



Arquiteturas e Serviços de Data Lakes e Data Warehousing



PUC Minas
Virtual

Unidade 2 – Data Warehouse

O que é um sistema de BI

O que é BI?

Abrange os processos, ferramentas e tecnologias necessárias para transformar dados corporativos em informações e informações em conhecimento que podem ser usados para aprimorar a tomada de decisões e criar planos acionáveis que conduzem a uma atividade comercial eficaz.



O que é BI?

O BI pode ser usado para adquirir:

- ✓ Insights táticos para otimizar processos de negócios, identificando tendências, anomalias e comportamentos que requerem ação de gerenciamento.
- ✓ Visão estratégica para alinhar vários processos de negócios aos principais objetivos de negócios por meio de gerenciamento e análise de desempenho integrados.

O que é BI?

O valor das informações cresce proporcionalmente ao investimento em extração e análise de dados.

✓ Dados:

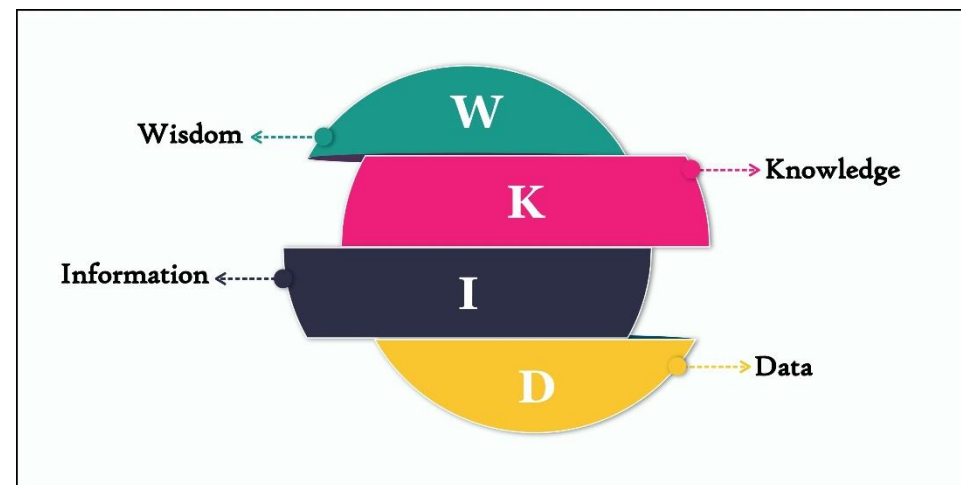
- ✓ Planilhas, sistemas, ERP, CRM, outros

✓ Informação: O que está acontecendo?

- ✓ Relatórios, gráficos e cruzamentos de dados.

✓ Conhecimento: Qual a minha estratégia?

- ✓ Sistemas de BI



O que é BI?

- ✓ Business Intelligence (BI) é algo como obter as informações certas, para os tomadores de decisão certos, no momento certo.
- ✓ BI é uma plataforma que suporta relatórios, análises e tomada de decisão.

O BI proporciona:

- ✓ Uma tomada de decisão baseada em fatos.
- ✓ Visão única dos dados.

O que é BI?

- ✓ Passado: O que e porquê aconteceu?
 - ✓ Análises diversas
- ✓ Presente: O que está acontecendo?
 - ✓ Performance, Indicadores, BSC
- ✓ Futuro: Qual a minha estratégia?
 - ✓ Está linkada ao planejamento estratégico.



O que é um Data Warehouse

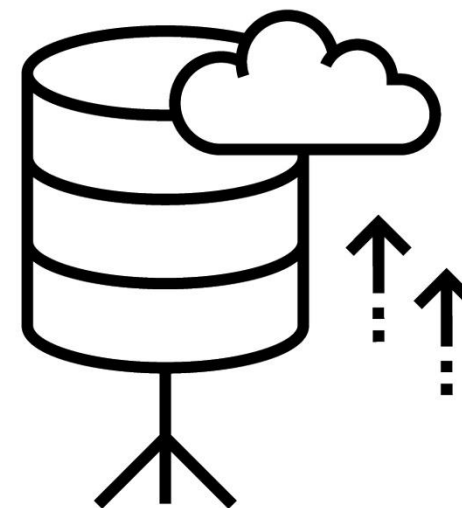
Data Warehouse - DW

- ✓ Data Warehouse é um ***depósito de dados digitais*** que serve para armazenar informações detalhadas relativamente a uma empresa, criando e organizando relatórios através de ***históricos*** que são, depois, usados pela empresa para ajudar a tomar decisões importantes com base nos fatos apresentados.



Data Warehouse - DW

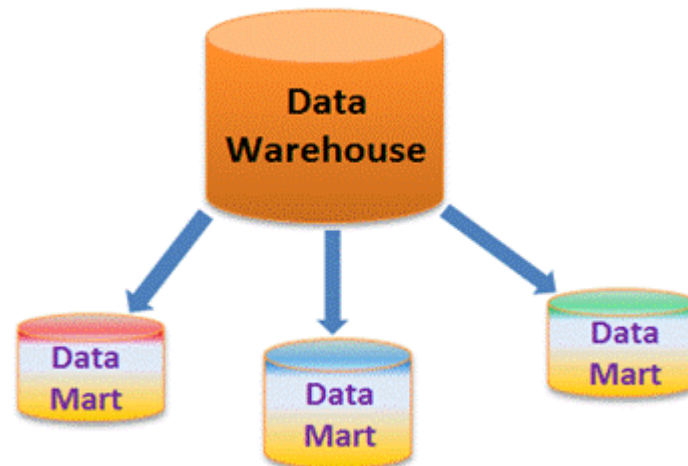
Todos os Data Warehouses compartilham um design básico no qual metadados, dados de resumo e dados brutos são armazenados no repositório central do Warehouse. O repositório é alimentado por fontes de dados em uma extremidade e acessado por usuários finais para análise, relatório e mineração na outra extremidade.



Data Warehouse

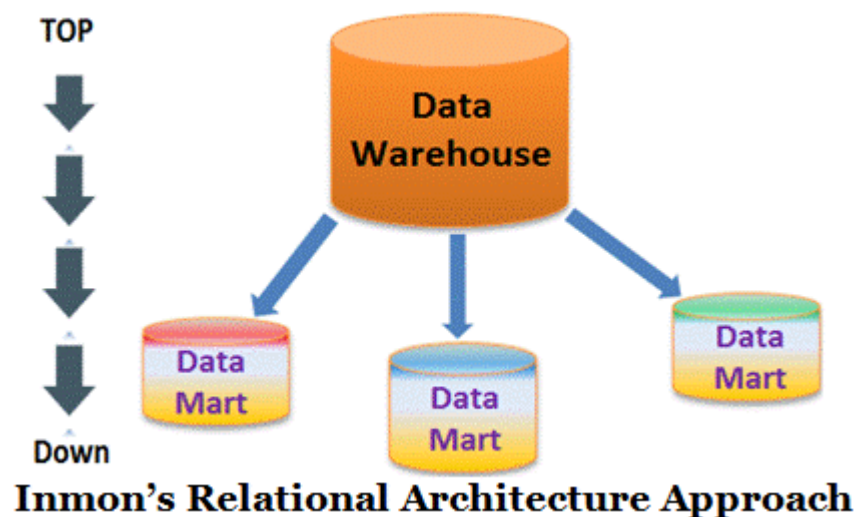
Data Mart

- ✓ Um Data Mart é uma **subdivisão ou subconjunto de um DW**. Os Data Marts são como pequenas fatias que armazenam subconjuntos de dados, normalmente organizados para um departamento ou um processo de negócio.
- ✓ Normalmente, o Data Mart **é direcionado para uma linha de negócios** ou equipe, sendo que a sua informação costuma pertencer à um único departamento.



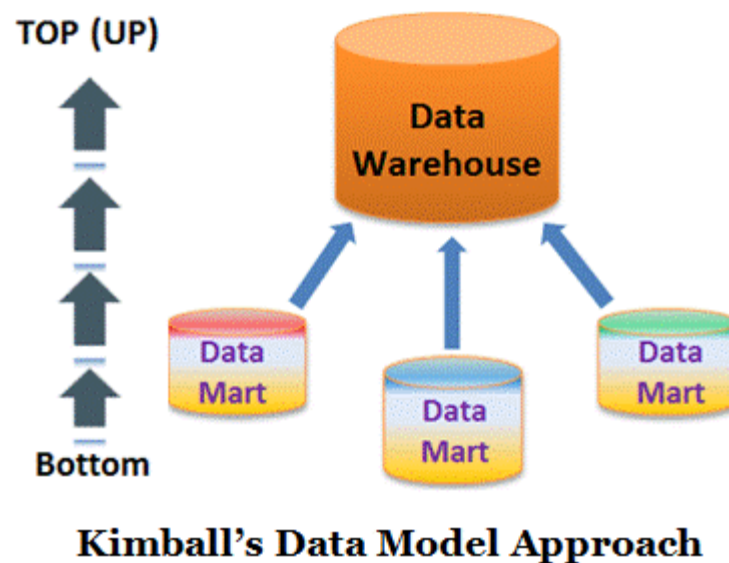
Abordagem Top - Down

- ✓ Bill Inmon defende uma abordagem de desenvolvimento de cima para baixo que adapta as ferramentas tradicionais de banco de dados relacional às necessidades de desenvolvimento de um data warehouse de toda a empresa.

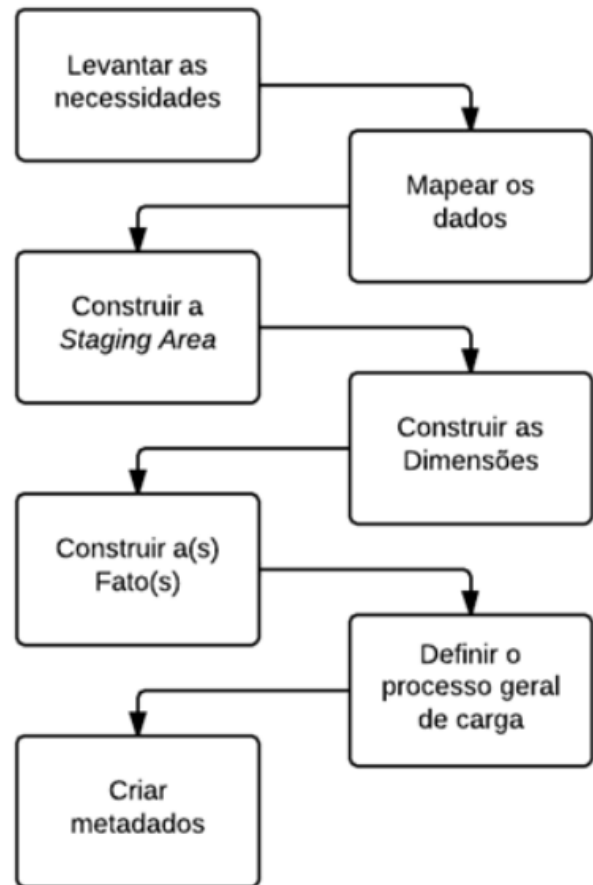


Abordagem Botton - Up

- ✓ Ralph Kimball recomenda uma abordagem de baixo para cima que usa modelagem dimensional, uma abordagem de modelagem de dados exclusiva para armazenamento de dados. Em vez de construir um único banco de dados corporativo, Kimball sugere a criação de um banco de dados (ou data mart) por processo de negócios principal.



Etapas na Construção de um DW

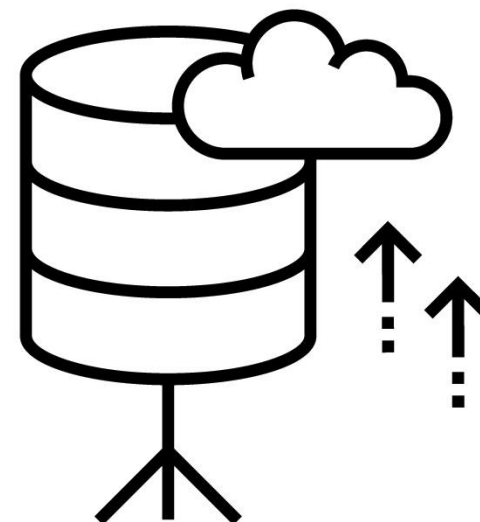


Fonte: <https://canaltech.com.br/infra/a-construcao-de-um-data-warehouse-em-7-etapas/>

Data Warehouse - DW

Data Warehouse é:

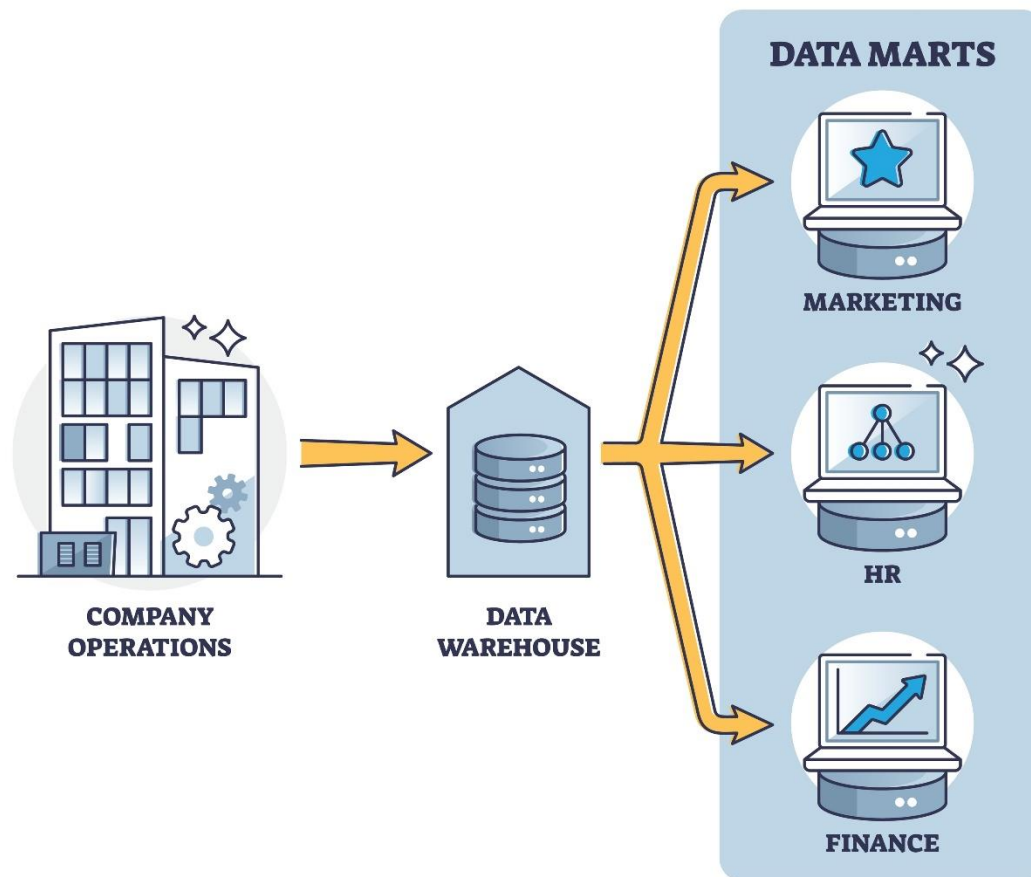
- Volatil ? Um Data Warehouse é uma coleção de dados orientados por assuntos, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo de tomada de decisão.
- Dados Futuros? Os dados na instância do Data Warehouse são dados processados e históricos.



Data Warehouse

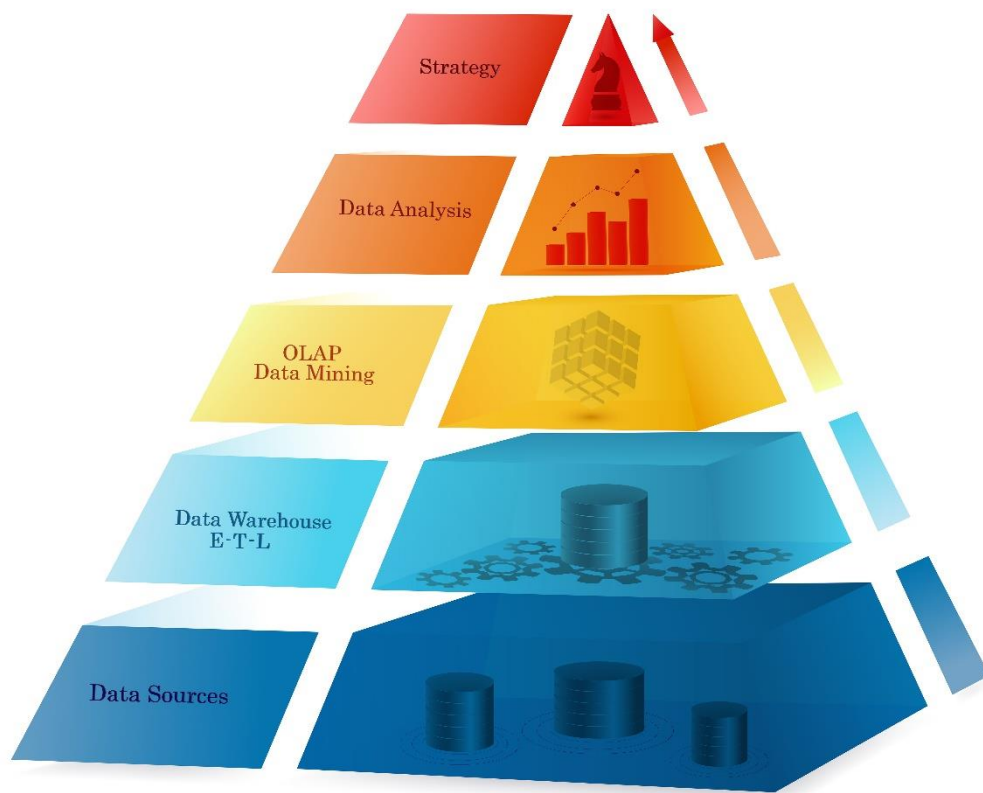
Data Warehouse - DW

- Read Only? Essa função deve ser reservada para o usuário ou usuários que precisam apenas revisar ou testar dados na instância do Data Warehouse sem permissões para fazer alterações. Acesse a instância do Data Warehouse no modo somente leitura. Leia todas as tabelas ou visualizações. Chame as funções que não modificam os dados.



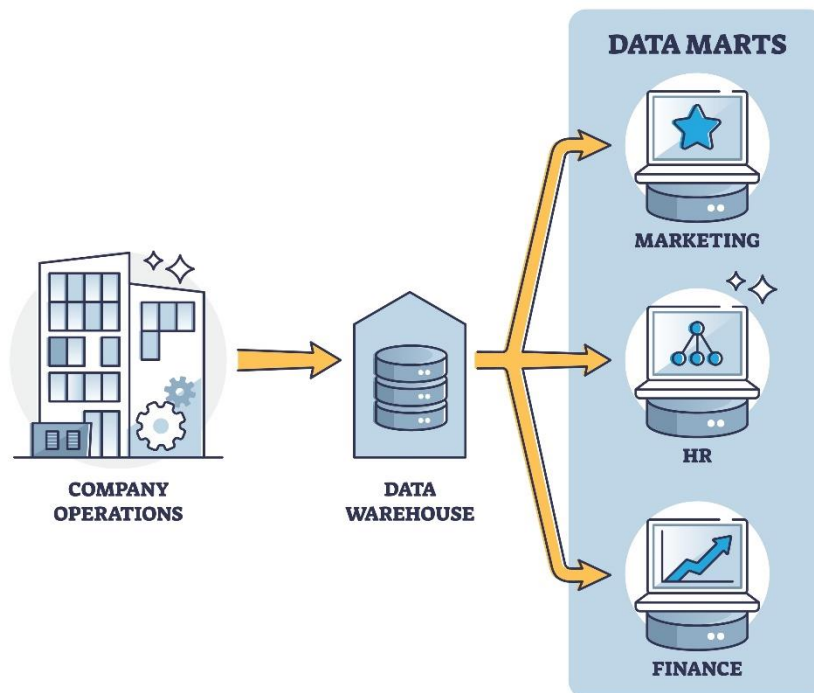
Vantagens do Data Warehouse

Estratégia



Inteligência de Negócios

- ✓ Processo de coleta, organização, análise, compartilhamento e monitoramento de informações para suporte a gestão de negócios.

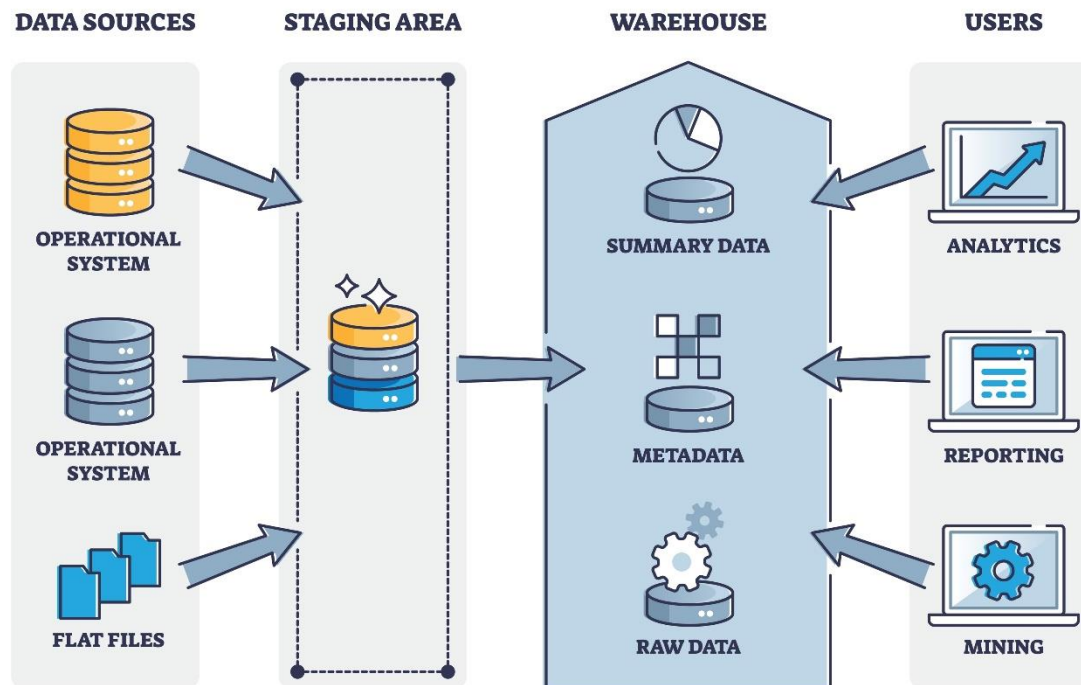


Como funciona um Data Warehouse?

- O desafio foi conseguir uma visão unificada entre sistemas, diferentes tecnologias, formatos de dados proprietários e controle de acesso.
- A arquitetura de dados se adaptou através de processos de ETL (Extract Transform Load), técnicas de modelagem, documentação/metadados e usando ferramentas e tecnologias proprietárias, otimizadas para rodar em servidores grandes e caros.

Como funciona um Data Warehouse?

DATA WAREHOUSE



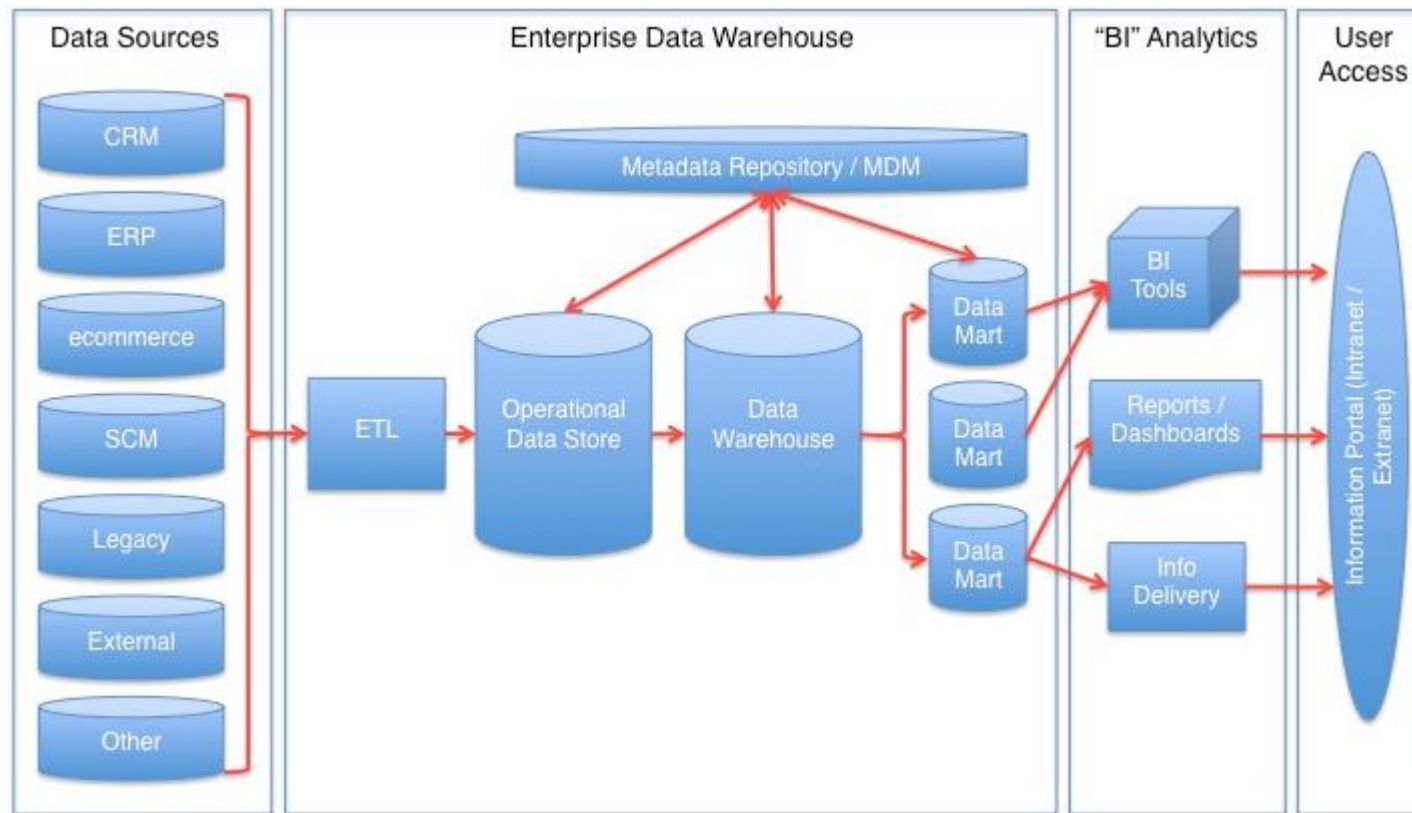
Como funciona um Data Warehouse?



ETL

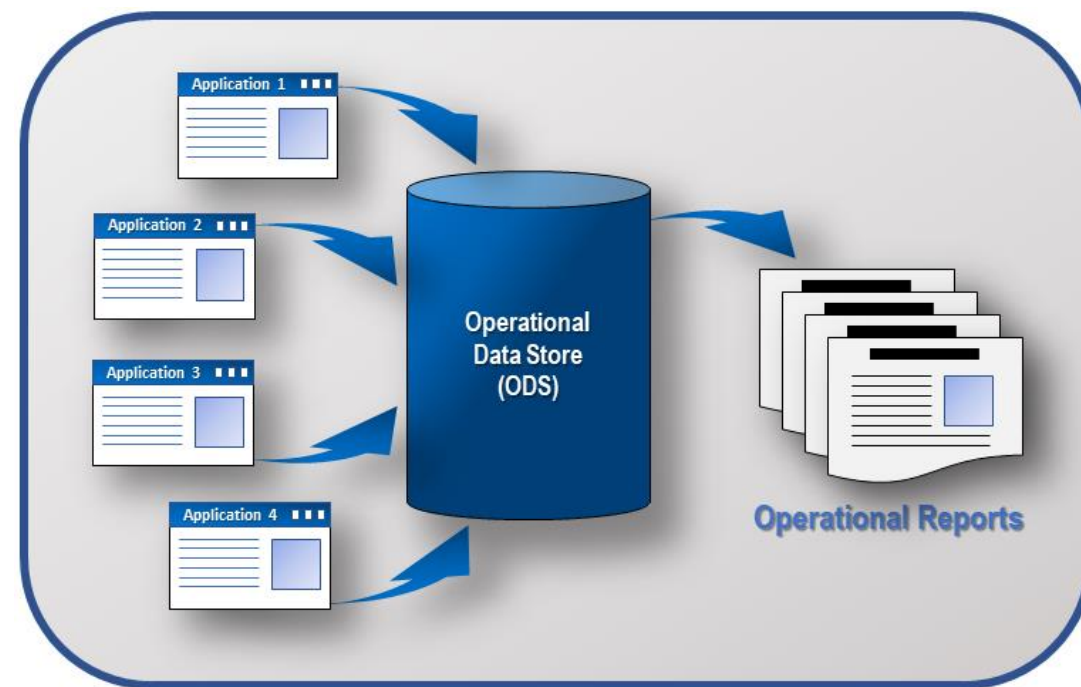
Business Intelligence – O que é ETL

- ✓ ETL significa "Extrair, Transformar e Carregar".



ODS

- ✓ ODS é um repositório de dados onde são colocados os dados para que sejam consultados por outros sistemas, ou por áreas de inteligência.
- ✓ Um ODS reúne dados de várias aplicações **e não é semelhante a um Data Warehouse, pois não tem o compromisso de armazenar um histórico de dados e de servir para processos de auditoria sobre esses dados.**



Fonte: <https://rankeronline.com/how-to-use-an-operational-data-store-for-maximum-efficiency/>

Importância do ETL

O processo ETL é uma das fases mais críticas na construção de um sistema DW, visto que:

- ✓ Grandes volumes de dados são processados.
- ✓ Serão implementadas as regras e fórmulas dos indicadores que irão compor as tabelas de Fato.
- ✓ É essencial para a criação das estruturas de Dimensões e Fatos no ambiente do DW. É ele que faz a “ponte” de ligação entre o operacional e o DW.



Importância do ETL

- ✓ Estudos relatam que o ETL e as ferramentas de limpeza de dados consomem um terço do orçamento num projeto de DW, podendo, no que tange ao tempo de desenvolvimento de um projeto de DW, chegar a consumir 80% desse valor.
- ✓ Outros estudos mencionam, ainda, que o processo de ETL consome 55% do tempo total de execução do projeto de DW.



Importância nas Empresas

- ✓ Quando utilizado em um Data Warehouse corporativo, o ETL fornece o contexto histórico completo para a empresa.
- ✓ Ao fornecer uma visão consolidada, o ETL facilita para os usuários corporativos a análise e a criação de relatórios sobre dados relevantes às suas iniciativas.
- ✓ O ETL pode melhorar a produtividade de profissionais analíticos, porque ele codifica e reutiliza processos que movem os dados sem que esses profissionais possuam a capacidade técnica de escrever códigos ou scripts.



Importância nas Empresas

- ✓ O ETL evoluiu ao longo do tempo para suportar os requisitos emergentes de integração para coisas como streaming data.
- ✓ As ferramentas de ETL migraram para a nuvem. Por exemplo na AWS temos:
 - ✓ Hevo Data
 - ✓ AWS Glue
 - ✓ AWS Data Pipeline
 - ✓ Stitch Data
 - ✓ Talend



AWS Glue

- ✓ O AWS Glue é uma das ferramentas AWS ETL mais populares no mercado atual.
- ✓ É muito fácil de usar, tudo o que você precisa fazer é criar e executar um trabalho ETL com apenas alguns cliques no Console de gerenciamento da AWS.
- ✓ Você só precisa configurar o AWS Glue para apontar para seus dados armazenados na AWS.

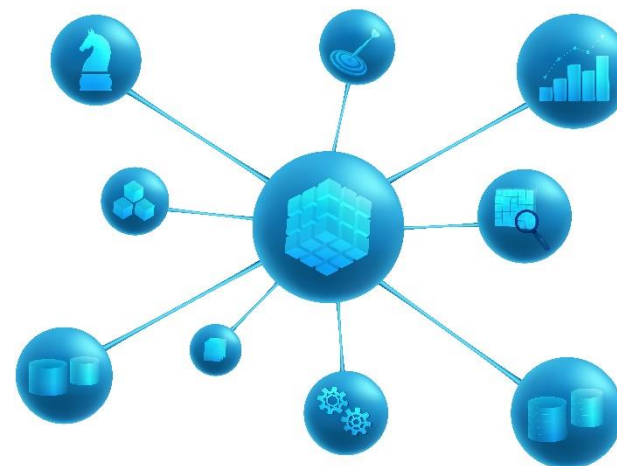


OLAP

OLAP – Online Analytical Processing

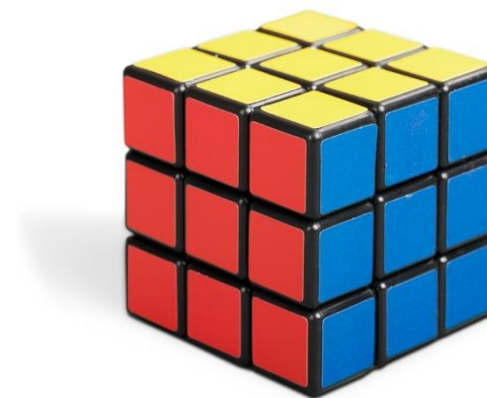
OLAP (*Online Analytical Processing* ou Processo Analítico em Tempo Real) é uma das ferramentas mais usadas para a exploração de um Data Warehouse. O OLAP possibilita alterar e analisar grandes quantidades de dados em várias perspectivas diferentes. As funções básicas do OLAP são:

- ✓ Visualização multidimensional dos dados.
- ✓ Exploração.
- ✓ Rotação.
- ✓ Vários modos de visualização.



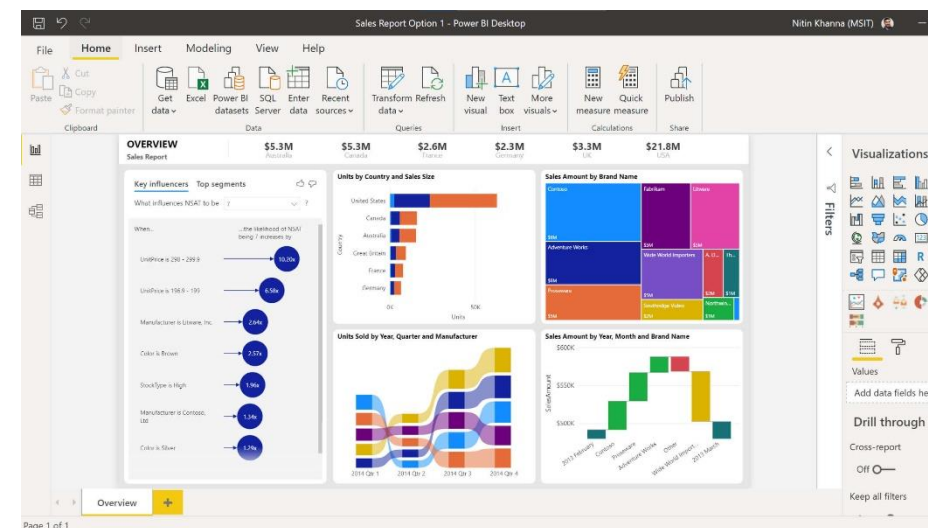
OLAP – Online Analytical Processing

- ✓ OLAP e o Data Warehouse são destinados a trabalharem juntos, enquanto o DW armazena as informações de forma eficiente, o OLAP deve recuperá-las com a mesma eficiência, porém com muita rapidez.
- ✓ Um cubo OLAP é uma estrutura de dados montada de forma multidimensional, e que proporciona uma rápida análise de valores quantitativos ou medidas relacionadas com determinado assunto, sob diversas perspectivas diferentes.



Power BI

Juntamente com o Tableau são os líderes no mercado de Business Intelligence. Começou a ser desenvolvido em 2010 com o codinome “Project Crescent”, porém só teve seu lançamento em 2013 e com outro nome Microsoft Power BI para o Office 365. No começo, era como uma parte do software de gerenciamento de planilhas da empresa, Microsoft Excel, com as funções PowerView, PowerQuery e Power Pivot.



Tableau

- ✓ <https://www.tableau.com>
- ✓ A empresa Tableau Software, foi fundada em 2003, por Chris Stolte, Pat Hanrahan e Christian Chabot que tinham como intenção fornecer as visualizações dos dados do departamento de computação da Universidade de Stanford. Pat Hanrahan foi um dos membros fundadores da empresa de animação digital Pixar, então eles tinham a visão que a computação gráfica poderia ser de ajuda enorme na compreensão dos dados nas organizações (Tableau, 2017).



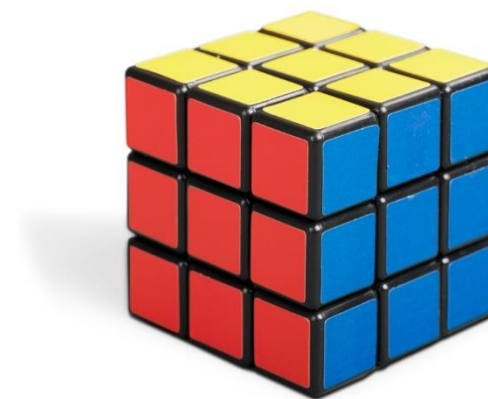
Modelagem Multi-Dimensional

Métricas no Data Warehouse

- ✓ Métrica é também chamada de quantificador ou medida. Alguns chamam de **KPI (Key Performance Indicator)**, mas KPI também pode ser considerado um cálculo entre duas métricas.
- ✓ A ideia por trás do KPI é simples: **definir** métricas para que uma empresa possa alcançar um determinado objetivo, seja ele qual for.
- ✓ Elas são utilizadas para metrificar algo e sempre são números, porque precisam ser contáveis. Esses números são provenientes de transações da empresa.

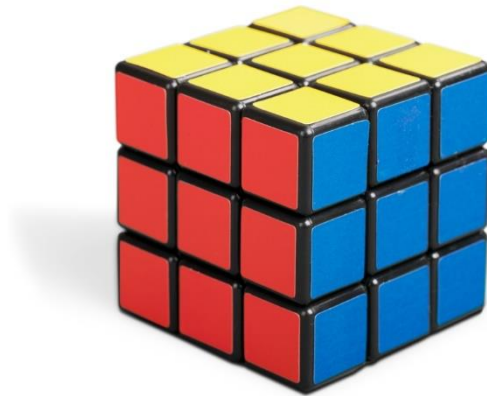
Modelagem Multidimensional

- ✓ Pode ser representado metaforicamente como um cubo.
- ✓ A Modelagem Multidimensional é uma técnica de estrutura de dados otimizada para armazenamento em um DW. O objetivo é otimização, visando a recuperação mais rápida e segura.
- ✓ O conceito de Modelagem Dimensional foi desenvolvido pelo professor Ralph Kimball, o precursor da ideia de Data Warehouse.



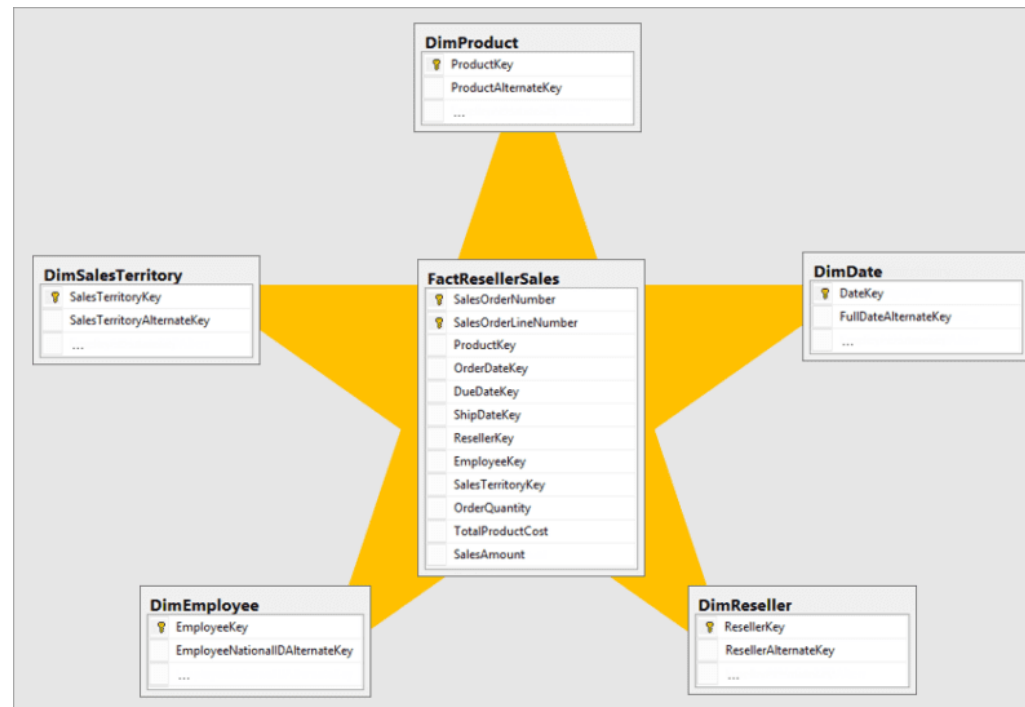
Modelagem Multidimensional

- ✓ Toda modelagem dimensional possui dois elementos imprescindíveis: as **tabelas Fatos** e as **tabelas Dimensões**. Ambas são obrigatórias e possuem característica complementares dentro de um Data Warehouse.



Modelo Star Schema

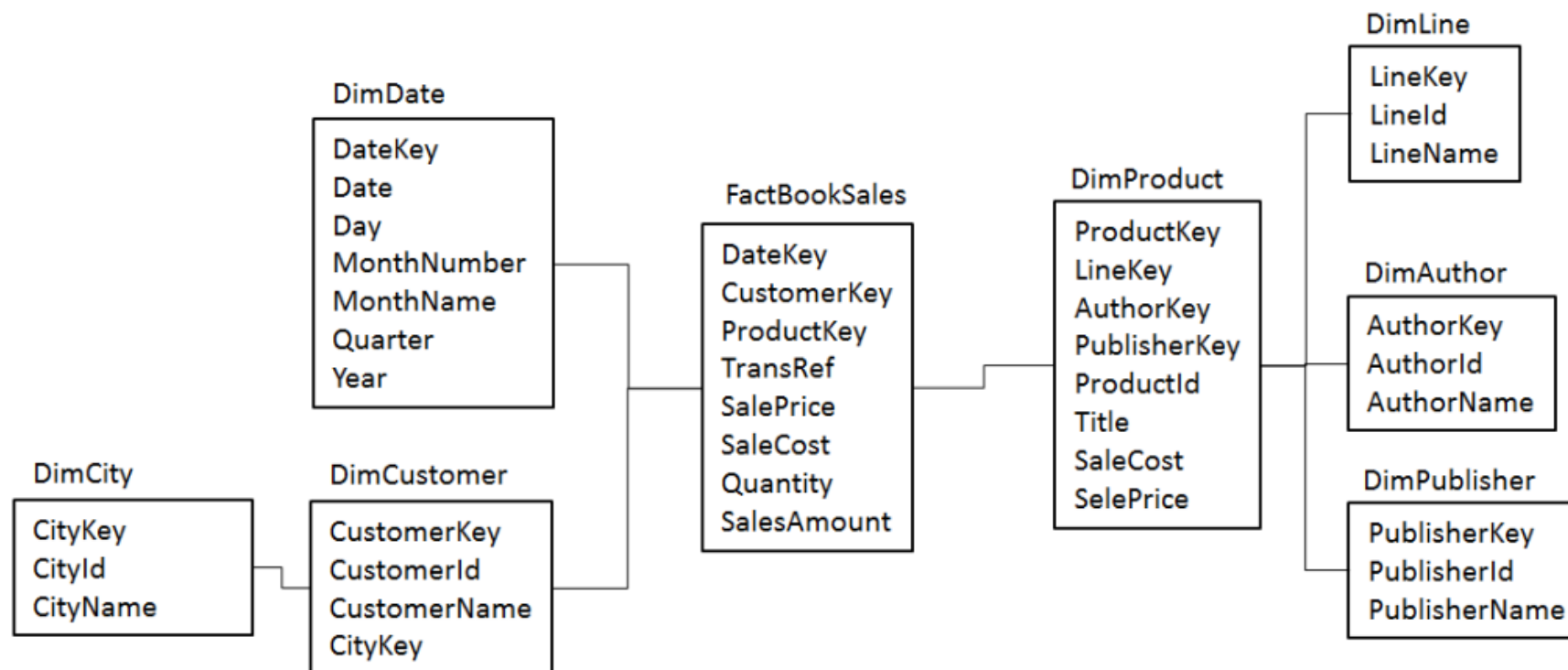
- ✓ Considerando um banco de dados de lojas, produtos e um data warehouse para executar relatórios de vendas agrupados por loja, data, ou categoria ou marca do produto.



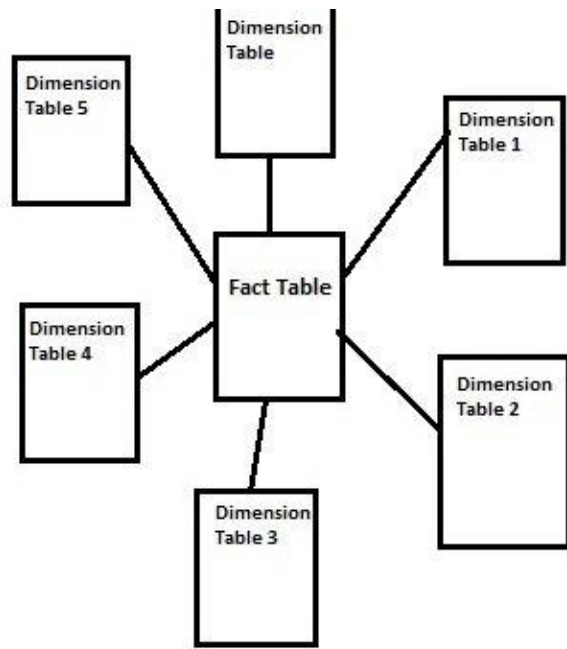
Fonte: <https://blog.math.group/metodologia-star-schema-de-modelagem-dimensional>

Modelo Snowflake

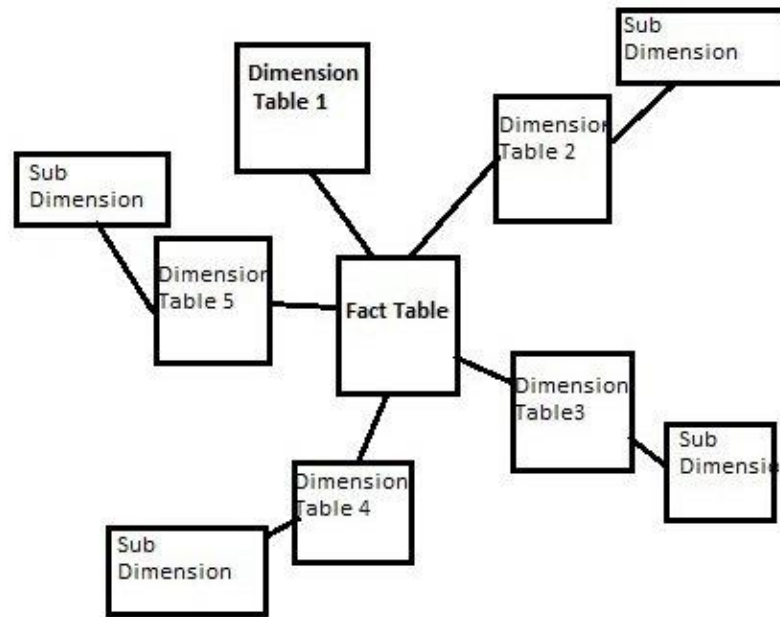
- ✓ O mesmo cenário no esquema Snowflake seria estruturado da seguinte maneira:



Modelos Star Schema x SnowFlake



Star Schema



Snow Flake Schema

Fonte: <https://www.databeetle.net/blog/dimensional-modelling/>

Considerações

- ✓ **Modelos Star Schema (mais usados):**
 - Dimensões Desnormalizadas.
 - Voltados para acessos com performance.
 - Hierarquias achatadas.
 - Mais simples e mais fácil navegação.
 - Utilizam mais espaço, repetindo as mesmas descrições ao longo de toda a tabela.

Considerações

✓ Modelo Snowflake:

- Normalizado.
- Hierarquias mantidas.
- Muitas tabelas → Muitas junções – 1:N..
- Reduz o espaço de armazenamento dos dados dimensionais, mas acrescenta várias tabelas, deixando-o mais complexo.
- **Acesso mais lento do que no StarSchema.**

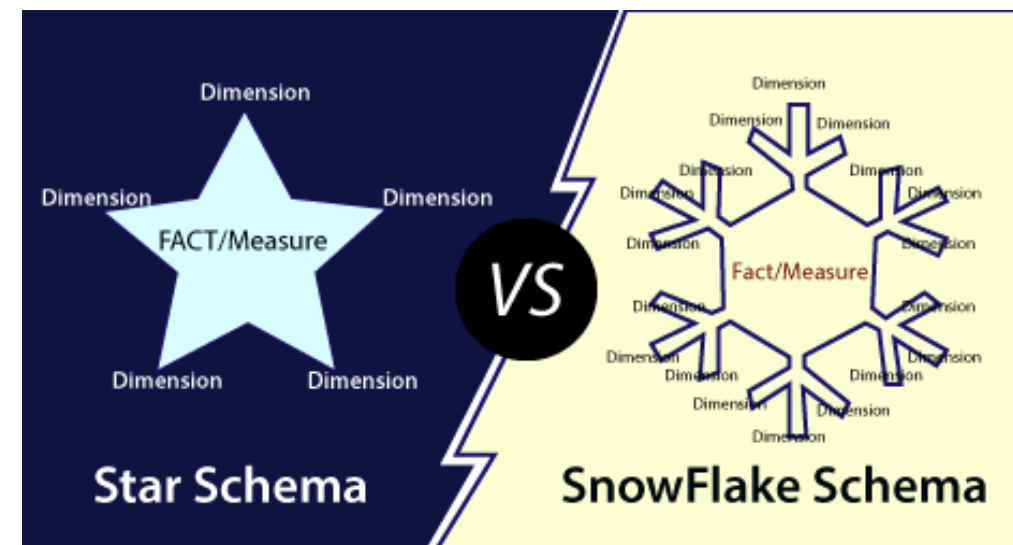
Considerações

✓ Modelos Star Schema *(mais usado)*

- Dimensões Desnormalizadas.
- Voltado para acessos com performance.
- Hierarquias achatadas.
- Mais simples e mais fácil navegação.
- Utiliza mais espaço repetindo as mesmas descrições ao longo de toda a tabela.

✓ Modelo Snowflake

- Normalizado
- Hierarquias mantidas.
- Muitas tabelas → Muitas Junções – 1:N.
- Reduz o espaço de armazenamento dos dados dimensionais, mas acrescenta várias tabelas, deixando-o mais complex.
- **Acesso mais lento que no StarSchema.**



Fonte: <https://www.javatpoint.com/data-warehouse-star-schema-vs-snowflake-schema>

Dimensões em DW

Dimensões

- ✓ As dimensões identificam um indicador de análise sobre um empreendimento, negócio ou ação feita.
- ✓ Através das dimensões é possível identificar **quando** (mês, ano), **onde** (estado, região) e **com quem** (segurado, produto) ocorreu um indicador de análise (prêmio emitido).



Fonte: Nardi (2007)

Tabela do Tipo Dimensão

- ✓ A tabela dimensão tem como finalidade armazenar informações como tempo, geografia, produto, cliente.
- ✓ É comum uma tabela dimensão possuir várias colunas de informação com o objetivo de representar sua **hierarquia**.
- ✓ Sua interação com as **tabelas fato** é feita através da relação 1:N.
- ✓ Possuem uma chave primária para garantir a unicidade de seus registros e está presente na tabela fato, consequentemente como parte de sua chave primária.
- ✓ As dimensões armazenam 3 coisas:
 - A Surrogate Key
 - A Natural Key
 - Os atributos

Surrogate Key em Dimensões

Surrogate Key

- ✓ Em um banco de dados, as chaves são usadas para identificar as linhas de uma tabela e fazer as conexões entre elas. No Data Warehouse, temos a **Surrogate Key** nas dimensões, que é a **chave artificial** utilizada para conectar a tabela na Fato.
- ✓ A **Surrogate Key** nada mais é que a **Primary Key** da dimensão.

Surrogate Key

- ✓ A Surrogate Key é uma chave artificial e auto incremental.
- ✓ A palavra artificial vem do tipo, porque ela não existe em lugar nenhum, não está lá no transacional como a Natural Key (PK que vem do legado), ela é criada no Data Warehouse.
- ✓ E é auto incremental porque toda vez que é chamada, troca de número, então ela começa com 1 e vai indo para 2, 3, 4, e assim por diante.

Dimensão Produto						
SK	Chave Natural	Descrição do Produto	Peso	Unidade	Data Inicio	Data Fim
	ou Chave de Negócios					
14	1000	Chocolate X	180	Gramas	01/01/2017	
13	1000	Chocolate X	200	Gramas	01/01/2015	31/12/2016
12	1000	Chocolate X	210	Gramas	01/01/2010	31/10/2014
15	1002	Chocolate Y	170	Gramas	01/01/2017	
9	1002	Chocolate Y	200	Gramas	01/01/2010	31/12/2016

Surrogate Key

- ✓ Ela é gerada automaticamente na hora da carga, quando você carrega a dimensão no ETL.
- ✓ Na tabela Fato, essa **Surrogate Key** vai ser uma **Foreign Key**, a chave que serve para relacionar os dados entre duas tabelas, sempre apontando para uma Primary Key em outra tabela, que no caso da dimensão, vai ser a Surrogate Key.
- ✓ Assim, a tabela Fato receberá apenas o código da Surrogate Key da linha que ela está referenciando e não os atributos.

Surrogate Key

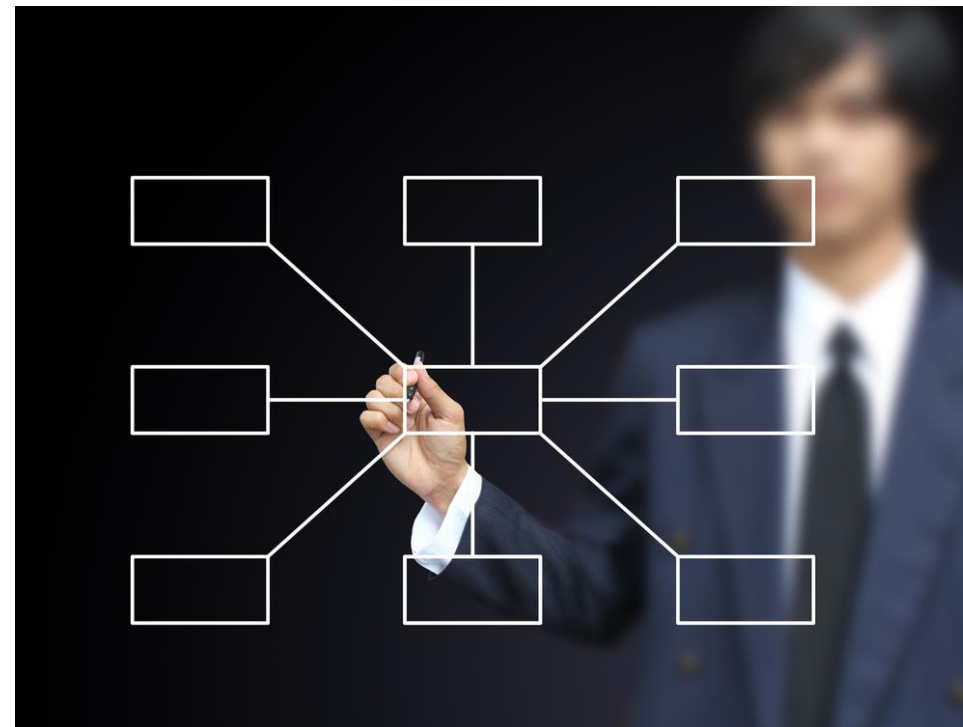
Resumindo, a Surrogate Key:

- ✓ Tem as características de uma Primary Key.
- ✓ É utilizada para referenciar a dimensão na Fato.
- ✓ É auto incremental.
- ✓ É uma chave artificial.
- ✓ É criada no Data Warehouse.
- ✓ Não pode se repetir.

Tabelas Fato

Tabelas Fato

- ✓ Principal tabela do Data Warehouse, ela vai se conectar nas dimensões.
- ✓ Podem existir uma ou mais tabelas fato.
- ✓ Armazenam principalmente:
 - **Métricas** - que são os fatos propriamente ditos (tudo que a empresa for mensurar é uma métrica).
 - **Foreign key** – chave estrangeira que serve para relacionar os dados das Dimensões com a Fato.



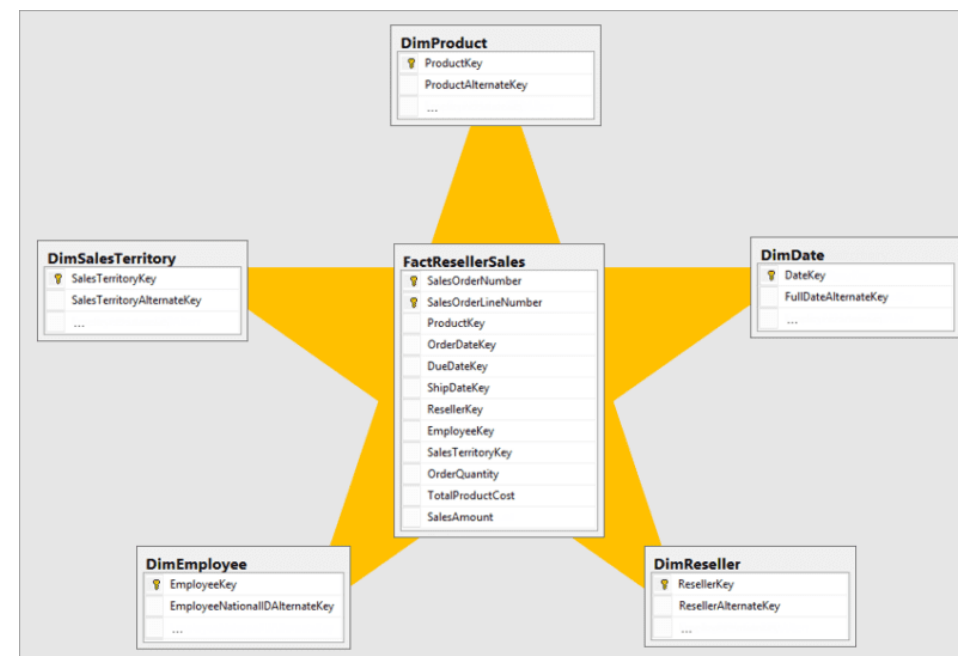
Tabelas Fato

O que é Métrica?

- ✓ É utilizada para medir, quantificar algo, são sempre números provenientes de transações da empresa.
- ✓ Tudo que a empresa quer mensurar é métrica, geralmente sendo o que o usuário quer medir. Por exemplo: número de vendas ou seguidores em determinada rede social.

O que é uma Foreign Key?

- ✓ É uma chave estrangeira que serve para relacionar os dados entre as tabelas fato e dimensão.



Fonte: <https://blog.math.group/metodologia-star-schema-de-modelagem-dimensional>

Tabelas Fato

Tabela fato é a tabela com uma quantidade enorme de linhas, e que representa os fatos de uma empresa, que podem se repetir ou não.

A tabela dimensão é a tabela que vai auxiliar a tabela fato com dados complementares ou explicativos, e que possui informações que não se repetem.

Engines de Data Warehousing

O que é uma Engine para Data Warehousing?

- Esses mecanismos são voltados para cargas de trabalho analíticas que exigem consultas maiores, mas menos frequentes do que suas contrapartes transacionais. Bancos de dados transacionais como o PostgreSQL são otimizados para fazer leituras e gravações rápidas em volumes incrivelmente altos para executar os aplicativos que eles atendem.
- Os casos de uso analíticos consultam dados com menos frequência, mas suas consultas geralmente são mais complexas e em conjuntos maiores de dados.



O que é uma Engine para Data Warehousing?

- Em 2005, foi lançado um documento inovador conhecido como C-Store paper, introduzindo uma nova arquitetura de armazenamento em colunas. Os muitos desenvolvimentos nesse documento levaram a uma nova classe de bancos de dados baseados em nuvem que podem lidar com grandes conjuntos de dados de maneira muito poderosa.



- C-Store é um sistema de gerenciamento de banco de dados (DBMS) baseado em um DBMS orientado a colunas desenvolvido por uma equipe da Brown University, Brandeis University, Massachusetts Institute of Technology e University of Massachusetts Boston, incluindo Michael Stonebraker, Stanley Zdonik e Samuel Madden.
- O C-Store difere da maioria dos projetos tradicionais de sistemas de gerenciamento de banco de dados relacional (RDBMS) de várias maneiras, principalmente porque armazena dados por coluna e não por linha, otimizando o banco de dados para leitura de dados em vez de gravação.

Modern Warehouse Engine

Hoje, existem três opções dominantes para mecanismos de armazenamento de dados baseados em nuvem:

- Amazon Redshift.
- Google BigQuery.
- Snowflake.

Todos estes são semelhantes e baseados no C-Store paper.

Amazon Redshift

- O Redshift tem os benefícios de facilidade de uso, velocidade e custo.
- Como parte da AWS, há integração completa de serviços para uma ampla variedade de serviços da AWS, como S3 para armazenamento e CloudWatch para monitoramento de infraestrutura.
- O Redshift geralmente é mais barato que o Snowflake ou o BigQuery
- É simples configurar e dimensionar adicionando nós ao cluster e aumentando o armazenamento e o desempenho.
- Redshift é provavelmente o mais popular, embora esteja perdendo terreno para o Snowflake.
- Ele se beneficia de ser semelhante em conexão e sintaxe SQL ao PostgreSQL.



Google BigQuery

O BigQuery não é limitado pela capacidade do cluster de armazenamento ou recursos de computação, por isso é dimensionado e funciona muito bem com demandas crescentes de simultaneidade (por exemplo, mais usuários e consultas acessando o banco de dados).

Para empresas que usam produtos do Google, o BigQuery se integra bem ao Google Drive e ao Google Analytics.



Assim como o BigQuery, o Snowflake tem uma arquitetura que separa o mecanismo de consulta de computação do armazenamento de dados. Como resultado, é altamente escalável em qualquer quantidade de volume e simultaneidade.

O preço é baseado no armazenamento e na computação usados com base no tempo com seus bancos de dados virtuais, em vez de por bytes verificados.

Ajuste, índices e chaves de distribuição não são necessários para que as consultas sejam otimizadas e tenham desempenho.

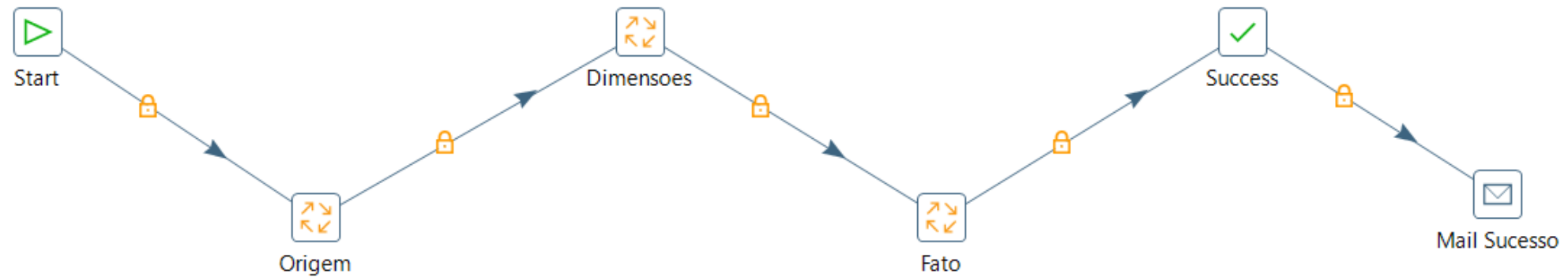
Por esses motivos, pode-se dizer que o Snowflake tem muitos dos benefícios do Redshift e do Big Query.

Orquestração de Transformações em DW

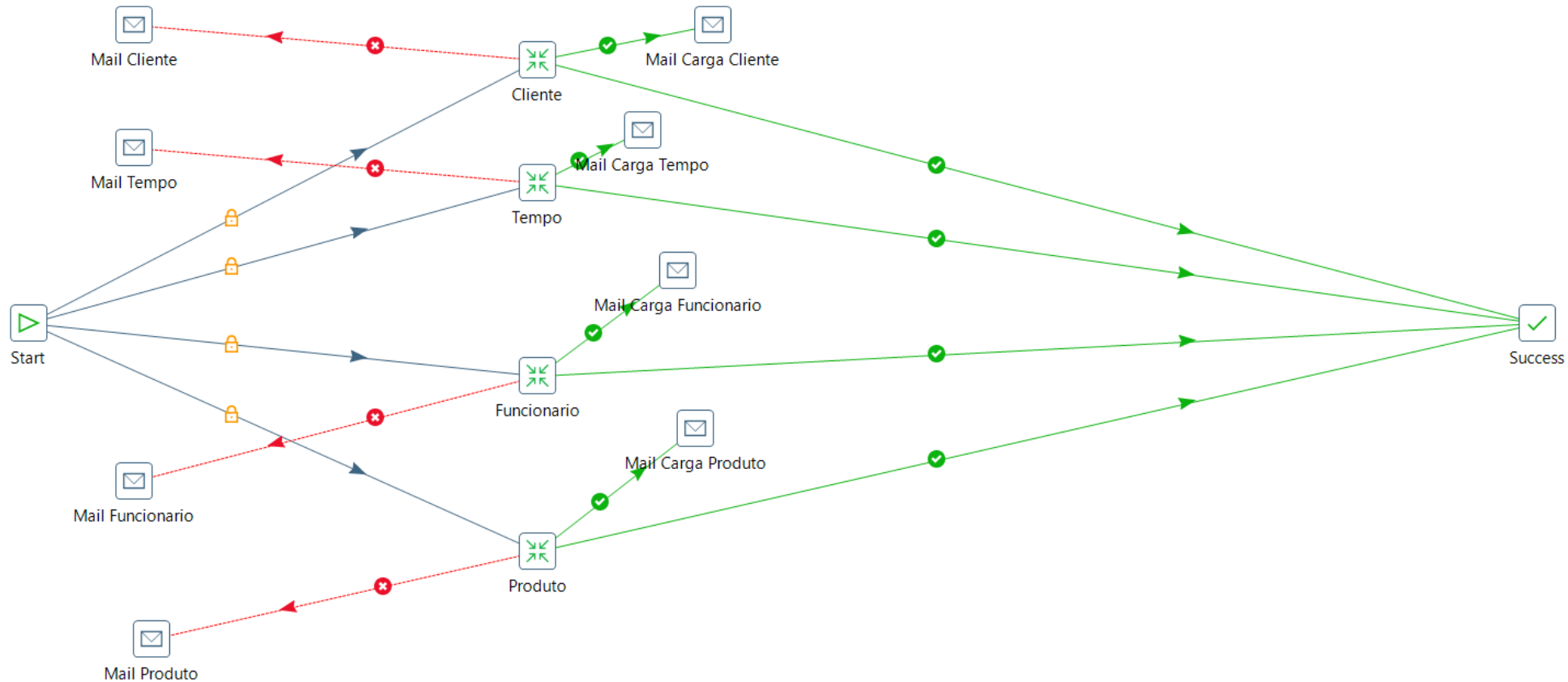
Orquestração de Transformações dos Dados

- A orquestração de Transformações de Dados é o processo de obter dados em silos de vários locais de armazenamento de dados, combiná-los e organizá-los e disponibilizá-los para ferramentas de análise de dados.
- A orquestração de dados permite que as empresas automatizem e simplifiquem a tomada de decisões orientada por dados.

Exemplo de Orquestração pelo Pentaho



Exemplo de Orquestração pelo Pentaho



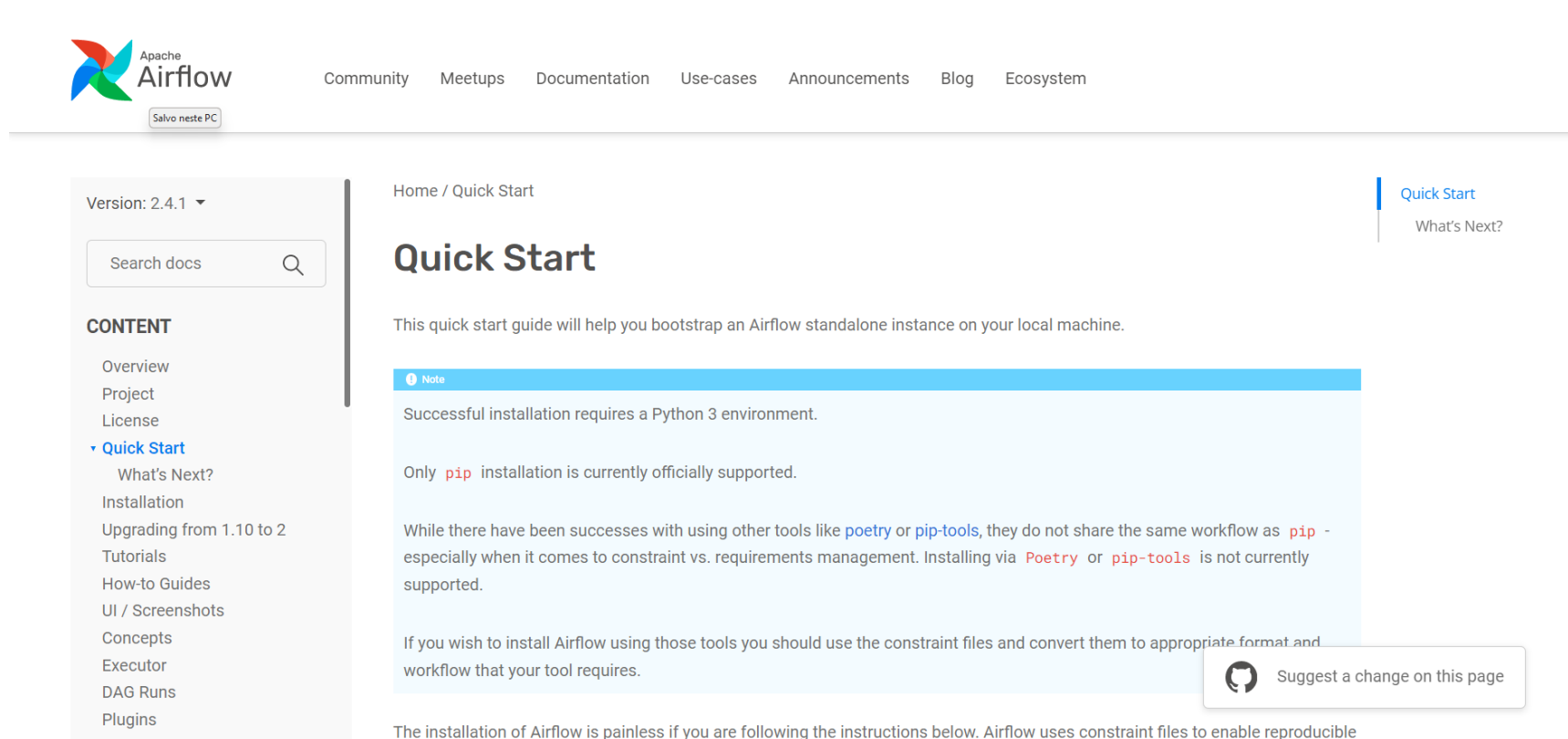
Apache Airflow

- <https://airflow.apache.org/>
- O Airflow é uma plataforma criada pela comunidade para criar, agendar e monitorar de forma agendada os fluxos de trabalho.
- O Airflow tem uma arquitetura modular e usa uma fila de mensagens para orquestrar um número arbitrário de trabalhadores. O fluxo de ar está pronto para escalar até o infinito.
- Os pipelines de fluxo de dados são definidos em Python, permitindo a geração dinâmica de pipeline. Isso permite escrever código que instancia pipelines dinamicamente.



Apache Airflow - Documentação

- <https://airflow.apache.org/docs/apache-airflow/stable/start.html>



The screenshot displays the Apache Airflow documentation website. At the top, the Apache Airflow logo is on the left, and navigation links for Community, Meetups, Documentation, Use-cases, Announcements, Blog, and Ecosystem are on the right. Below the logo is a 'Salvo neste PC' button. The main content area is titled 'Quick Start' and includes a sub-header 'Home / Quick Start'. A sidebar on the left lists the table of contents, with 'Quick Start' highlighted. The main text explains that the quick start guide helps bootstrap an Airflow standalone instance on a local machine. A blue note box contains important installation details: it states that successful installation requires a Python 3 environment, only `pip` installation is officially supported, and while other tools like `Poetry` or `pip-tools` can be used, they are not currently supported. A 'Suggest a change on this page' button is visible at the bottom right of the note box. The page also includes a version selector (2.4.1) and a search bar.

Version: 2.4.1 ▾

Search docs 🔍

CONTENT

- Overview
- Project
- License
- **Quick Start**
 - What's Next?
 - Installation
 - Upgrading from 1.10 to 2
 - Tutorials
 - How-to Guides
 - UI / Screenshots
 - Concepts
 - Executor
 - DAG Runs
 - Plugins
 - Security

Home / Quick Start

Quick Start

This quick start guide will help you bootstrap an Airflow standalone instance on your local machine.


Note

Successful installation requires a Python 3 environment.

Only `pip` installation is currently officially supported.

While there have been successes with using other tools like `Poetry` or `pip-tools`, they do not share the same workflow as `pip` - especially when it comes to constraint vs. requirements management. Installing via `Poetry` or `pip-tools` is not currently supported.

If you wish to install Airflow using those tools you should use the constraint files and convert them to appropriate format and workflow that your tool requires.

 Suggest a change on this page

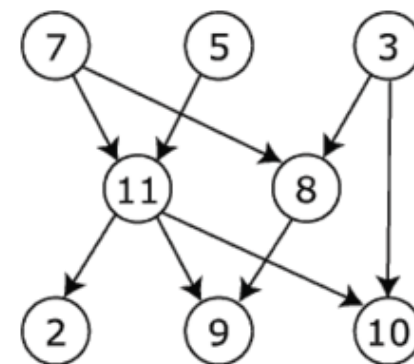
The installation of Airflow is painless if you are following the instructions below. Airflow uses constraint files to enable reproducible

Apache Airflow

- DAG - Grafos de tarefas/ usos.
- OPERATOR - O operador refere-se à etapa de transformação ainda dividida em:
 - Sensor - Este tipo de operador realiza uma função de polling com frequência/timeout.
 - Executor - Este tipo de operador realiza operações de trigger, por exemplo, HiveOperator, Pig Operator.
 - TASK - Task é a principal entidade do DAG. O principal aqui é a instância de tarefa considerada para executar uma tarefa em um ponto do tempo.
 - HOOK - É considerado como a Interface para o Sistema externo como um gancho de JDBC e HTTP.

Apache Airflow - DAGs

- Em ciência da computação e matemática, um grafo acíclico direcionado (DAG – directed acyclic graphs) é um grafo direcionado que não possui ciclos direcionados.
- Na teoria dos grafos, um grafo refere-se a um conjunto de vértices conectados por linhas chamadas arestas.
- Em um grafo direcionado, cada aresta está associada a uma direção de um vértice inicial a um vértice final.
- Se viajarmos ao longo da direção das arestas e descobrirmos que não se formam laços fechados ao longo de qualquer caminho, diz-se que não há ciclos direcionados.
- E o gráfico formado é um DAG.





PUC Minas
Virtual