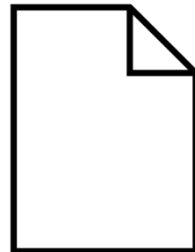
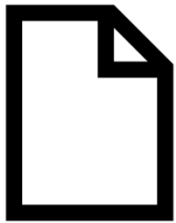


Formatos de Arquivos



- Dados podem ser armazenados em arquivos
- Estruturas específicas facilitam o armazenamento:
 - Escrita
 - Leitura
- Normalmente arquivos podem ser lidos por qualquer tipo de aplicativo

Formatos de Arquivos



- Texto/Csv (flat file)
- Json/XML
- ORC, Parquet
- Planilhas

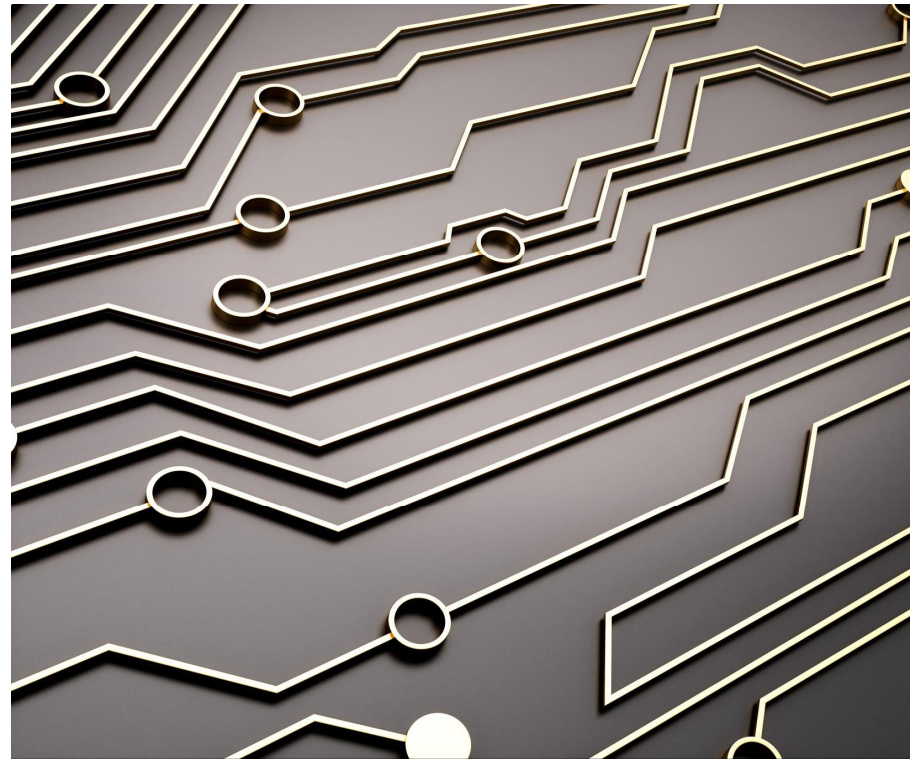


Sistemas de Arquivos

- Gerencia dados armazenados fisicamente:
 - Segurança
 - Versionamento
 - Metadados
 - Exemplo: NTFS
- Sistema de Arquivos Distribuídos
 - Características especiais para sistemas distribuídos
 - Exemplo: HDFS



**PROF.
FERNANDO
AMARAL**
www.datascientist.com.br



Sistemas Gerenciadores de Dados



Um software que faz a gestão dos dados, cuidando de aspectos como:

- Operações de leitura, inclusão, alteração, exclusão
- Segurança
- Disponibilidade
- Otimização

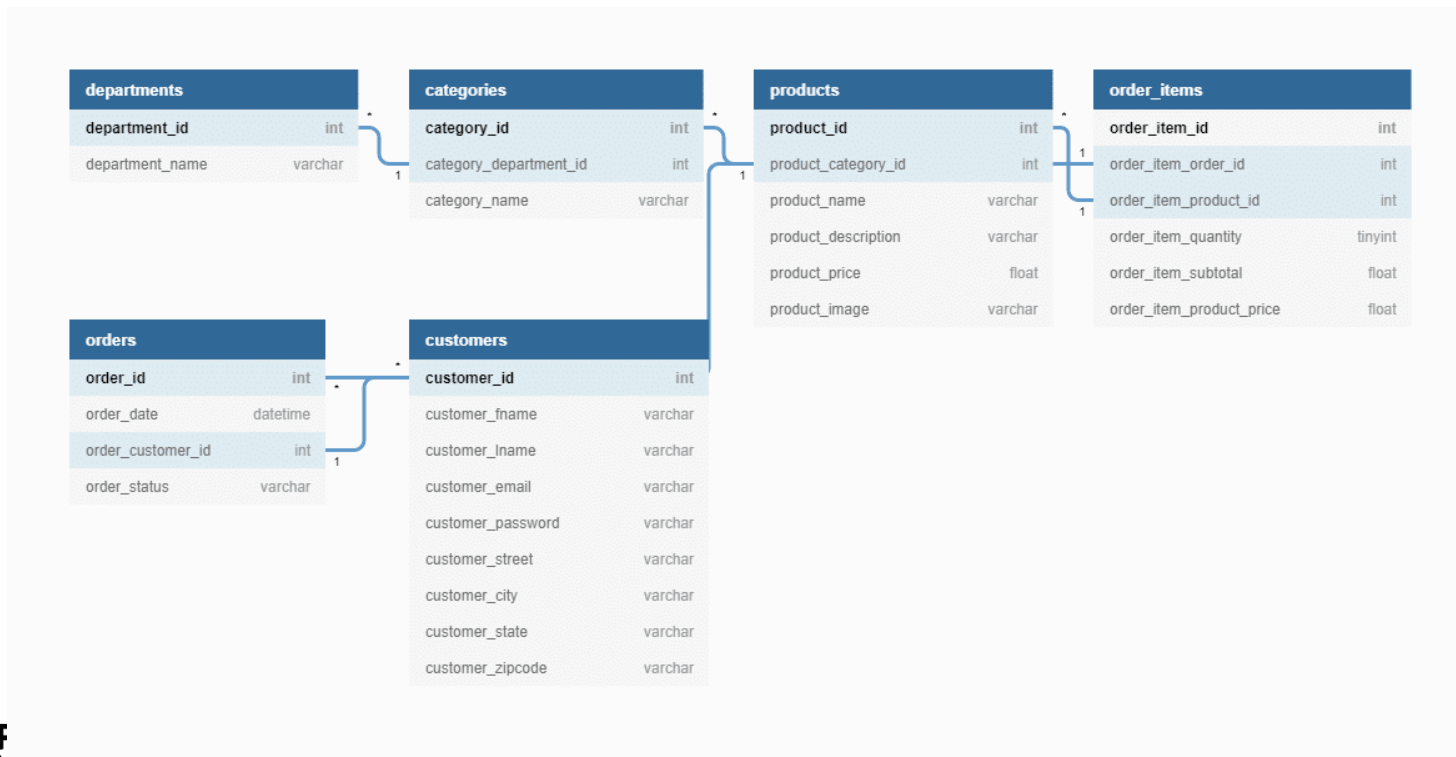


Os dados só podem ser operados pelo sistema gerenciador



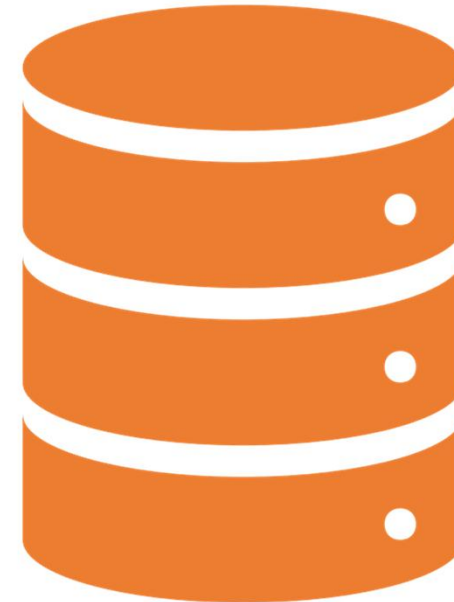
Exemplo: Sistemas de Gerenciadores de Banco de Dados Relacionais, como Oracle

Relacional



NoSQL

- Documento
- Chave-valor
- Orientado a Coluna
- Grafos



Documento

- Estrutura de dados em formato Json
- Exemplo: Mongodb



```
[
  {
    "description": "quarter",
    "mode": "REQUIRED",
    "name": "qtr",
    "type": "STRING"
  },
  {
    "description": "sales representative",
    "mode": "NULLABLE",
    "name": "rep",
    "type": "STRING"
  },
  {
    "description": "total sales",
    "mode": "NULLABLE",
    "name": "sales",
    "type": "INTEGER"
  }
]
```

Chave-valor

- Armazena estrutura simples: chave valor
- Usamos por exemplo, para sessões de web sites
- Exemplo: Redis



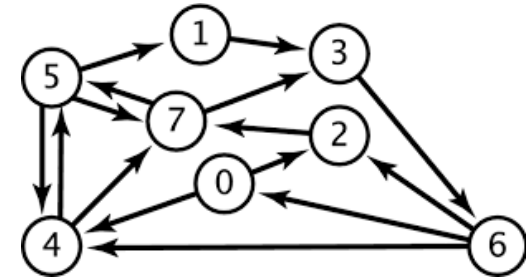
Orientado a Coluna

- Armazena dados por colunas, ao invés de linhas
- Ótimo para aplicações analíticas
- Exemplo: Amazon Redshift



Grafos

- Estruturas compostas por vertices e arestas
- Analise de redes sociais, dissiminação de doenças, menor trajeto
- Exemplo: Neo4j



Busca

- Sistemas que indexam documentos e permitem buscas
- Suportam erros de digitação, sinônimos etc
- Exemplo: Elasticsearch



elasticsearch

Processamento

- Sistemas especializados em computação
- Alta performance
- Não são gerenciadores de banco de dados
- Exemplos: Spark, Pig



Streaming

- Sistemas capazes de processar dados a medida que são produzidos
- Em tempo real ou próximo a tempo real
- Podem capturar mudanças ou processar em mini-batches
- Exemplo: Kinesis (Aws), Flink



Flink

Ecosystem Hadoop

