



## A aula interativa do Módulo 3 – Bootcamp Desenvolvedor Business Intelligence em breve!

### Importante:

- 1) Não se esqueça de acessar a aula com seu e-mail cadastrado no ambiente de aprendizagem do IGTI e seu nome completo. Entrou com os dados errados? Saia da sala e entre novamente com os dados corretos.
- 2) Para sua frequência ser computada, responda a enquete no ambiente de aprendizagem, no horário indicado pelo professor. A enquete ficará disponível por 10 minutos.
- 3) Utilize o Chat para interagir com os colegas durante a aula, a ferramenta Raise Hands para pedir a palavra, e, em caso de dúvidas sobre o conteúdo, utilize o Q&A (perguntas e respostas) para que o professor tutor possa respondê-las.

# Informações importantes

- Acesse a aula com o seu nome completo e com o mesmo e-mail utilizado no cadastro do ambiente de aprendizagem do IGTi. Entrou com os dados errados? É só sair da sala e entrar com os dados corretos!
- As presenças das Aulas Interativas são computadas através de uma enquete, que será realizada no ambiente de aprendizagem do IGTi. Para sua frequência ser computada, quando solicitado pelo professor, você deverá ir até a seção “Enquete de presença da Aula Interativa”, localizada na Área Acadêmica da disciplina, e responder a enquete. Essa seção estará logo abaixo da que contiver o link para a Aula Interativa em questão. A enquete ficará no ar por 10 minutos e sua nota estará disponível ao término desse tempo.
- Utilize o chat para interagir com os colegas durante a aula interativa! Em caso de dúvidas sobre o conteúdo, é só postá-las no Q&A que o tutor irá respondê-las! Lembre-se que você ainda pode esclarecer as suas dúvidas nos fóruns disponibilizados no Ambiente de Aprendizagem.

# Informações importantes

- Ah! E se você não conseguir assistir a aula interativa, não se preocupe! Sua gravação ficará disponível no Área Acadêmica, juntamente com os slides utilizados pelo professor, em até 24 horas úteis após o término da aula. Você também poderá realizar a atividade de reposição para recuperar os pontos de presença!
- Se você precisar solicitar prorrogação e/ou 2ª oportunidade para entrega de atividades, saiba que isso é realizado somente mediante a apresentação de atestado médico ou de óbito de parentes de 1º grau.
- Para melhor experiência nas aulas interativas, sugerimos que você baixe o aplicativo do Zoom no seu computador.

# Aplicações em ETL

Primeira Aula Interativa

Prof. Ricardo Brito Alves

# Nesta aula



- ☐ Apresentação do Professor.
- ☐ Tópicos da Disciplina e Temas Interessantes.
- ☐ Dúvidas.
- ☐ 2ª Aula Interativa.

# Apresentação do Professor

---



# Apresentação do Professor



Ricardo Brito Alves

## Formação Acadêmica:

- Graduado em Ciência da Computação pela Pontifícia Universidade Católica de Minas Gerais ,1994.
- Especialista em Gestão de Negócios pela Una, 2008.
- MBA em Gestão Estratégica de Projetos pela Una, 2009.
- Mestrado em Engenharia Elétrica pela Pontifícia Universidade Católica de Minas Gerais, 2018.
- Doutorando em Ciência da Computação pela Universidade Federal de Minas Gerais, 2024.

# Apresentação do Professor



Ricardo Brito Alves

## Experiência Profissional:

- Desde 1993 venho atuando no setor de tecnologia, com desenvolvimento de projetos de Software.
- Desde 2002 atua com projetos de Data mining e BI.
- Atua há 7 anos na área de Inteligência Artificial.
- Ocupa atualmente o cargo de IT Manager em uma empresa de tecnologia e é docente de cursos de pós-graduação em tecnologia.



# **Tópicos da Disciplina e Temas Interessantes**

---


# Material




<https://drive.google.com/drive/folders/1AJutZfoazPGdPdGQsE8Sf6jIPlqsdw-f>

Name ↑

 Conector MySql

 Java

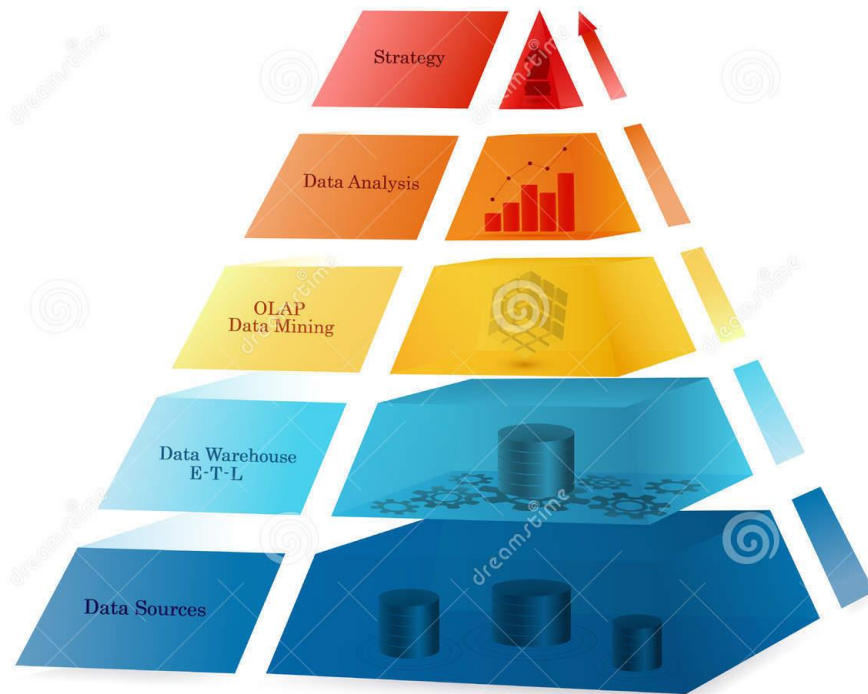
 MySql

 PDI - Pentaho Data Integrator

 PME - Pentaho Metadata Editor

 XAMP

# Estratégia

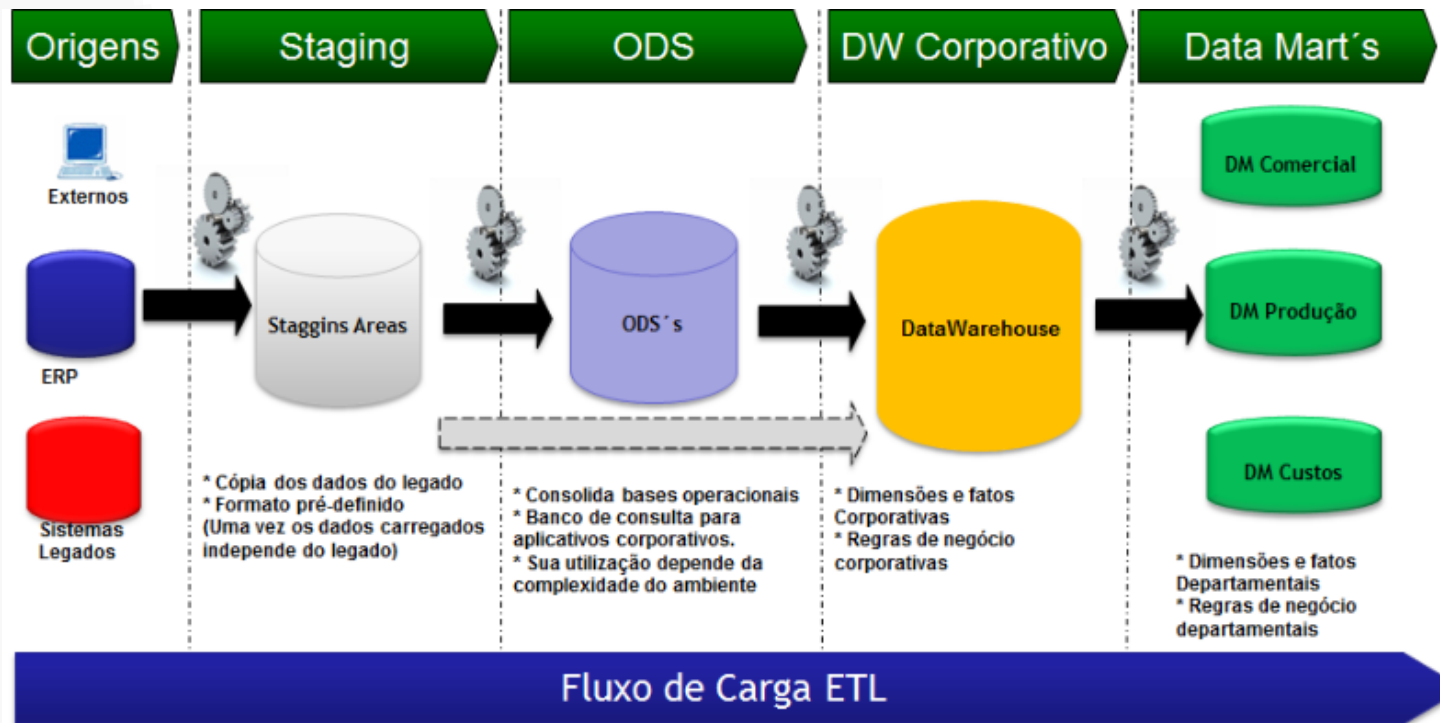


Download from  
**Dreamstime.com**  
This watermarked comp image is for previewing purposes only.

91453009

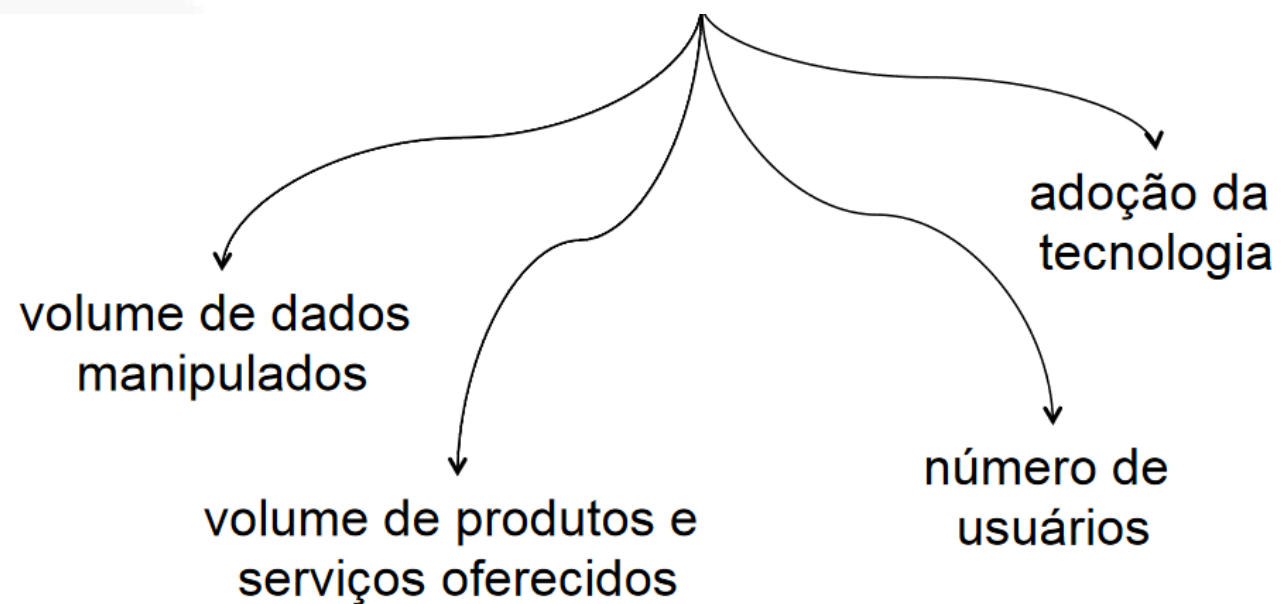
Eugenia Stan | Dreamstime.com

# Business Intelligence



# Conceitos Básicos

## Arquitetura



# Tuning em Banco de Dados

Em TI, Tuning refere-se basicamente ao conceito de propor e aplicar mudanças visando otimizar o desempenho na recuperação ou atualização de dados.

Em curtas palavras, Tuning é sinônimo de otimização.

Para fazer um bom trabalho de Tuning, é necessário executar criteriosamente os seguintes processos:

- Entender o problema;
- Elaborar o diagnóstico;
- Aplicar as dicas e técnicas de otimização (que se aplicam ao diagnóstico elaborado).

# Tuning em Banco de Dados

## 1. Planejamento de performance:

Definição e configuração do ambiente em que o BD será instalado, considerando-se os seguintes itens: Hardware, Software, Sistema Operacional e Infraestrutura de rede.

## 2. Tuning de instância e BD:

Ajuste de parâmetros e configurações do BD (atividades que fazem parte do trabalho de um DBA).

## 3. SQL Tuning:

Otimização de instruções SQL.



# SQL Tuning: otimização de instruções SQL



- Identificar uma query lenta ou “pesada” no seu banco de dados.
- Identificar as consultas que utilizam um determinado índice através de alguma ferramenta - rebuild de índices.
- Uso de comandos como order by, colunas calculadas, delete x truncate, etc.

# Etapas na Construção de um DW

**Levantamento das necessidades:** devemos antes de tudo fazer o levantamento de todas as informações desejadas pelo usuário. Nesse primeiro momento fazemos o cruzamento de Dimensões e Fatos necessários para alcançar os anseios dos gestores. Trabalhamos **em O QUÊ** o DW terá, **e não O COMO**, por isso não devemos nos preocupar com a existência efetiva dos dados e sim com os desejos requisitados.

**Mapeamento dos dados:** nessa etapa fazemos o mapeamento dos dados, identificando a fonte e como chegar até eles. Aqui vemos a viabilidade dos desejos elencados na primeira etapa, verificando a existência ou não dos dados para o alcance das necessidades solicitadas.

**Construção da Staging Area:** após o mapeamento, construímos a estrutura chamada Staging Area, que se trata da área de transição dos dados do DW. Nessa área os dados são copiados e desacoplados dos sistemas de operação (OLTP) e recebem o devido tratamento para as futuras cargas nas tabelas de Fatos e Dimensões.

# Etapas na Construção de um DW

**Construção das Dimensões:** construímos nessa etapa a estrutura das Dimensões que farão parte do DW. Definimos também a historicidade que os dados irão possuir nas Dimensões.

**Construção das tabelas Fato:** construímos nessa etapa (após a construção das Dimensões) a estrutura da tabela Fato. Aqui é avaliado e definido a granularidade da informação que será armazenada em cada Fato. Avaliamos também a expectativa de crescimento e de armazenamento que serão utilizados.

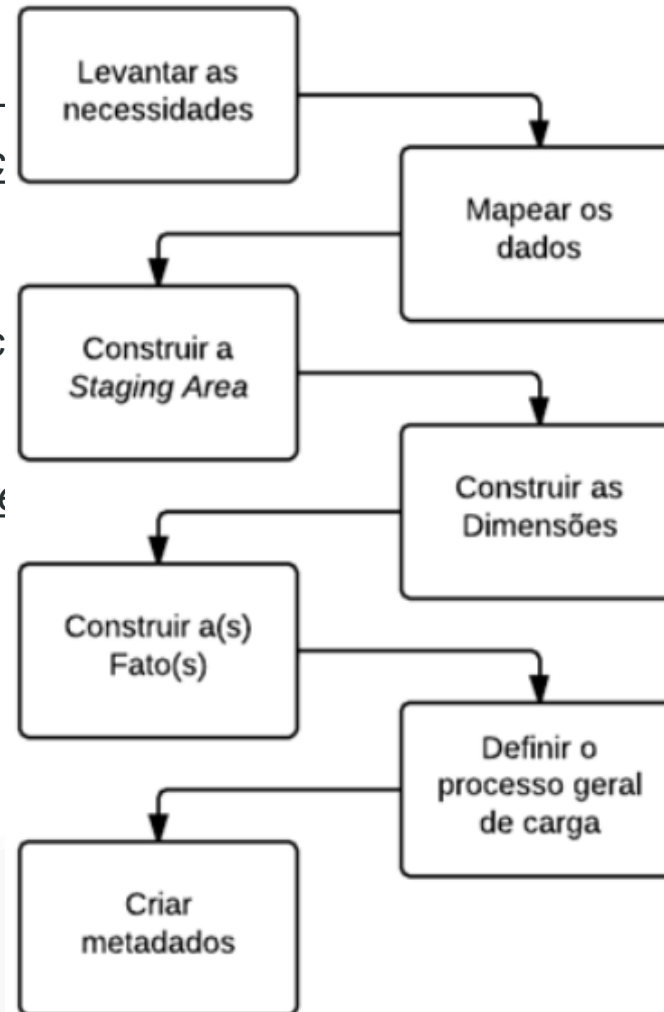
**Definição do processo geral de carga:** após concluirmos as etapas anteriores, precisamos criar o motor para que tudo seja carregado, atualizado, orquestrado e processado de forma automática e ordenada. Por isso, a necessidade do processo geral de carga que é o “cérebro” do DW.

**Criação dos metadados:** por fim, precisamos desenvolver toda a documentação dos metadados, que incluem o processo de construção e o dicionário de dados. Os metadados fornecem apoio importante para a gestão do conhecimento.

# Aprendizado por Reforço

Lembrando que o Data Mart é em subconjuntos de informação por assuntos específicos.

Logo, todas essas etapas, com o “levantamento das necessidades”, provavelmente deverão ser recriadas no novo Data Mart criado.



# Características dos Dados

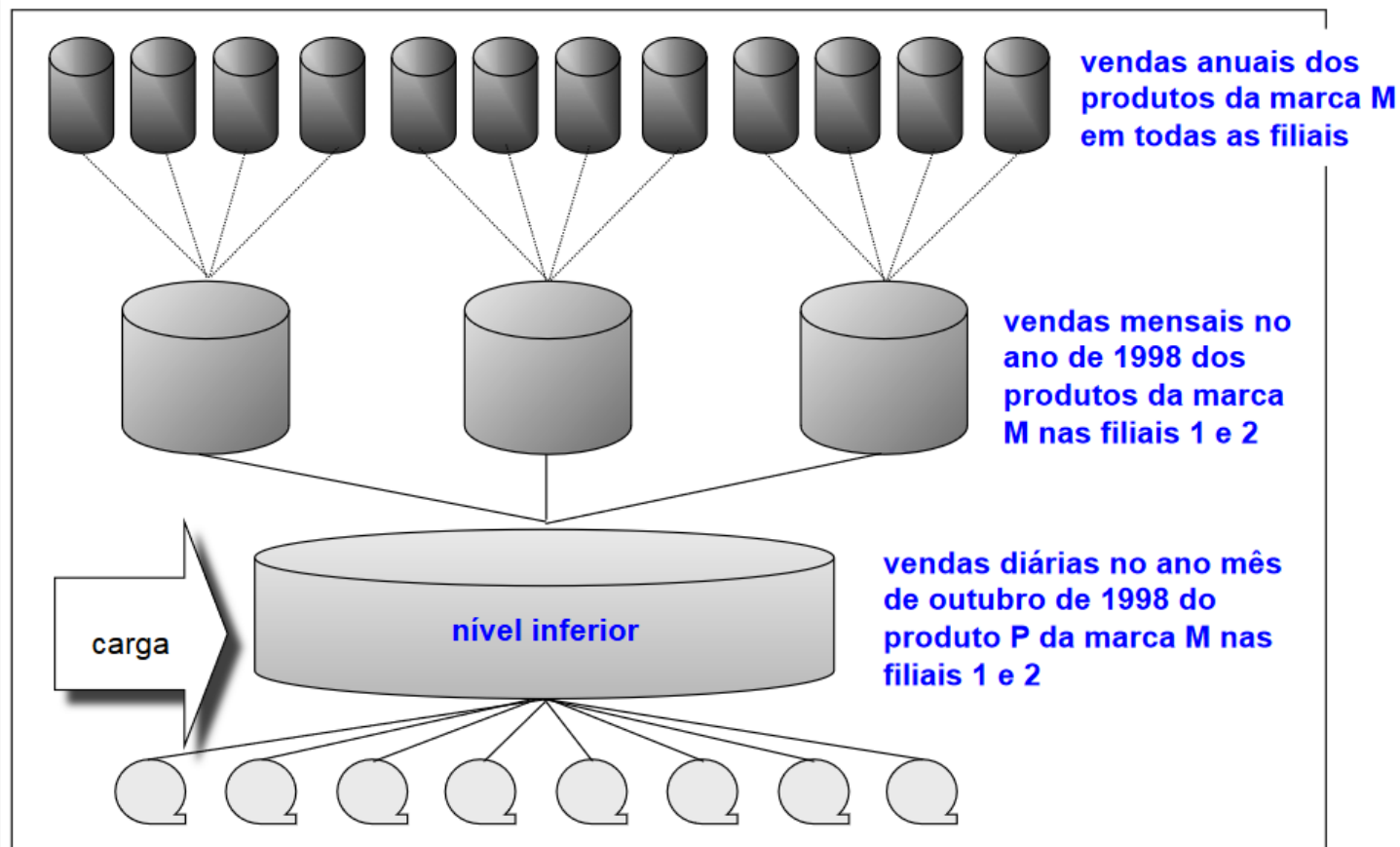
## São não-voláteis

- O conteúdo do DW permanece estável por longos períodos de tempo.

## São históricos

- Relevantes a algum período de tempo, por exemplo: usualmente dados relativos a um grande espectro de tempo (5 a 10 anos) encontram-se disponíveis.

# Características dos Dados



# Granularidade

Grau de detalhamento em que os dados são armazenados em um nível.

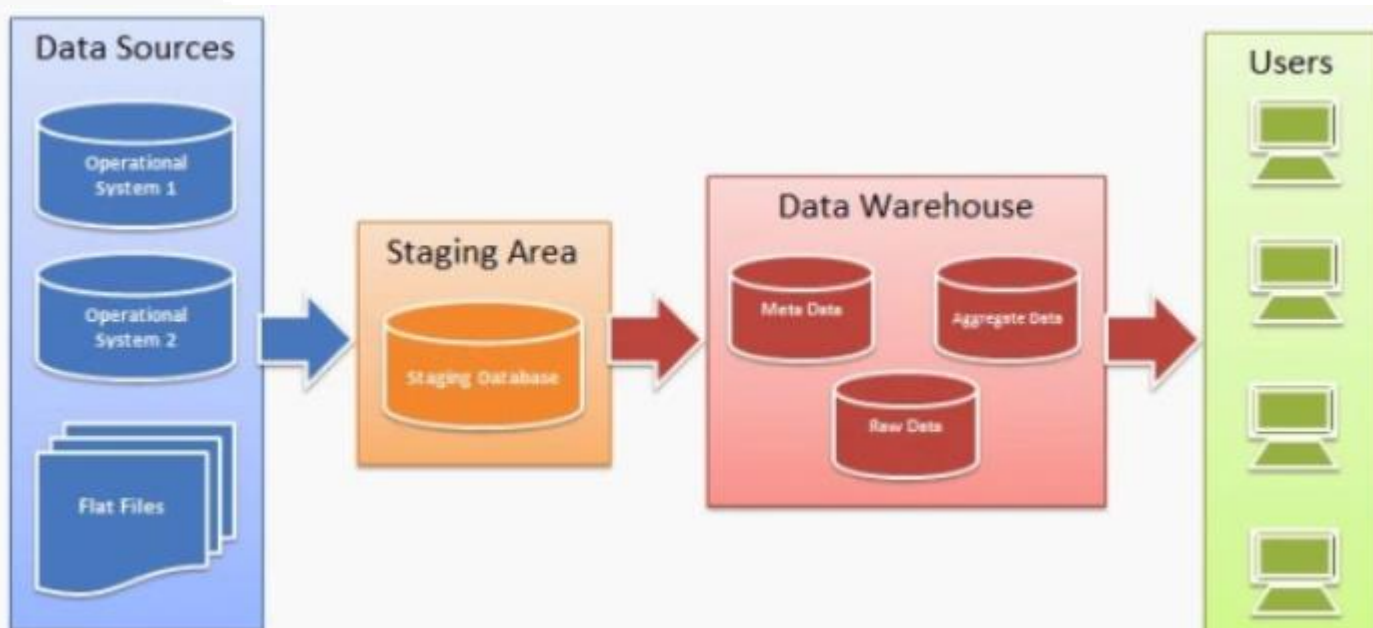
Questão de projeto muito importante, pois:

- Impacta no volume de dados armazenado.
- Afeta as consultas que podem ser respondidas.



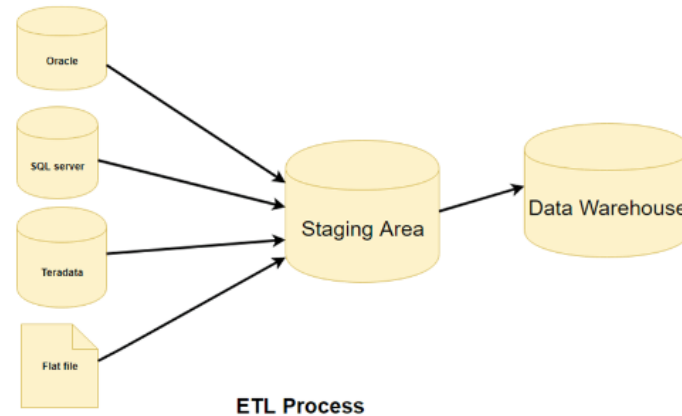
# Stage Area

É uma área de tratamento, padronização e transformação das informações operacionais para carga na arquitetura de dados BI (DW, ODS, DM).



# Stage Area

ST1 (staging 1)



## Características:

1. Query de carga idêntica à tabela de origem.
2. Fonte de Origem é a tabela origem.
3. Levam-se todos os campos da tabela origem.
4. Não existe chave primária na ST1.
5. Método de carga escolhido Truncate.
6. A cada processamento a tabela será esvaziada e carregada novamente.

# Stage Area



## ST2 (staging 2)

### Características:

1. Query de carga com transformações e cálculos.
2. Fonte de origem é a ST1;
3. Levam-se os campos que serão utilizados nas Dimensões e Fatos;
4. Existe chave primária na ST2. A chave montada é a chave de negócio;
5. Método de carga escolhido Update/Insert;
6. A latência dessa tabela será de todo o período de carga do DW.
7. A cada processamento a tabela será atualizada com alterações de registros já existentes e com novos registros.

# Stage Area

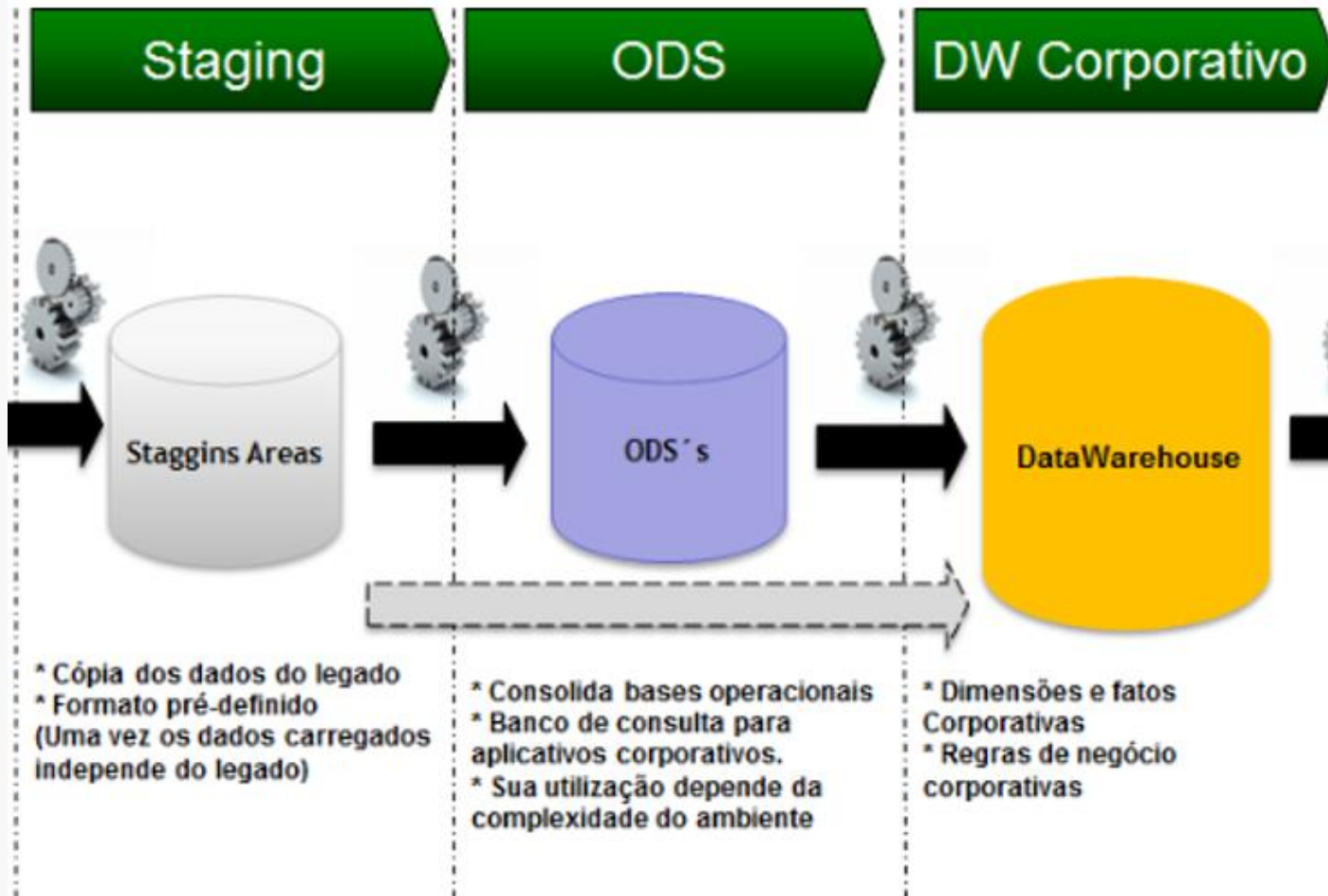


**ST2 Aux (staging 2 aux):** tem como objetivo otimizar o processo de carga diário.

## **Características:**

1. Query de carga com transformações e cálculos.
2. Fonte de origem é a ST1.
3. Levam-se os campos que serão utilizados nas Dimensões e Fatos.
4. Existe chave primária na ST2. A chave montada é a chave de negócio.
5. Método de carga escolhido Truncate.
6. Latência diária.
7. A cada processamento a tabela será esvaziada e carregada novamente.

# ODS



# Surrogate Key



A Surrogate Key nada mais é que o campo de Primary Key da dimensão.

O que é uma Primary Key?

É a coluna utilizada para identificar cada linha na tabela de forma única.

**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, 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.

- Tem as características de uma Primary Key.
- É utilizada para referenciar a dimensão na fato.
- É auto incremental.
- É uma chave artificial.
- Não se repete.

# Carga das Dimensões



As cargas das dimensões serão originadas a partir da ST2.

## Características:

1. Query apenas de leitura da ST2, pois as transformações já foram feitas.
2. Fonte de Origem é a ST2 Aux para carga diária e a ST2 para carga full.
3. Altera-se algum nome de campo para se adequar as regras da corporação de acordo com o Dicionário de Dados.
4. Para cada chave de negócio será gerada uma SRK.
5. A SRK é um campo numérico sequencial.



# Slowly Changing Dimension

## Tipo 1

O valor anterior é sobreposto pelo valor atual, perdendo-se o histórico. Usado principalmente para correção de informações, como nome de segurado e descrição de produto. Exemplo de uma tabela fornecedor.

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	Abc	Acme Supply Co	CA

# Slowly Changing Dimension

Tipo 2

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	START_DATE	END_DATE
123	Abc	Acme Supply Co	CA	01-Jan-2000	21-Dez-2004
124	Abc	Acme Supply Co	IL	22-Dez-2004	

# Slowly Changing Dimension



## Tipo 3

Supplier_Key	Supplier_Code	Supplier_Name	Original_Supplier_State	Effective_Date	Current_Supplier_State
123	Abc	Acme Supply Co	CA	22-Dez-2004	IL

# OLAP



O OLAP é uma interface com o usuário e não uma forma de armazenamento de dados, porém se utiliza do armazenamento para poder apresentar as informações.

Os métodos de armazenamento são:

- ROLAP (OLAP Relacional): Os dados são armazenados de forma relacional.
- MOLAP (OLAP Multidimensional): Os dados são armazenados de forma multidimensional.
- HOLAP (OLAP Híbrido): Uma combinação dos métodos ROLAP e MOLAP.

Os métodos mais comuns de armazenamento de dados utilizados pelos sistemas OLAP são ROLAP e MOLAP. O ROLAP usa a tecnologia RDBMS (Relational DataBase Management System), na qual os dados são armazenados em uma série de tabelas e colunas. Enquanto o MOLAP usa a tecnologia MDDB (MultiDimensional Database), onde os dados são armazenados em arrays multidimensionais.

Microsoft SQL Server oferece suporte a todos os três modos de armazenamento básico.



# OLAP



ROLAP é mais indicado para DATA WAREHOUSE pelo grande volume de dados, a necessidade de um maior número de funções e diversas regras de negócio a serem aplicadas.

MOLAP é mais indicado para DATA MARTS, onde os dados são mais específicos e o aplicativo será direcionado na análise com dimensionalidade limitada e pouco detalhamento das informações.

# Business Intelligence x Analytics

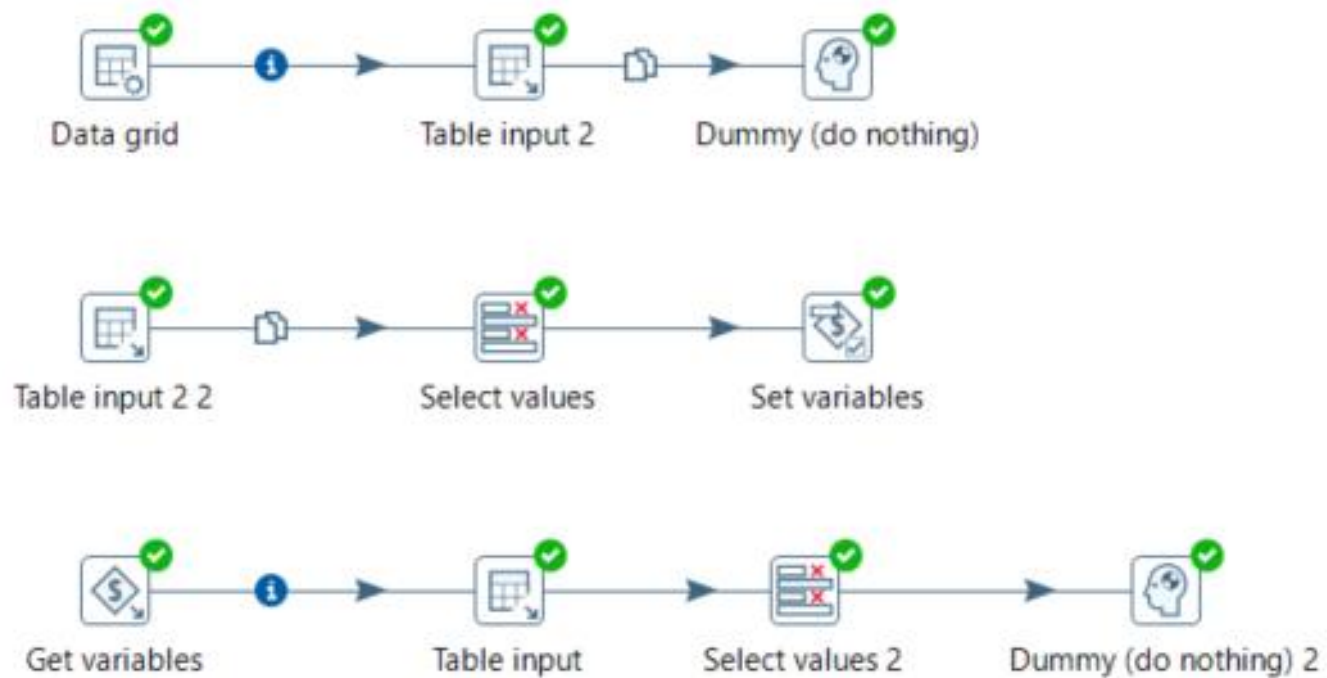


Área	Analista de BI	Cientista de Dados
Foco	Relatórios, KPI's, Tendências	Padrões, Correlações, Modelos Preditivos
Processo	Estático, Comparativo	Exploratório, Experimental, Visual
Fontes de Dados	Data Warehouses, Bancos Transacionais	Big Data, Dados Não-Estruturados, Bancos Transacionais e NoSQL, Dados Gerados em Tempo Real
Qualidade dos Dados na Fonte	Alta	Baixa ou Média (requer processo de limpeza e transformação)
Modelo de Dados	Esquema de dados bem definido na fonte	Esquema de dados definido no momento da consulta
Transformações nos Dados	Pouca ou nenhuma (dados já organizados na fonte)	Transformação sob demanda, necessidade de complementar os dados
Análise	Descritiva, Retrospectiva	Preditiva, Prescritiva
Responde à pergunta:	O que aconteceu?	O que pode acontecer?

# Ensaio no Pentaho



Trabalhando com a passagem de parâmetros.





# Ensaio no Pentaho



Trabalhando com Rest.

