

FIAP GRADUAÇÃO

# DATA SCIENCE

## DATA GOVERNANCE & DATA SECURITY MANAGEMENT

Prof. Dr. Renê de Ávila Mendes

# Objetivos da disciplina

**DISCIPLINA:** Data Governance & Data Security Management

**OBJETIVOS:** Descubra como funciona um **projeto de banco de dados** dentro de um ambiente corporativo, aplicando **técnicas de levantamento e documentação de requisitos**, aderente aos projetos de bancos de dados e aprenda a representar esses requisitos em arquiteturas de solução tecnológica para Data distribution e Data integration, modelos de estruturas de dados e dicionários de dados buscando **Data quality**. Garanta a qualidade dos dados de uma empresa para prover os melhores subsídios à tomada de decisão de negócio, praticando **Data cleaning** para limpar, harmonizar, complementar e corrigir dados inconsistentes, incompletos ou incorretos. Compreenda como funciona o **ciclo de vida da informação** e as responsabilidades administrativas sobre os dados de negócio, buscando qualidade, segurança e compatibilidade com políticas de administração de informação corporativas auditáveis, aplicando práticas atuais de **Data profiling** e conhecendo os princípios de **Data auditing**, de forma a atender a **Lei Geral de Proteção de Dados (LGPD)**.

## **Assuntos – 2º Semestre**

- Qualidade em metadados
- Arquiteturas de integração e distribuição física de banco de dados
- Master Data Management e Data Hub
- Qualidade de dados
- Enterprise Data Management
- LGPD

An abstract graphic featuring several overlapping, colorful cloud-like shapes in shades of blue, green, yellow, orange, and purple. These shapes are filled with various white icons representing different concepts: a globe, a key, a flag, a download arrow, an eye, a document, a gear, a clock, a speech bubble, a person with a headset, a camera, a musical note, a lightbulb, a padlock, a heart, a magnifying glass, a calendar with the number 15, a stack of coins, a Wi-Fi symbol, a checkmark, and a shield. Arrows of various colors (blue, green, yellow, orange, purple) point in different directions, suggesting a flow or process. The entire graphic is set against a dark gray background.

# PADRÃO DE NOMENCLATURA

# **PADRÃO DE NOMENCLATURA DE DADOS**

- Abra o arquivo “padrao\_nomenclatura\_dados.txt”
- Leia o arquivo para entender o padrão
- 10’

An abstract graphic featuring several overlapping speech bubbles in blue, green, yellow, orange, and purple. Each bubble is filled with various white icons representing different concepts: a globe, a key, a flag, a download arrow, an eye, a document, a shopping cart, a dollar sign, a gear, a lightbulb, a padlock, a heart, a magnifying glass, a calendar with the number 15, a database cylinder, a person with a headset, a camera, a water drop, a network of nodes, musical notes, and a Wi-Fi symbol. Arrows of various colors and styles (solid, dashed, double-lined) point in different directions across the composition. The entire graphic is set against a solid grey background.

# ORACLE DATA MODELER EXERCÍCIO

## Exercício Oracle Data Modeler

- Abra o arquivo “dd\_dicionado\_dados.doc” e implemente a estrutura proposta usando o Oracle Data Modeler
- 30’



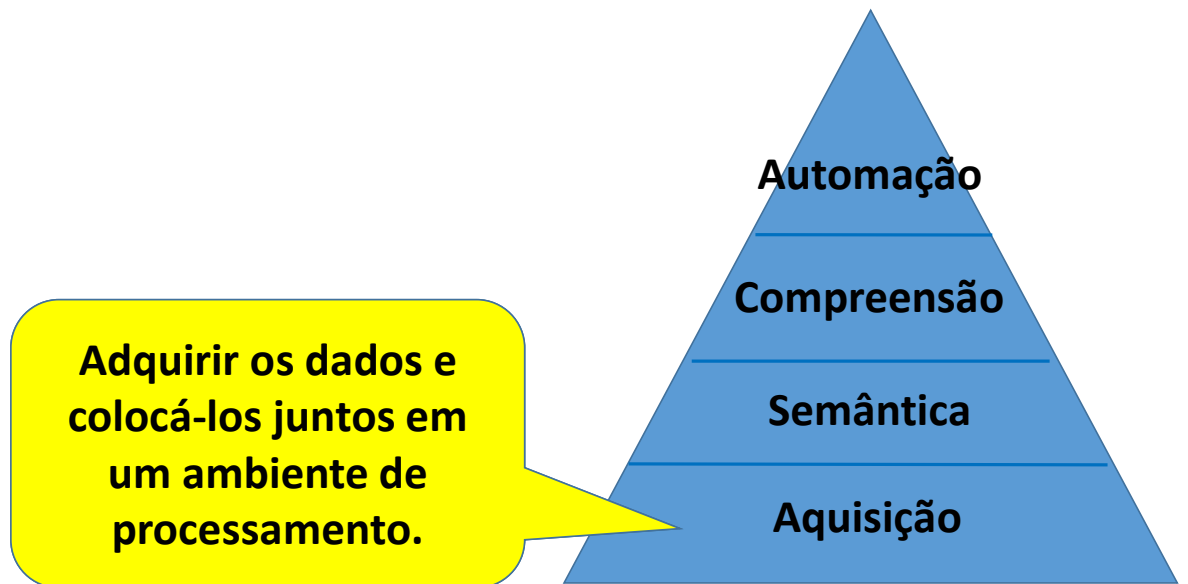
# INTEGRAÇÃO DE DADOS ENTRE BASES



# Técnicas de integração de dados

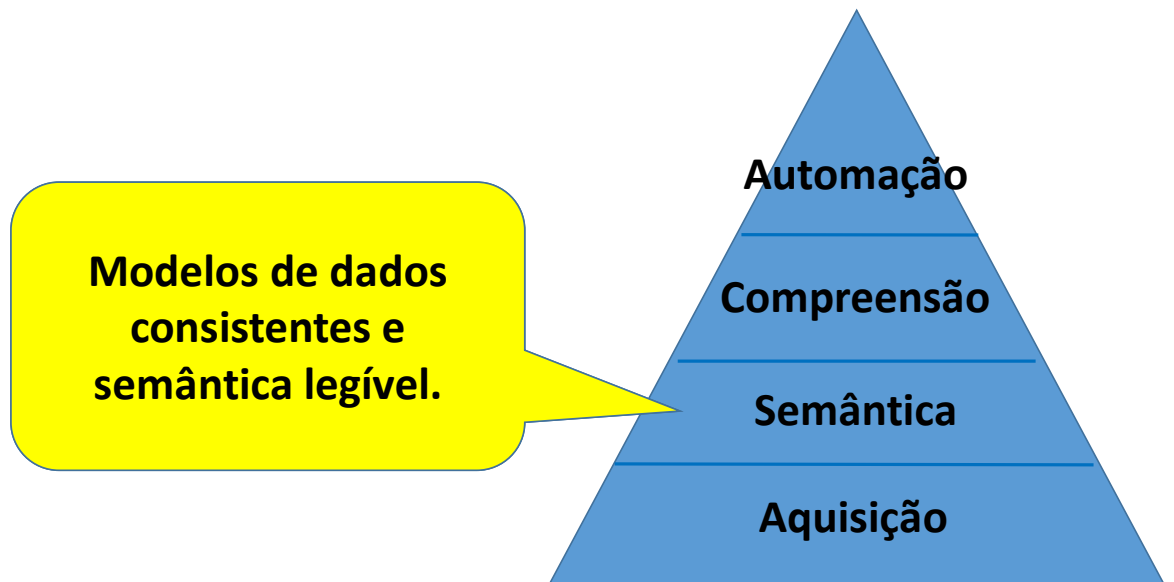
- Dados são considerados ativos corporativos
- Os ativos corporativos devem ser gerenciados
- O compartilhamento dos dados entre sistemas é necessário para que o dado seja:
  - Adquirido
  - Enriquecido
  - Corrigido
  - Consumido

## Hierarquia de uso dos dados



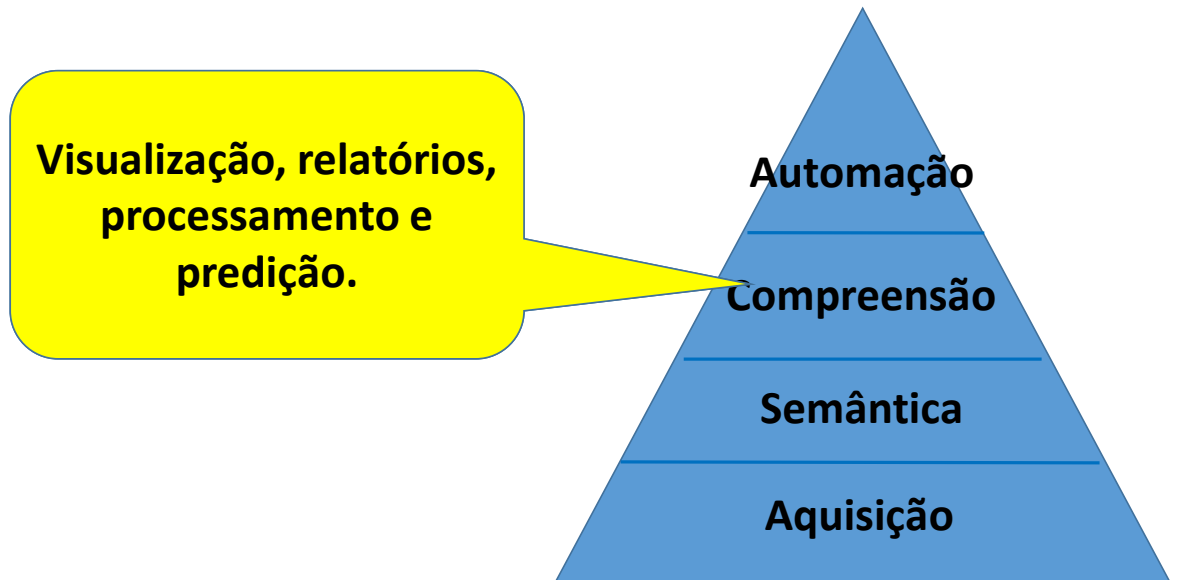
KREPS, Jay. *I Heart Logs: Event Data, Stream Processing, and Data Integration*. "O'Reilly Media, Inc.", 2014, p. 12.

## Hierarquia de uso dos dados



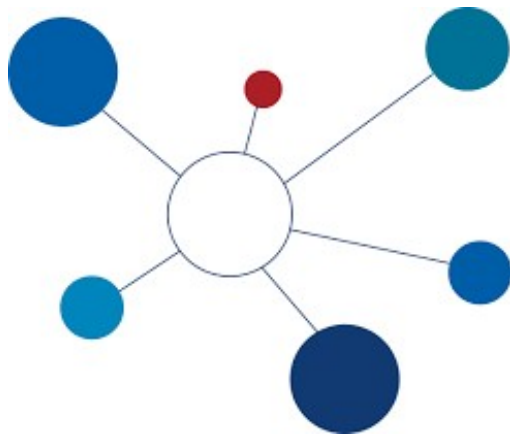
KREPS, Jay. **I Heart Logs: Event Data, Stream Processing, and Data Integration.** "O'Reilly Media, Inc.", 2014, p. 12.

## Hierarquia de uso dos dados



KREPS, Jay. **I Heart Logs: Event Data, Stream Processing, and Data Integration.** "O'Reilly Media, Inc.", 2014, p. 12.

## Integração de Dados - Definição



**Tornar disponíveis todos os dados que uma organização possui a todos os serviços e sistemas que necessitem deles.**

KREPS, Jay. **I Heart Logs: Event Data, Stream Processing, and Data Integration.** "O'Reilly Media, Inc.", 2014, p. 11.

## Cópia em mídia externa

- Método útil para situações em que os dados precisam chegar ao destino mais rápido do que outros métodos podem resolver
- Casos de uso:
  - Migração de dados offline entre data centers (devido a mudança de fornecedor de cloud, por exemplo)
  - Importação de dados para cloud
  - Transporte de dados críticos (diminui o risco de interceptação ou violação)

<https://docs.aws.amazon.com/snowball/latest/developer-guide/whatisedge.html>

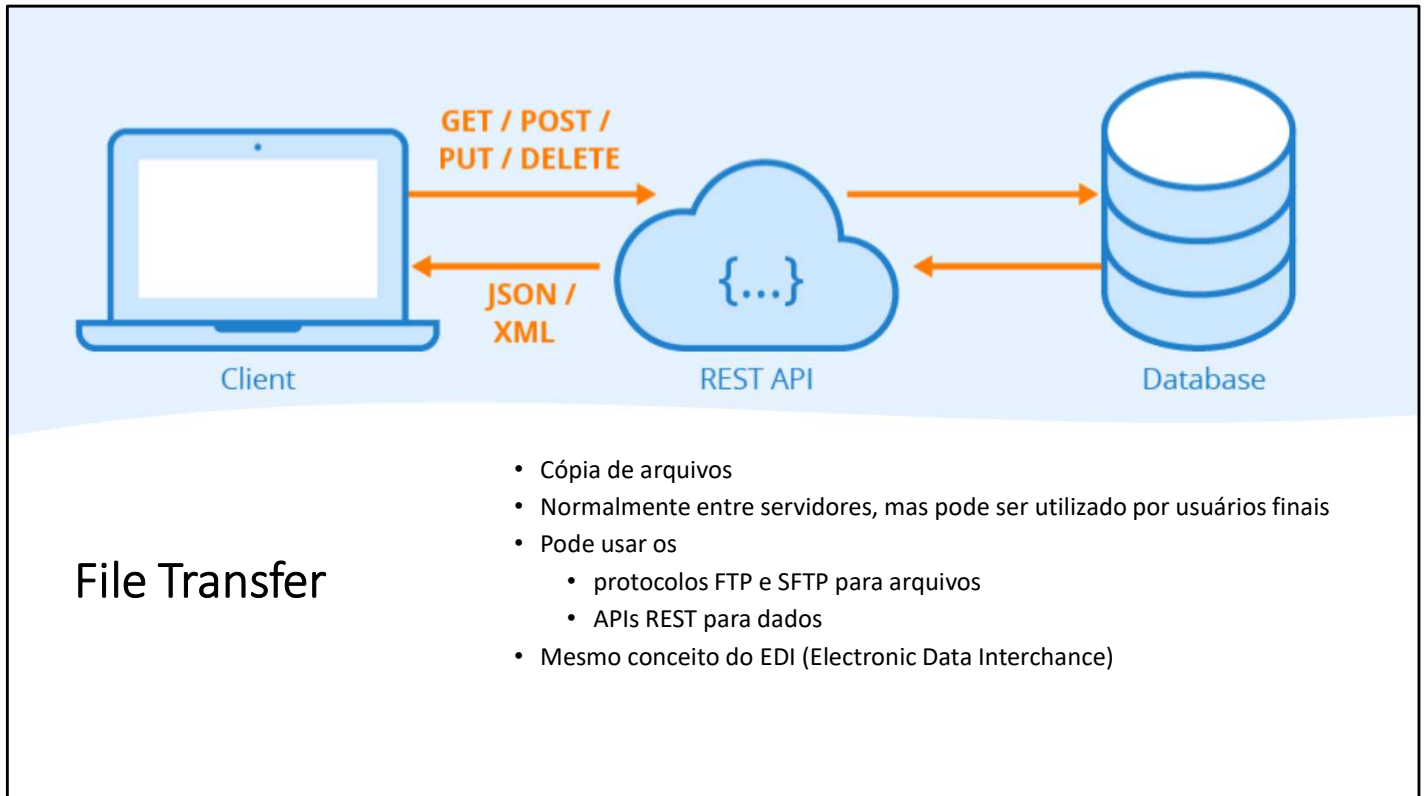
## Cópia em mídia externa - Exemplos

- AWS Snowball
  - 210 Terabytes
- AWS SnowMobile
  - 100 Petabytes
  - 26 anos a 10 Gbits



<https://docs.aws.amazon.com/snowball/latest/developer-guide/whatisedge.html>



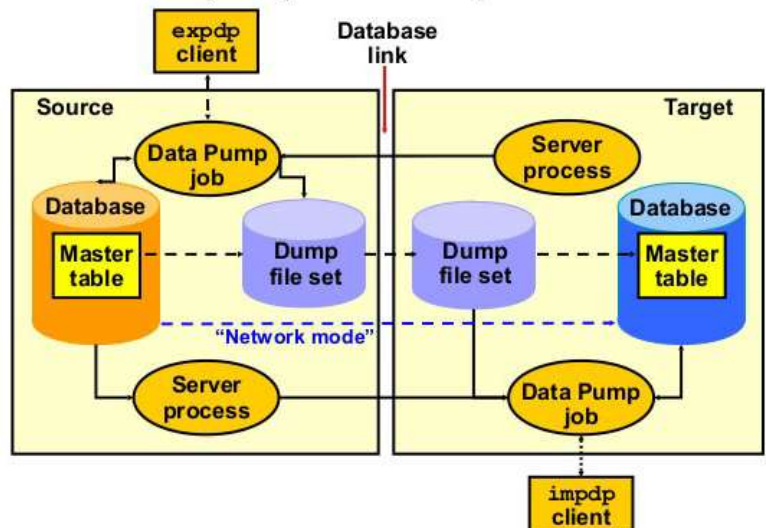


<https://docs.aws.amazon.com/snowball/latest/developer-guide/whatisedge.html>

# Data export/import

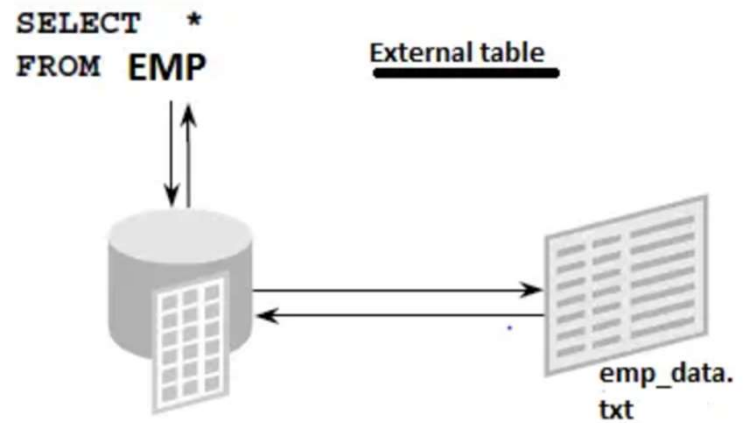
- Integração assíncrona de dados entre bases de dados
- Dados em formato binário proprietário ou SQL
- DDL em formato SQL

## Data Pump Export and Import: Overview

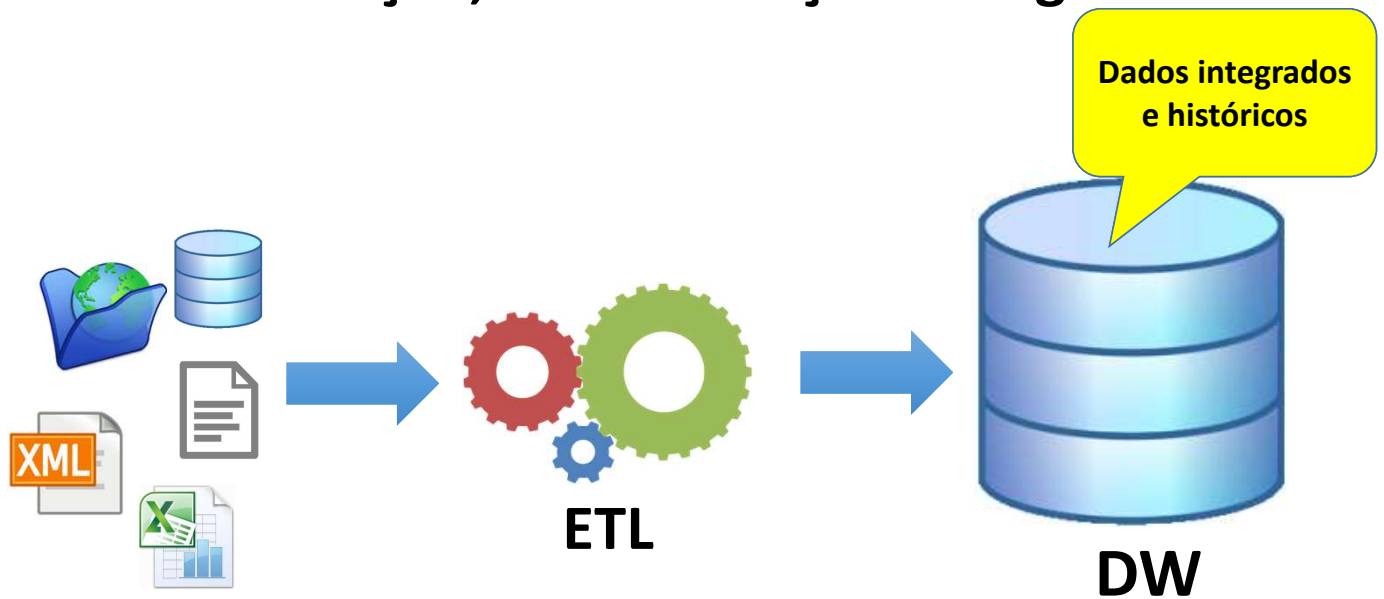


# External table

- Integração offline de dados entre bases de dados
- Os dados são lidos do arquivo quando um SELECT é executado sobre a tabela externa



## ETL – Extração, transformação e carga



## Comparando ETL com ELT

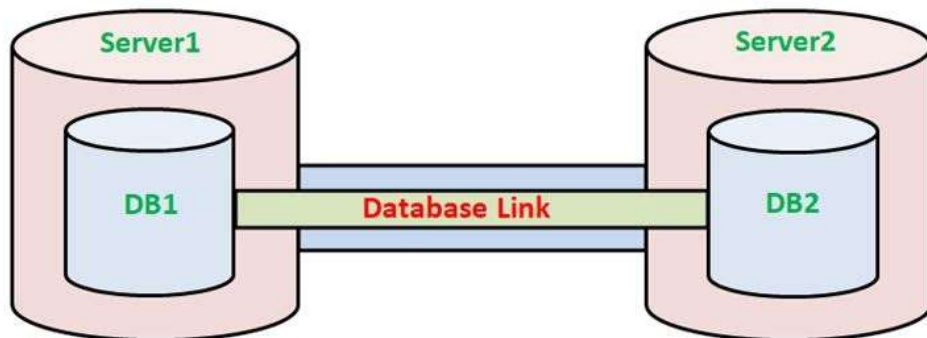


	ETL	ELT
Origem dos dados	Algumas	Todas
Transferência de dados	Lote	“Bulk” e streaming
Limpeza dos dados	Antes da carga	Prorrogada
Padronização de dados mestre	Antes da carga	Prorrogada

<https://www.lynda.com/Hadoop-tutorials/Comparing-big-data-ELT-traditional-ETL/385663/424483-4.html>

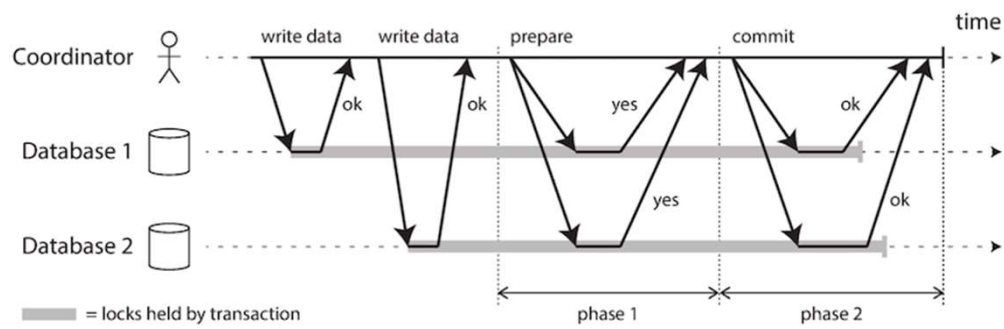
# Database link

- Integração síncrona de dados entre bases de dados
  - Mesmo produto (Oracle X Oracle)
  - Produtos diferentes (Oracle x SQL Server)
- Requer uma conexão aberta entre os bancos



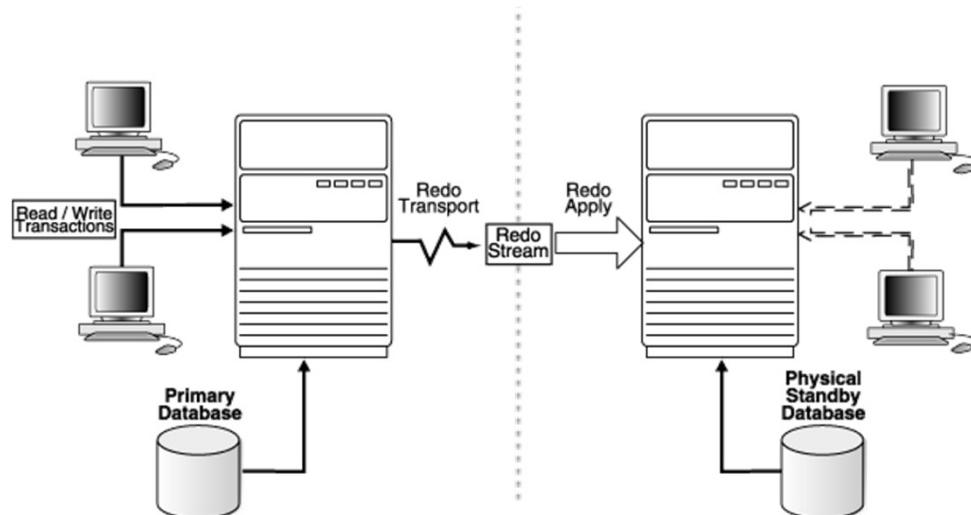
# Replicação síncrona

- Síncrona (two-phase commit)



# Replicação assíncrona

- Assíncrona (standby)

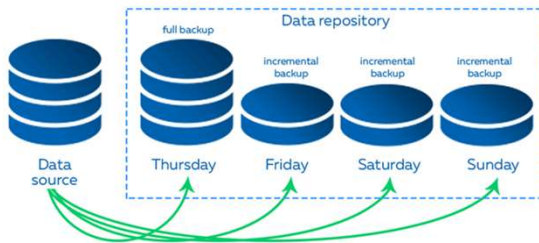




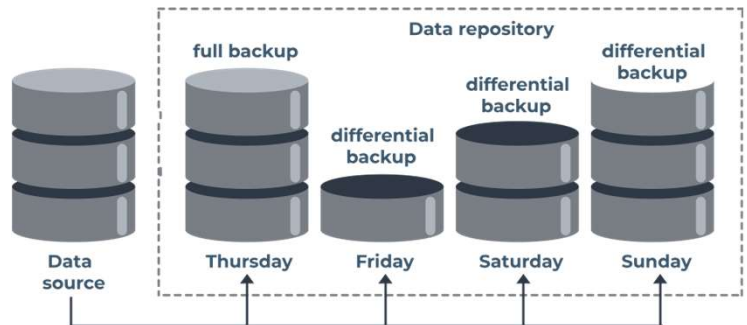
# Backup e restore

- Cópia dos dados para uma mídia
- Permite a restauração dos dados
- Incremental e diferencial

## Incremental



## Differential



# Snapshot

- Duplicação idêntica do banco de dados ou de um sistema inteiro
- Funcional: utilizado para duplicar um ambiente ou servidor ou uma base de dados

# Web services

- Transferência de dados entre aplicações
- Utilizam protocolos como SOAP (Simple Object Access Protocol) e REST (Representationl Transfer Protocol)
- Trocam dados em XML, JSON, CSV etc.

