



Análise de Dados e TI aplicado à Gestão



Módulo 3

Engenharia de Dados

Teoria

1. DATA SOURCES (FONTES DE DADOS)

O que é



São os locais de onde os dados vêm. É o ponto de origem da informação.

Exemplos:

1. Meta Ads (anúncios do Facebook/Instagram)
2. Google Analytics
3. CRM (como HubSpot, RD Station)
4. Planilhas do Excel
5. Sistemas internos (como ERP)
6. Sites, bancos de dados, APIs

Em resumo: é onde os dados “nascem”.



Teoria

2. DATA COLLECTION (COLETA DE DADOS)

O que é —————→

É o processo de extrair dados dessas fontes e levá-los para um local onde possam ser armazenados e analisados.

Exemplos:

1. Como funciona?
 - a. Pode ser feito manualmente (exportar planilhas);
 - b. Ou automaticamente com ferramentas como:
 - i. Airbyte
 - ii. Fivetran
 - iii. Zapier (via APIs simples)
 - iv. Scripts em Python

A coleta pode ser em tempo real ou em blocos (batch, por exemplo a cada dia).



Teoria

3. DATA STORAGE (ARMAZENAMENTO EM NUVEM)

O que é —————→

Data Storage é a etapa onde os dados são armazenados após serem coletados, antes de serem analisados.

1. Cloud Storage

a. Um “HD online” onde você armazena arquivos como planilhas, CSVs, imagens, PDFs, etc. Serve para guardar com segurança e acessar de qualquer lugar.

b. Ferramentas:

- i. Google Cloud Storage
- ii. Amazon S3
- iii. Azure Blob Storage.



Teoria

3. DATA STORAGE (ARMAZENAMENTO EM NUVEM)

3. Data Lake

a. O que é? Um repositório de dados brutos (não estruturados ou semi-estruturados). Aceita qualquer tipo de dado: logs de sistemas, vídeos, JSON, cliques de usuários, etc. Não impõe estrutura rígida: você guarda tudo para usar quando/quiser.

a. **Ferramentas:**

- i. AWS Lake Formation
- ii. Google Cloud BigLake
- iii. Azure Data Lake Storage
- iv. Databricks Delta Lake

Ideal para empresas que querem armazenar tudo e só depois decidir como e quando organizar.



Teoria

3. DATA STORAGE (ARMAZENAMENTO EM NUVEM)

4. Data Processing (Processamento de Dados)

a. O que é? É a etapa de transformação dos dados armazenados, onde eles são:

- i. Limpados
- ii. Organizados
- iii. Combinados (joins)
- iv. Preparados para análise

Ferramentas:

- Google BigQuery (data warehouse com processamento embutido)
- Apache Spark
- Dataflow

O que acontece aqui?

Você aplica regras de negócio, remove duplicatas, trata erros e deixa os dados prontos para serem analisados (por dashboards, IA, relatórios, etc.).



Teoria

4. DATA WAREHOUSE (ARMAZEM DE DADOS)

O que é? Um ambiente de armazenamento de dados estruturados, criado para consultas e análises em grande escala.

- i. Usa tabelas organizadas (como planilhas grandes);
- ii. Otimizado para leitura e agregação;
- iii. Ideal para gerar relatórios, dashboards e análises.

Ferramentas:

- Google BigQuery
- Snowflake
- Amazon Redshift
- Azure Synapse

O que acontece aqui?

Você aplica regras de negócio, remove duplicatas, trata erros e deixa os dados prontos para serem analisados (por dashboards, IA, relatórios, etc.).



5. DATA ANALYSES (ANÁLISE DE DADOS)

A etapa de **Data Analyses** transforma dados em inteligência acionável. Com os dados já limpos e organizados, é aqui que aplicamos modelos de inteligência artificial, geramos relatórios analíticos e criamos visualizações que facilitam a interpretação.

AI Models

- São modelos de **inteligência artificial** e **machine learning** aplicados sobre os dados.
- Usados para prever tendências, identificar padrões ou gerar alertas automáticos.

Reports

- Relatórios analíticos que mostram métricas, indicadores e resultados com base nos dados tratados.
- Podem ser dashboards automáticos, PDFs, apresentações ou relatórios enviados por e-mail.

Data Visualization

- Conecta os modelos de IA e os relatórios em formatos visuais (gráficos, painéis, dashboards).
- Ajuda na compreensão rápida dos dados para quem toma decisões.
- Ferramentas típicas: Looker Studio, Power BI, Tableau.



Teoria

ONDE O SQL ENTRA

O SQL (Structured Query Language) entra como a linguagem declarativa principal para consultar, organizar e preparar os dados que serão analisados.

Por que o SQL é usado no Data Warehouse?

Porque o Data Warehouse (como o BigQuery, Snowflake, Redshift) organiza os dados em tabelas estruturadas, exatamente como um banco de dados relacional. E o SQL foi feito para consultar esse tipo de estrutura. O SQL é como a “chave de leitura” do Data Warehouse.



Teoria

ESTRUTURA

[Data Sources]



[Data Storage] ← (dados brutos)



[Data Processing] ← (limpa, organiza, transforma)



[Data Warehouse] ← (dados prontos para análise)



Dashboards / IA / Decisões



SQL

(utilizado para realizar
consultar e análises)

Consultar

Analisar

Data Warehouse

Data Sources

Meta Ads

Google Analytics

CRM

Data Collection

Data Storage

Cloud Storage

Data Lake

Data Processing

Data Analyses

Data Analyses

AI Models

Reports

Data Visualization

Decision Making

Strategic Adjustments

Fim

Início

Data Sources

Data Storage

Data Analyses

AI Models

Reports

Data Visualization

Decision Making

Strategic Adjustments

Fim



Obrigado

Pedro Stob
Professor Convidado
dadoseti@preparaportugal.com
Instagram: @stobpedro
Linkedin: Pedro Stob

