comecem a importar, enquanto coleções de objeto se comportem cada vez menos como tabelas e linhas, no qual mais abstração começa a importar. Por esta razão o SQLAlchemy adotou o padrão mapeado de dados (como o Hibernante para Java) em vez do padrão de registro ativo usado por vários outros mapeados objeto-relacional[1]. Entretanto, *plugins* adicionais como Elixir e *declarative* permitem aos usuários desenvolverem usando sintaxe declarativa.

O SQLAlchemy foi lançado em fevereiro de 2006 e rapidamente se tornou uma das ferramentas de mapeamento objeto-relacional mais utilizadas juntamente com o ORM do Django, na comunidade Python.

### 2.4.3 Petl

O Pelt é um pacote utilizado para realizar os processos de extrair, transformar e carregar tabelas de dados. Além de realizar algumas das extrações necessárias para construir o *Data Warehouse*, de forma consistente. Pode-se citar a utilização do *fromdb* e *tobd* para realizar a cópia do banco de dados em uma área temporária.

## 2.4.4 Panda

A biblioteca de software pandas foi desenvolvida para a linguagem Python com a utilização na manipulação de dados e análise de dados em tabelas numéricas e séries temporais. Essa tem objetivo de tornar o código aberto mais poderosa e flexível disponível [15]. O pandas são adequados para vários tipos de dados:

- Dados tabulares com colunas heterogêneas.
- Dados de séries temporais.
- Dados de matriz arbitrária com rótulos de linha e coluna.
- Conjunto de dados observacionais e estatísticos.

A biblioteca pandas tem uma diversidade de vantagens, alguns deles são o manuseio fácil de dados ausentes, variação no tamanho devido a inserção e exclusão de colunas no *DataFame*, automatização explícito de dados, remodelagem e dinamização de dados, agrupamento por mecanismo, mesclagem de dados, hierarquia para rotulação nos eixos, carregamento de vários tipos de dados desde os simples (CSV) até o robusto (HDF5)[15].

## 2.5 PostgreSQL

O PostgreSQL é um sistema de gerenciamento de banco de dados objeto-relacional (ORDBMS) baseado no PostgreSQL, versão 4.21 [17], desenvolvido na Universidade da Califórnia no Departamento de Ciência da Computação de Berkeley.

O PostgreSQL foi pioneiro em muitos conceitos que só se tornaram disponíveis em alguns sistemas comerciais de bancos de dados muito mais tarde. O PostgreSQL é um sistema de gerenciamento de banco de dados objeto-relacional baseado no PostgreSQL, versão 4.21, desenvolvido na Universidade da Califórnia no Departamento de Ciência da Computação de Berkeley [9]. O PostgreSQL é um proveniente de código aberto desse código original de Berkeley. Ele suporta grande parte do padrão *Structured Query Language* (SQL) e oferece muitos recursos modernos: O PostgreSQL é um descendente de código aberto desse código original de Berkeley. Ele suporta grande parte do padrão SQL e oferece muitos recursos modernos:

- Consultas complexas.
- Chaves estrangeiras.
- Gatilhos.
- Visualizações atualizáveis.
- Integridade transacional.
- Controle de simultaneidade multiversão.

Além disso, o PostgreSQL pode ser estendido pelo usuário de várias maneiras, por exemplo, adicionando novos.

- Tipos de dados.
- Funções.
- Operadores.
- Funções agregadas.
- Métodos de indexação.
- Linguagens processuais.

## 2.6 Power BI

O Power BI é um conjunto de serviços, aplicativos e conectores. Ou seja, ele torna possível a conexão com seus dados em qualquer lugar que eles estejam, permite que você faça filtros, crie visualizações atraentes e compartilhe com quem quer que seja [19].

O Power BI é construído em cima dos seguintes blocos:

- Visualizações – Representações visuais dos dados;
- Conjunto de dados – Conjunto de dados utilizada pela ferramenta para criar as visualizações;
- Relatórios – Conjunto de visuais de um conjunto de dados;
- *Dashboards* – Conjunto de página única de visuais, criado com base em relatórios;
- Blocos – Única visualização encontrada em um *Dashboards*.

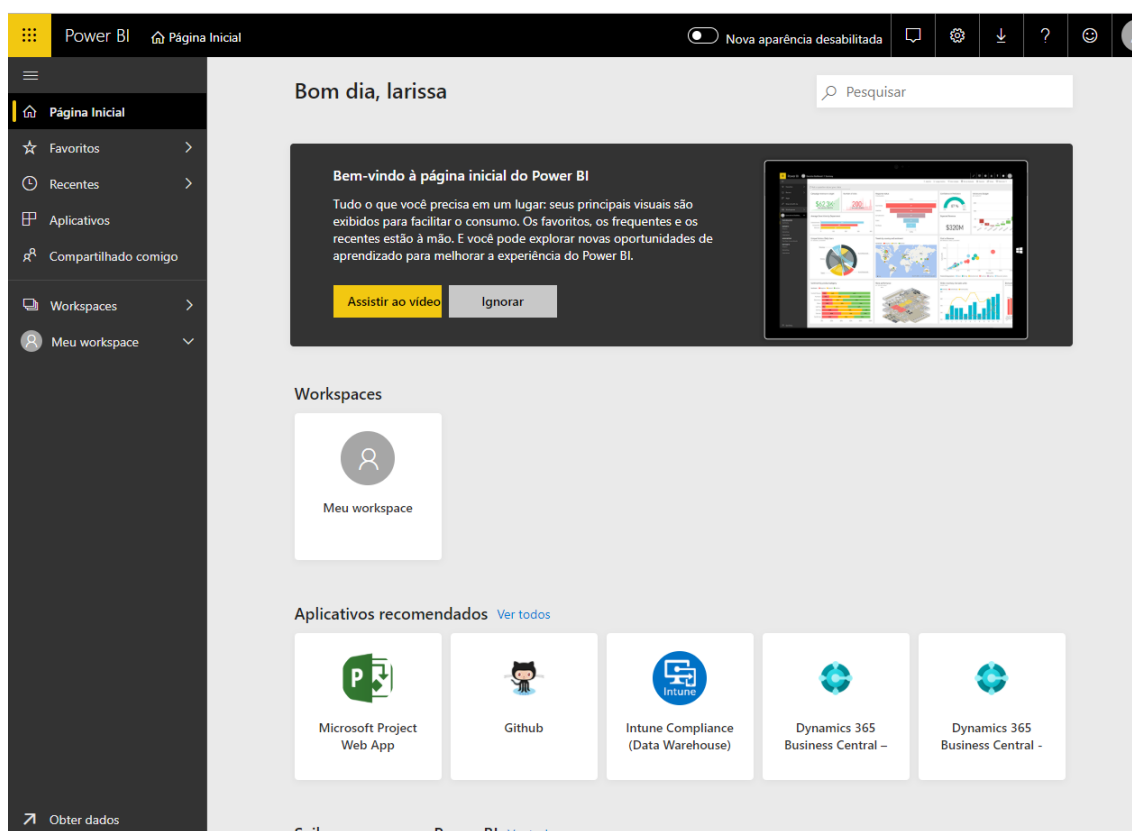
O fluxo comum de trabalho do Power BI é baseado na seguinte sequência: criar um relatório no - Power BI *Desktop*, publicá-lo no serviço do Power BI e compartilhá-lo com demais *stakeholders*, este é apenas um dos fluxos de utilização, o mais comum.

Neste fluxo, para criar um relatório, primeiro é necessário inserir os dados no Power BI. As fontes de dados disponíveis vão desde arquivos do Excel a bancos de dados. Além disso, o Power BI também se conecta a serviços de software, como provedores de *Software as a Service* (SaaS) ou serviços de nuvem como *Salesforce*, *Google Analytics* e Facebook [19].

Abaixo estão alguns dos pacotes de conteúdo para os tipos de serviços. Como os demonstrados na figura 2.6.

Além disso, é possível criar um *Dashboard* para visualização de solicitações *Pull* feitos no código do Github, os contribuidores, tempo gasto, etc. Além disso é possível fazer perguntas sobre seus dados por meio de linguagem natural. O Power BI se encarregará de criar uma visualização em tempo real com as perguntas mais realizadas [19].

Figura 2.6: Portal PowerBI service.



### Implementação do estudo de caso

Neste capítulo, será apresentado a implementação da construção do *Data Warehouse*. Para isso, foi apresentada no capítulo 2 a fundamentação teórica necessário para o desenvolvimento deste projeto.

A implementação deste modelo facilita a aplicação de fundamentos relacionados a análise dos dados. No decorrer deste capítulo, será descrito o processo de execução e aplicação do modelo nas seções 3.2 referente a modelagem do *Data Warehouse*, as diferenças de conexão para extrair os dados da fonte.

Um dos grandes desafios encontrados nas empresas de TI é o controle e o uso eficiente dos dados armazenados. A solução para sistemas desintegrados está em reorganizar a maneira como as informações são armazenadas, disponibilizada e acessada. Este cenário auxilia no desenvolvimento do *Data Warehouse*, na qual é descrito como um ambiente de suporte à decisão oportuna a ligar dados armazenados em fontes distintas, organizá-los e entregá-los aos responsáveis por tomar as decisões.

O conjunto de dados é composto por informações internas à a UFG, por dados históricos em um nível de granularidade. O SGBD no ambiente de apoio à decisão tem a função de permitir que usuários definam, construam e manipulem os dados integrados e compartilhados no banco de dados. Uma das representações do SGBD é a agrupamento de uma variedade de arquivos, que de outra forma, seriam diferentes, a fim de eliminar total ou parcialmente a redundância entre dados.

No projeto fotovoltaico, diversas pesquisas estão sendo realizadas, a fim de gerar conhecimento acerca dos verdadeiros benefícios que essa iniciativa impacta na redução de gastos relacionados à energia. Além disso, procura-se saber diversas informações referentes a quais ambientes dentro da universidade ocorre mais gastos em quais equipamentos, o que pode ser trocado.

Portando, para a realização desse objetivo é necessário dados obtidos através dos equipamentos, de faturas, do tempo sejam extraídos e transformados para que esses sejam utilizados para obter conhecimento acerca de todo processo. Então, primeiramente foi necessário conectar aos bancos de dados extraídos pelo Rafael Ratachesky em [3] através dos sistemas de medição da enel, sistema de monitoramento da UFG.

Além dos dados das estações meteorológicas da UFG inseridas dentro do kibana. Estes dados serão unificados em um único sistema para que sejam acessados de forma

mais viável para a abstração de informações. Através da criação de uma DW denominado Husky, e serão descritos abaixo.

## 3.1 Pré-processamento

Na etapa de pré-processamento e transformação tem o objetivo de detectar erros de cadastro e inconsistência de dados extraídos do ambiente operacional. é realizado a limpeza dos dados a fim de adequar e carregar apenas os dados necessário no *Data Warehouse*. Houve limpeza e transformação dos dados em que se encontravam nulas (em branco), algumas colunas precisaram ter seu tipo de dado alterado.

## 3.2 Modelagem do *Data Warehouse*

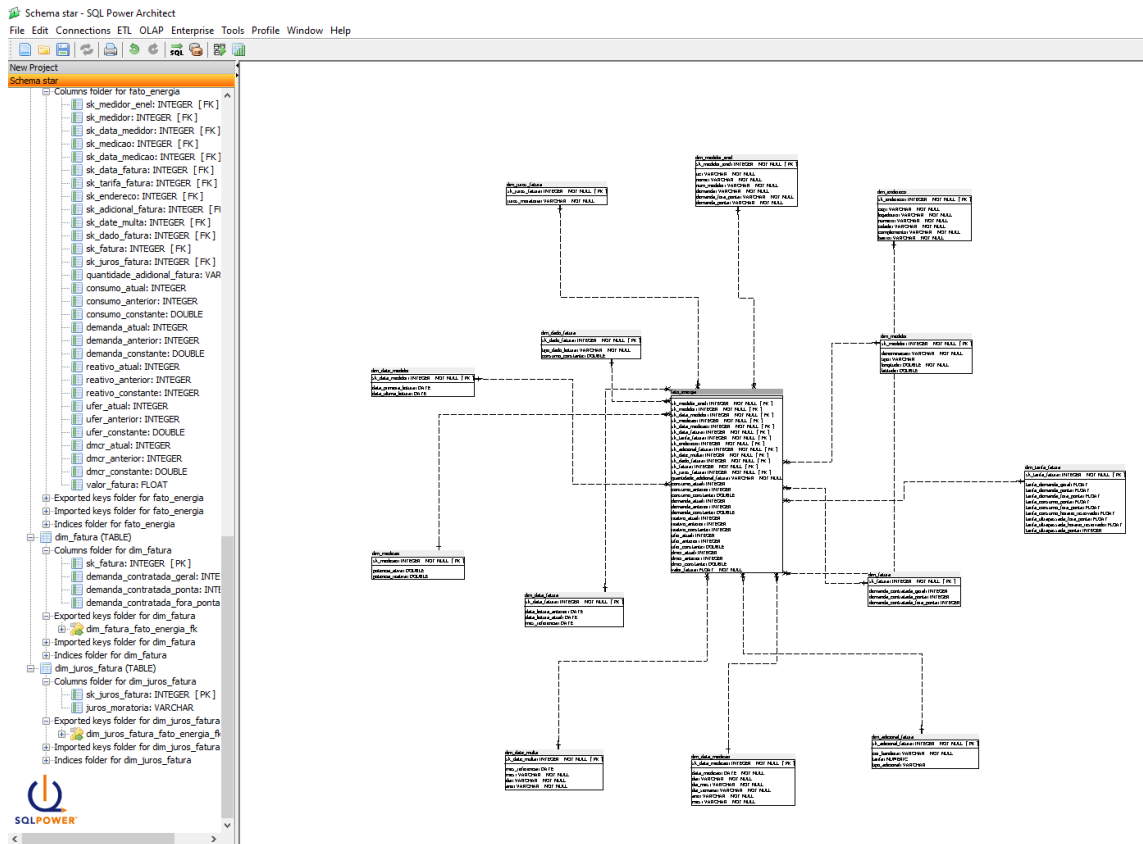
Inicialmente foi desenvolvido um banco de dados onde todos os atributos foram inseridos em algumas tabelas. Em seguida, o *Data Warehouse* deste trabalho foi implementado no SGBD PostgreSQL, utilizando a modelagem dimensional do esquema em estrela. Este *Data Warehouse* corresponde aos dados do setor de energia da UFG. A ferramenta utilizada para o apoio a decisão é um software que auxilia na manipulação de dados extraídos do *Data Warehouse*. A ferramenta utilizada neste trabalho foi o SQL Power Architect: ferramenta responsável pela criação das tabelas de fatos, dimensões e métricas do esquema dimensional em estrela. Neste trabalho foi criado uma tabela fato energia. A figura 3.1 ilustra o esquema dimensional do esquema estrela. 3.1

Foi realizado coleta dos requisitos para a modelagem do *Data Warehouse*, produziu-se as tabelas dimensões como descrito na seção 2.3.1 do capítulo 2. Em seguida, foi realizado a construção da tabela fato semelhante ao relatado na teoria do capítulo 2 na seção 2.3.3 e sua representação na figura 3.2. Vê-se necessário interligar as tabelas em esquema estrela como referenciado na seção 2.3.4, então o desenho esquemático está descrito na figura 3.5.

A modelagem dimensional do *Data Warehouse* desenvolvida neste trabalho foi implementada fisicamente no SGBD relacional PostgreSQL conforme ilustra a figura 3.4 e seu detalhamento na figura 3.3.

A modelagem lógica em modelo estrela conforme ilustrado na figura 3.5 possui uma tabela fato "energia" e 14 tabelas dimensões dentre elas: tarifa fatura, medidor enel, medidor, data dimensão, entre outras.

**Figura 3.1:** Ferramenta utilizada para a modelagem do esquema estrela.














**Figura 3.2:** Tabela fato.

fato_energia	
sk_medidor_enele	INTEGER NOT NULL [FK]
sk_medidor	INTEGER NOT NULL [FK]
sk_data_medidor	INTEGER NOT NULL [FK]
sk_medicao	INTEGER NOT NULL [FK]
sk_data_medicao	INTEGER NOT NULL [FK]
sk_data_fatura	INTEGER NOT NULL [FK]
sk_tarifa_fatura	INTEGER NOT NULL [FK]
sk_endereco	INTEGER NOT NULL [FK]
sk_adicional_fatura	INTEGER NOT NULL [FK]
sk_data_multa	INTEGER NOT NULL [FK]
sk_dado_fatura	INTEGER NOT NULL [FK]
sk_fatura	INTEGER NOT NULL [FK]
sk_juros_fatura	INTEGER NOT NULL [FK]
quantidade_adicional_fatura	VARCHAR NOT NULL
consumo_atual	INTEGER
consumo_constante	DOUBLE
demandas_atual	INTEGER
demandas_anterior	INTEGER
demandas_constante	DOUBLE
reativo_atual	INTEGER
reativo_anterior	INTEGER
reativo_constante	INTEGER
ufer_atual	INTEGER
ufer_anterior	INTEGER
ufer_constante	DOUBLE
dmc_atual	INTEGER
dmc_anterior	INTEGER
dmc_constante	DOUBLE
valor_fatura	FLOAT NOT NULL

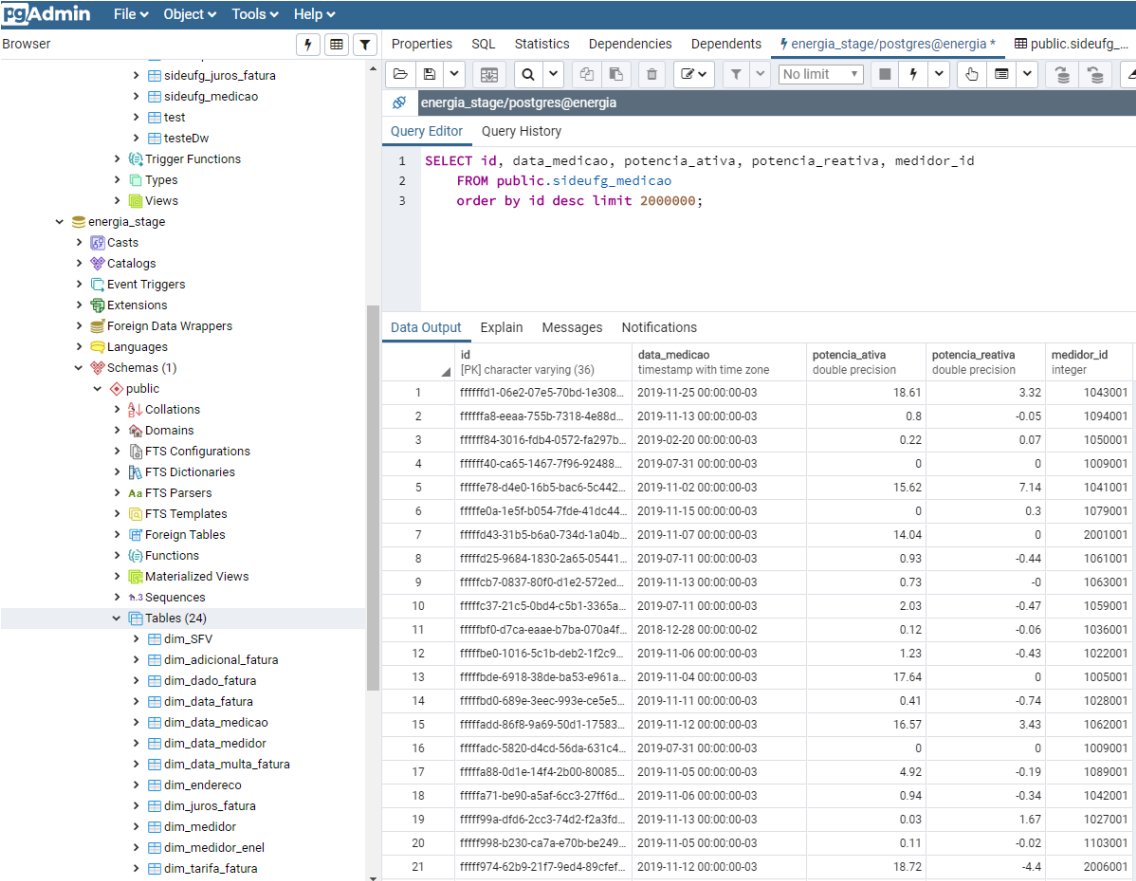
### 3.3 Utilização do esquema estrela

O principal motivo de usar o esquema em estrela para a modelagem do Banco de Dados é que ele favorece a otimização ao realizar as consultas no BD fazendo com que os comandos SQL resultantes tenham o mínimo de junções possível. Tendo em

**Figura 3.3:** Detalhamento das tabelas no PostgreSQL

- >  dim\_SFV
- >  dim\_adicional\_fatura
- >  dim\_dado\_fatura
- >  dim\_data\_fatura
- >  dim\_data\_medicao
- >  dim\_data\_medidor
- >  dim\_data\_multa\_fatura
- >  dim\_endereco
- >  dim\_juros\_fatura
- >  dim\_medidor
- >  dim\_medidor\_enel
- >  dim\_tarifa\_fatura

**Figura 3.4:** Desenvolvimento da modelagem dimensional no SGBD PostgreSQL



The screenshot shows the pgAdmin interface with a query editor and a data output table. The query is:

```

1 SELECT id, data_medicao, potencia_ativa, potencia_reativa, medidor_id
2 FROM public.sideufg_medicao
3 order by id desc limit 2000000;

```

The data output table shows the following columns and values:

	id	data_medicao	potencia_ativa	potencia_reativa	medidor_id
	[PK] character varying (36)	timestamp with time zone	double precision	double precision	integer
1	fffffd1-06e2-07e5-70bd-1e308...	2019-11-25 00:00:00-03	18.61	3.32	1043001
2	fffffa8-eeaa-755b-7318-4e88d...	2019-11-13 00:00:00-03	0.8	-0.05	1094001
3	fffff84-3016-fdb4-0572-fa297b...	2019-02-20 00:00:00-03	0.22	0.07	1050001
4	fffff40-ca65-1467-7f96-92488...	2019-07-31 00:00:00-03	0	0	1009001
5	ffffe78-d4e0-16b5-bac6-5c442...	2019-11-02 00:00:00-03	15.62	7.14	1041001
6	ffffe0a-1e5f-b054-7fde-41dc44...	2019-11-15 00:00:00-03	0	0.3	1079001
7	ffffd43-31b5-b6a0-734d-1a04b...	2019-11-07 00:00:00-03	14.04	0	2001001
8	ffffd25-9684-1830-2a65-05441...	2019-07-11 00:00:00-03	0.93	-0.44	1061001
9	ffffcb7-0837-80f0-d1e2-572ed...	2019-11-13 00:00:00-03	0.73	-0	1063001
10	ffffc37-21c5-0bd4-c5b1-3365a...	2019-07-11 00:00:00-03	2.03	-0.47	1059001
11	ffffbf0-d7ca-eeaa-b7ba-070a4f...	2018-12-28 00:00:00-02	0.12	-0.06	1036001
12	ffffbe0-1016-5c1b-deb2-1f2c9...	2019-11-06 00:00:00-03	1.23	-0.43	1022001
13	ffffbde-6918-38de-ba53-e961a...	2019-11-04 00:00:00-03	17.64	0	1005001
14	ffffbd0-689e-3eec-993e-ce5e5...	2019-11-11 00:00:00-03	0.41	-0.74	1028001
15	ffffad9-86f8-9a69-50d1-17583...	2019-11-12 00:00:00-03	16.57	3.43	1062001
16	ffffadc-5820-d4cd-56da-631c4...	2019-07-31 00:00:00-03	0	0	1009001
17	ffffa88-0d1e-14f4-2b00-80085...	2019-11-05 00:00:00-03	4.92	-0.19	1089001
18	ffffa71-be90-a5af-6cc3-27ff6d...	2019-11-06 00:00:00-03	0.94	-0.34	1042001
19	ffff99a-dfd6-2cc3-74d2-f2a3fd...	2019-11-13 00:00:00-03	0.03	1.67	1027001
20	ffff998-b230-ca7a-e70b-be249...	2019-11-05 00:00:00-03	0.11	-0.02	1103001
21	ffff974-62b9-21f7-9ed4-89cfe...	2019-11-12 00:00:00-03	18.72	-4.4	2006001

vista construir a tabela de fatos em esquema estrela otimizado para as consultas e que representasse claramente o setor energético da UFG, foram definidas as tabelas fatos e suas respectivas dimensões. Cada tabela de dimensão é definida com uma única chave primária (PK), baseada na integridade de relacionamento com a tabela fato, ou seja, a chave primária da dimensão é a chave estrangeira na tabela fato (FK). A modelagem da

**Figura 3.5:** Esquema estrela do Data Warehouse.

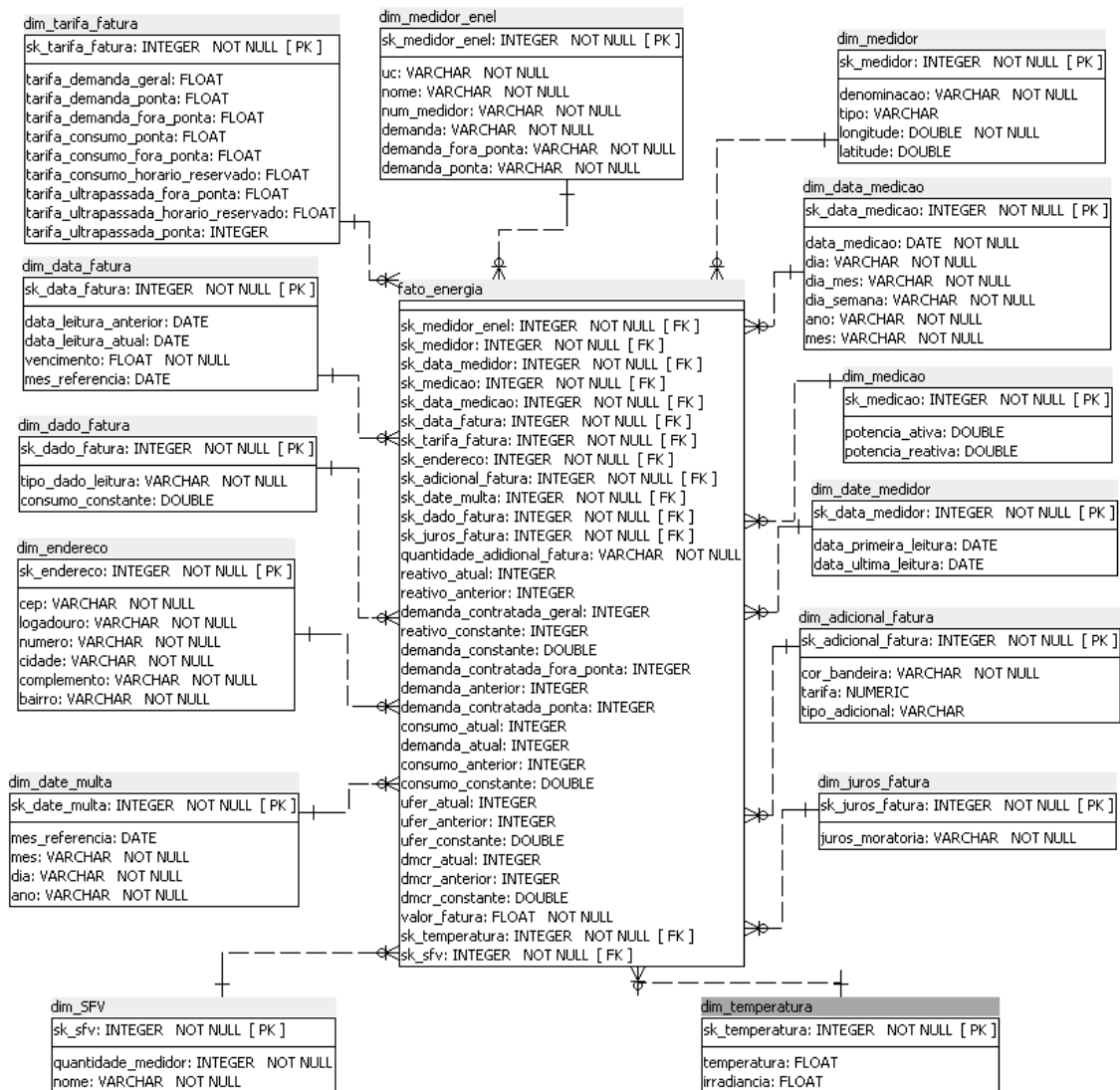


tabela fato e suas dimensões encontra-se ilustrada na figura 3.5.

## 3.4 Script em python para realizar o ETL

Para a realização do processo de ETL foi utilizado *Scripts* em Python juntamente com a biblioteca *pandas*. Essa biblioteca é uma das mais usadas para realizar esse tipo de processo. Além disso, foi usado o pacote *sqlalchemy* para fazer a conexão com o banco de dados. Inicialmente foi-se utilizado o *Psycopg2* para essa conexão, porém além de não poder ser usado junto com o *pandas* ele não possui alguns tipos de conversões como o *timestamp*.

A conversão das tabelas data foi usado *week* para semana, *year* para ano, *month* para mês, *day* para dia. E, dessa forma foi-se criado a tabela contendo os dados separados das datas para melhor utilização na parte de visualização. Além disso, foi excluído colunas de tabelas, criado novas tabelas.

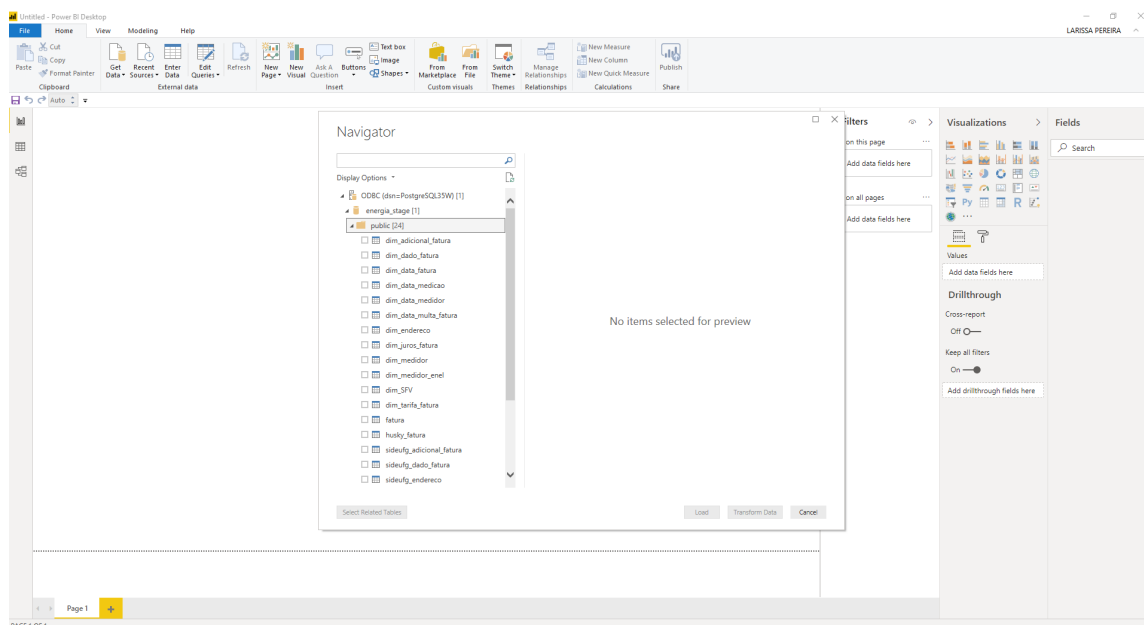
## Resultados experimentais

Neste capítulo vamos apresentar alguns testes feitos para verificar os componentes do *Data Warehouse*. Para efeito de análise, vamos criar dois gráficos, em que um terá as informações referentes a irradiância versus temperatura, e em outro gráfico os valores da tarifa versus valor da fatura versus medição.

Foram verificados a conexão de todos os dados através do DW desenvolvido. Além disso, todas as tabelas foram inseridas no DW. Alguns tratamentos foram necessários de ser realizados no próprio Power BI para finalizar a configuração correta de todos os dados. Os benefícios da utilização da ferramenta Power BI foram descritos no capítulo 2 seção 2.6. Nas ilustrações 4.3 4.4 4.1 demonstram que o processo de ETL foi realizado. E que esse facilita no processo de visualização.

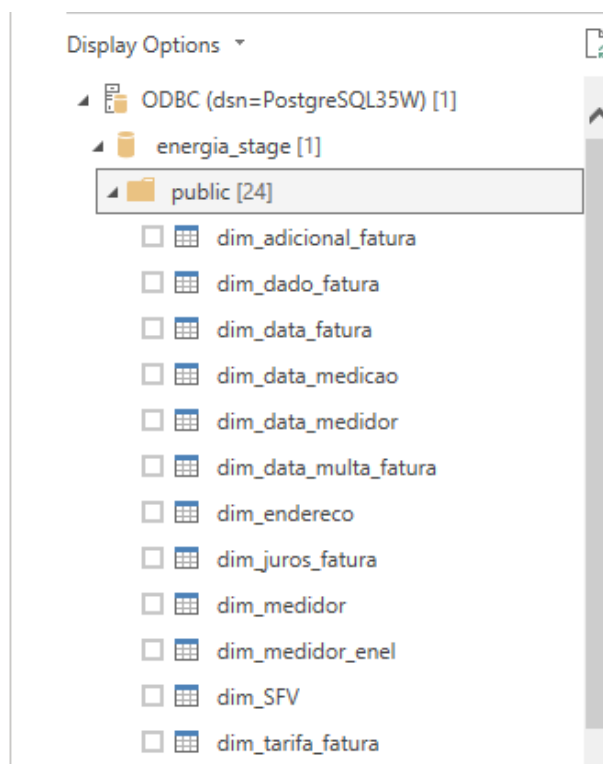
Na ilustração 4.1 mostra as tabelas criadas dentro do DW conectadas através do Power BI e na imagem 4.2 mostra com mais detalhes as tabelas do DW.

**Figura 4.1:** Conexão do PowerBI com o DW.



Ao executar o processo de extração dos dados da fonte foi percebido que o tempo de execução da tabela de medição foi consideravelmente longo, em razão da grande quantidade de linhas, isso também ocorreu ao extrair a tabela de temperatura.

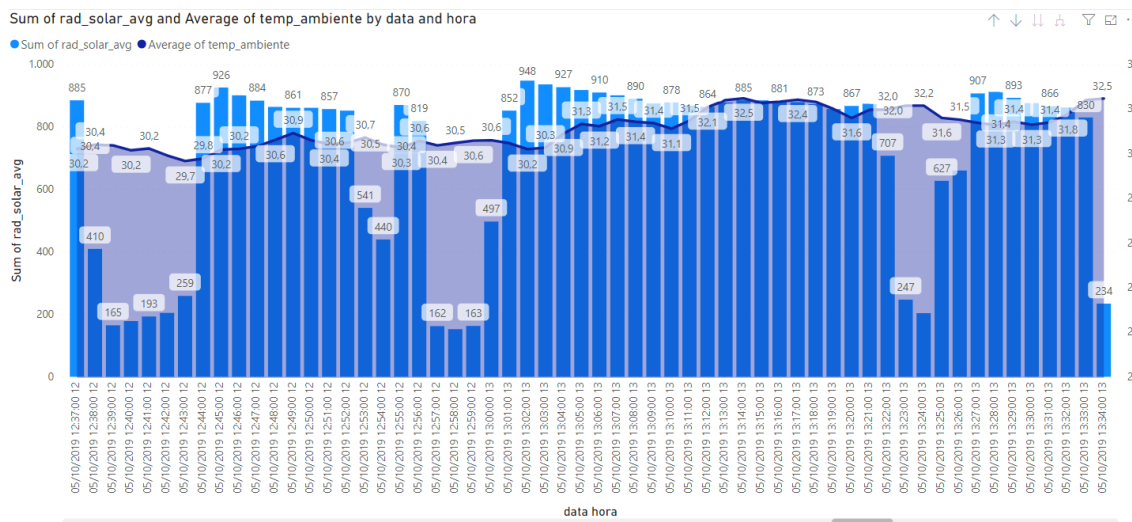
**Figura 4.2:** Detalhe da figura 4.1 de conexão do PowerBI com o DW.



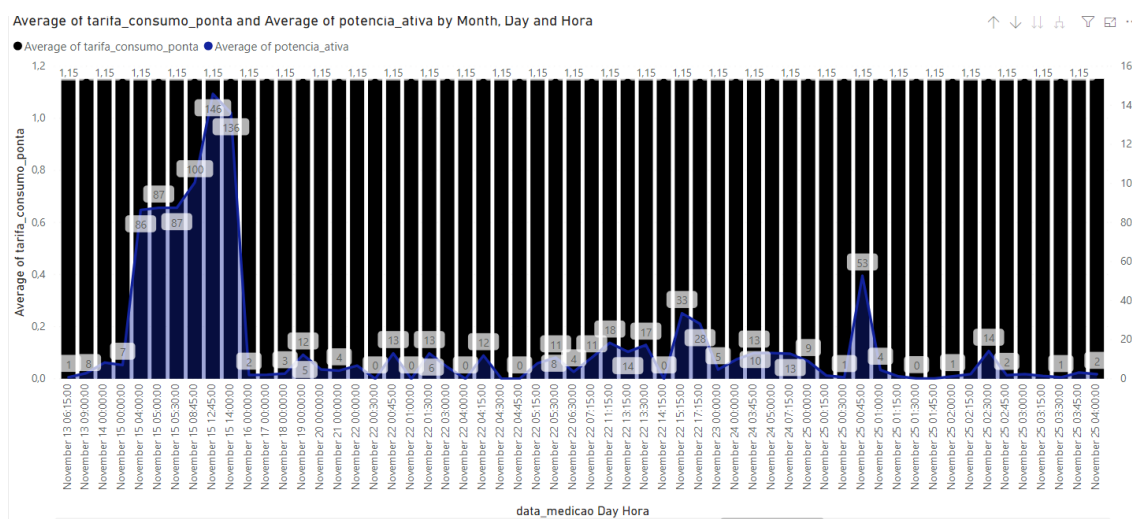
Além disso, para o preenchimento dos dados no DW foi necessário eliminar os campos Null e em alguns casos houve a necessidade de substituição por valores igual a zero ou não cadastrado ou mesmo pela média dos valores. Sabemos que os valores em branco dificultam no processo de análise dos dados.

No Gráfico 4.4 temos a representação dos dados da fatura referenciando os valores obtidos através das medições das placas solares. No gráfico 4.3 temos a representação dos dados em relação a temperatura no decorrer do tempo. Esse foi descrito de minuto em minuto. O gráfico criado para representar o valor em reais da produção de energia através das placas fotovoltaicas, isso foi demonstrado na figura 4.4.

**Figura 4.3: Relatório referente a temperatura por irradiância.**



**Figura 4.4: Gráfico referente ao valor da fatura pelo consumo no tempo.**



## Considerações Finais

Este trabalho apresentou resultados práticos e teóricos referente ao *Data Warehouse* e sua implementação. Foi avaliado os bancos de dados, verificando a necessidade de implantar alterações no *Data Warehouse* em um banco postgresSQL. Integrando os esquemas das fontes de dados das estações meteorológicas, Faturas e Medição.

Além disso, foram analisadas as variáveis de interesse para geração de gráficos que permitissem análises sobre o consumo e a geração de energia elétrica na UFG, resultando em uma maior otimização. Esse conjunto inicial de variáveis pode ser complementando, para isso é necessário acrescentar novos campos no mecanismo para inserção no DW criado inicialmente. Foi necessário realizar o processo de extração, transformação e carregamento (ETL).

A criação desse *Data Warehouse* (DW) permitiu a para detecção de desvio de comportamento no uso de energia elétrica, bem como estudos de comportamento de consumo. Também, tem como benefícios a facilidade no acesso aos dados, a garantia da integridade dos dados, facilidade para outros usuários acessarem os dados sem a necessidade de envolver-se diretamente com o banco de dados.

A formatação da data realizada dentro do banco de dados não foi interpretada pelo Power BI então foi-se necessário fazer a conversão dentro do próprio software. O Power BI se mostrou uma ferramenta robusta, para auxiliar na agregação de valor os dados e também na melhoria e refinamento da qualidade dos dados que não foram implementadas dentro do DW. Sendo essa melhor que a realização de relatórios de visualização utilizando o Excel que se vê muito incompleta para esse tipo de realização. Ao utilizar conjuntos de dados massivos é necessário realizar um carregamento indireto no Power BI, reduzindo assim o tempo de carregamento.

Os resultados validam o modelo desenvolvido e implementado permitindo sua utilização na UFG, sendo utilizado como mais uma ferramenta de Gestão para auxiliar e somar-se as a ações já existentes na UFG.

No entanto, a área do *Business Intelligence* é ampla e complexa, de difícil implementação. Principalmente, quando houve a necessidade de trabalhar com tabela com mais de cinquenta milhões de linhas, assim necessário outras tecnologias para o carregamento desses dados. Então para a realização deste trabalho foi utilizado a carga da tabela medição compondo um milhão de linhas para facilitar no carregamento.

## 5.1 Trabalhos Futuros

Uma primeira proposta para a implementação do DW é a busca de melhores ferramentas que tratem um grande volume de dados de forma otimizada para a criação desse. Já que os algoritmos usando no pandas se mostraram ineficientes, quando se trata de volume massivo de dados.

Outra proposta seria pesquisar a cerca de utilizar outros tipos de banco de dados para armazenar esse grande volume de dados. Como, por exemplo, o Hadoop e o Spark podem mostrar uma maior eficiência quanto a questão de tempo para executar as consultas e gravação dos dados.

Criar outras tabelas de relacionamento entre os bancos de dados com intuito de melhorar a análise de dados e visualização.

Por último, comparar ferramentas de análises a fim de verificar qual delas possuem melhor eficácia no tratamento de dados.

## Referências Bibliográficas

- [1] BAYER, M. **Gerrit is Awesome**. <https://techspot.zzzeek.org/>, 2016. Online; acessado em 25/10/2019.
- [2] BÍBLIA, A. T. **Provébios**. In: BÍBLIA., editor, *Bíblia Sagrada: Antigo testamento. Tradução em português por João Ferreira de Almeida. Revista Atualizada e Revisada, Segunda edição*, volume 21.872 de **NPI**, p. 867, Barueri-São Paulo: Sociedade Bíblica do Brasil, 2012.
- [3] DE SOUSA RAULINO, R. R. **Desenvolvimento de Sistema para gerenciar e auxiliar na tomada de decisão envolvendo consumo de energia elétrica**. UFG, Goiânia, GO, 2017.
- [4] DEVELOPED.; MAINTAINED BY THE PYTHON COMMUNITY, F. T. P. C. **Psycop2**. <https://pypi.org/project/psycopg2/>, 2019. Online; acessado em 04/11/2019.
- [5] EL-SAPPAGH, S. H. A.; HENDAWI, A. M. A.; BASTAWISSY, A. H. E. **A proposed model for data warehouse etl processes**. *Journal of King Saud University*, 23:91 – 104, 2011.
- [6] ELIAS, D. **Entendendo a Modelagem Multidimensional**. <https://canaltech.com.br/business-intelligence/entendendo-a-modelagem-multidimensional-19988/>, 2014. Online; acessado em 25/10/2019.
- [7] FOUCHÉ, G.; LANGIT, L. **Foundations of SQL Server 2008 R2 Business Intelligence, second edition**. Apress, United States of America, 2011.
- [8] GAL, R. **Power BI – Muito além de Business Intelligence**. <https://www.microsofttech.com.br/pt-br/microsofttech/power-bi-muito-alem-de-business-intelligence/>, 2018. Online; acessado em 25/11/2019.
- [9] GROUP, T. P. G. D. **PostgreSQL 11.5 Documentation**. University of California, Estados Unidos da America, 1996-2019.
- [10] HOČEVAR, B.; JAKLIČ, J. **Assessing benefits of business intelligence systems – a case study**. *Management*, 15:87–119, 2010.

- [11] INMON, W. H. **Building the Data Warehouse, Third Edition**. Robert Ipsen, New York, USA, 2002.
- [12] KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit, Third Edition**. John Wiley Sons, Inc, Canadá, 2013.
- [13] KOOMEY, J. G. **Turning Numbers into Knowledge**. quoting Clifford Stoll, Oakland, CA, 2001.
- [14] LANE, P. **Data Warehousing Guide**. [https://docs.oracle.com/cd/E11882\\_01/server.112/e25554/title.htm](https://docs.oracle.com/cd/E11882_01/server.112/e25554/title.htm), United States of America, 2007.
- [15] MANAGEMENT, A. C.; LLC.; FOUNDRY, L.; TEAM, P. D. **Visão geral do pacote**. [https://pandas.pydata.org/pandas-docs/stable/getting\\_started/overview.html](https://pandas.pydata.org/pandas-docs/stable/getting_started/overview.html), 2012. Online; acessado em 05/11/2019.
- [16] PONNIAH, P. **Data Warehousing Fundamentals**. JOHN WILEY SONS, INC, New York, USA, 2001.
- [17] POSTGRES, U. **University Postgres, Version 4.2**. <http://db.cs.berkeley.edu/postgres.html>, 1999. Online; acessado em 05/11/2019.
- [18] SULTAN, O. **Why is learning Python important in Data Science?** <https://www.hackread.com/why-is-learning-python-important-in-data-science/>, 2019. Online; acessado em 04/11/2019.
- [19] TECH, M. **Why Choosing Python For Data Science Is An Important Move**. <https://www.smartdatacollective.com/why-choosing-python-for-data-science-important-move/>, 2018. Online; acessado em 04/11/2019.
- [20] VUCEVIC, D.; YADDOW, W. **Testing the Data Warehouse Practicum Assuring Data Content**. traffo publishing, United States of America, 2012.
- [21] ZELEN, A. **Facts about Facts: Organizing Fact Tables in Data Warehouse Systems**. <https://www.vertabelo.com/blog/facts-about-facts-organizing-fact-tables-in-data-warehouse-systems/>, 2016. Online; acessado em 25/10/2019.
- [22] ZELEN, A. **Warehousing's Most Common Dimensional Table Types**. <https://www.vertabelo.com/blog/dimensions-of-dimensions-a-look-at-data-warehousings-most-common-dimensional-table-types/>, 2016. Online; acessado em 25/10/2019.