



Introdução a Banco de Dados e Linguagem SQL

Big Data e NoSQL

Agenda

- ❑ Big Data
 - ❑ MapReduce
- ❑ Banco de Dados NoSQL:
 - ❑ Banco de Dados Modelo Chave-Valor
 - ❑ Banco de Dados Modelo Documentos
 - ❑ Banco de Dados Modelo Família de Colunas
 - ❑ Banco de Dados Modelo Grafos



Big Data

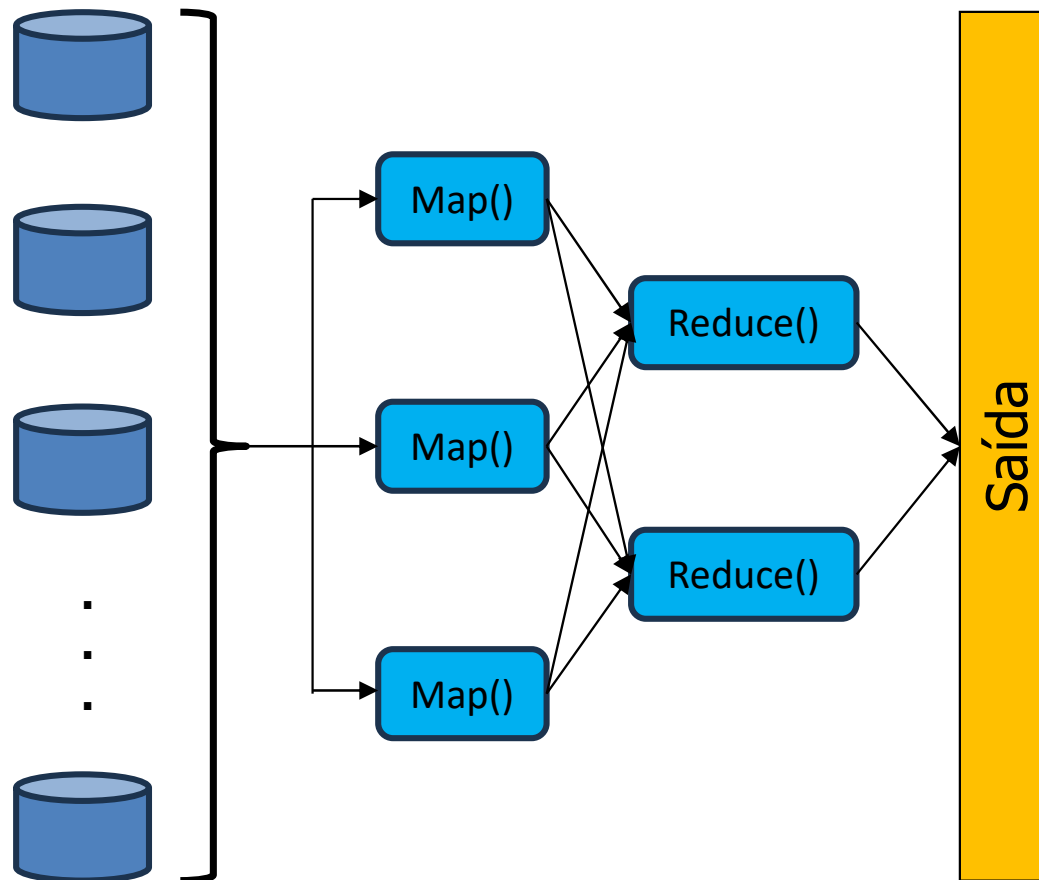
O big data é caracterizado por um grande `volume` de dados, adquiridos e processados em uma `velocidade` apropriada para não se tornarem obsoletos e disponibilizados em uma grande gama de `variedade` de formatos.

Big Data

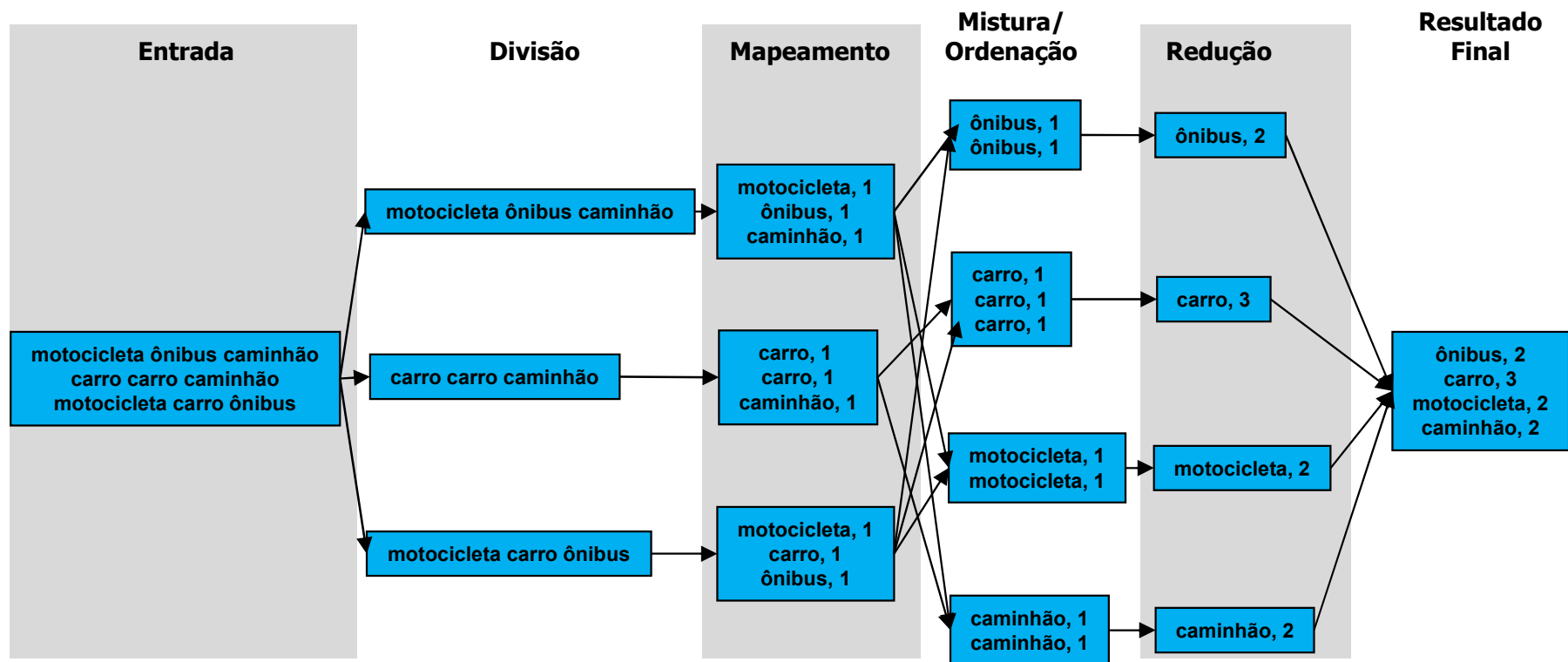
- ❑ O MapReduce, pode ser dividido em dois estágios principais:
 - ❑ **Mapeamento:** o nó mestre divide os dados em vários subconjuntos menores, um nó trabalhador processa um subconjunto de dados menor sob o controle de um rastreador de trabalho e armazena o resultado no sistema de arquivos local, onde um redutor será capaz de acessá-lo.
 - ❑ **Redução:** esta etapa analisa e reúne os dados de entrada a partir das etapas de mapeamento. Pode haver múltiplas tarefas de redução para paralelizar a agregação, e essas tarefas são executadas nos nós trabalhadores sob o controle do rastreador de trabalho.

MapReduce

Big Data



MapReduce



Bancos de Dados NoSQL

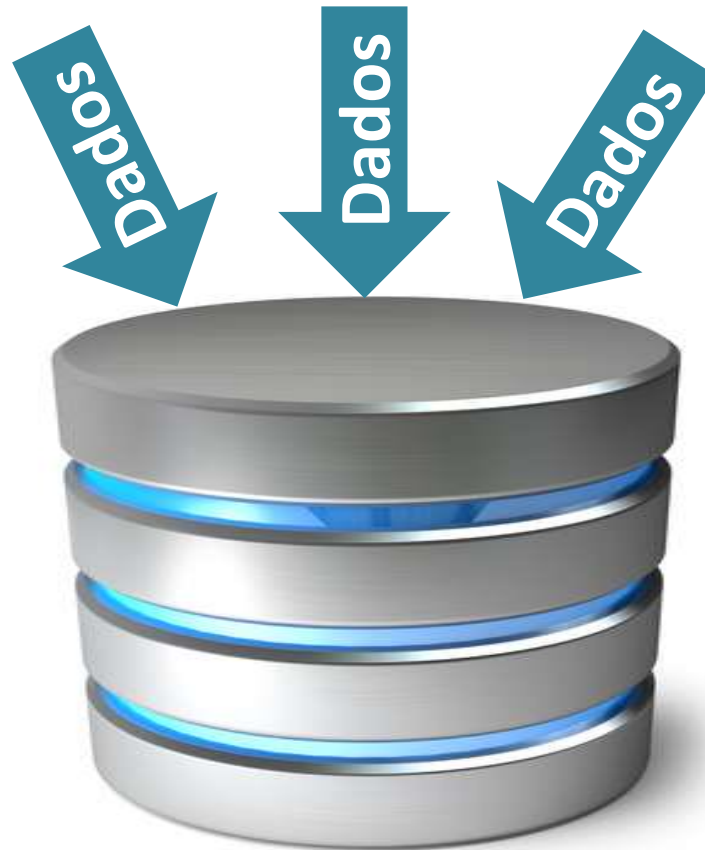
- ❑ Sofisticação crescente dos serviços disponibilizados na web:
 - ❑ Facebook, uma rede social com mais de 2 bilhões de usuários;
 - ❑ Google, um serviço de busca com mais de 1 bilhão de visitantes únicos por mês;
 - ❑ Amazon, uma loja de varejo on-line com picos que chegam a mais de 1 milhão de transações por segundo.
- ❑ O grande desafio: armazenar, recuperar e processar, de forma eficaz, a grande quantidade de dados estruturados e não-estruturados disponíveis.

Bancos de Dados NoSQL

❑ Necessidades:

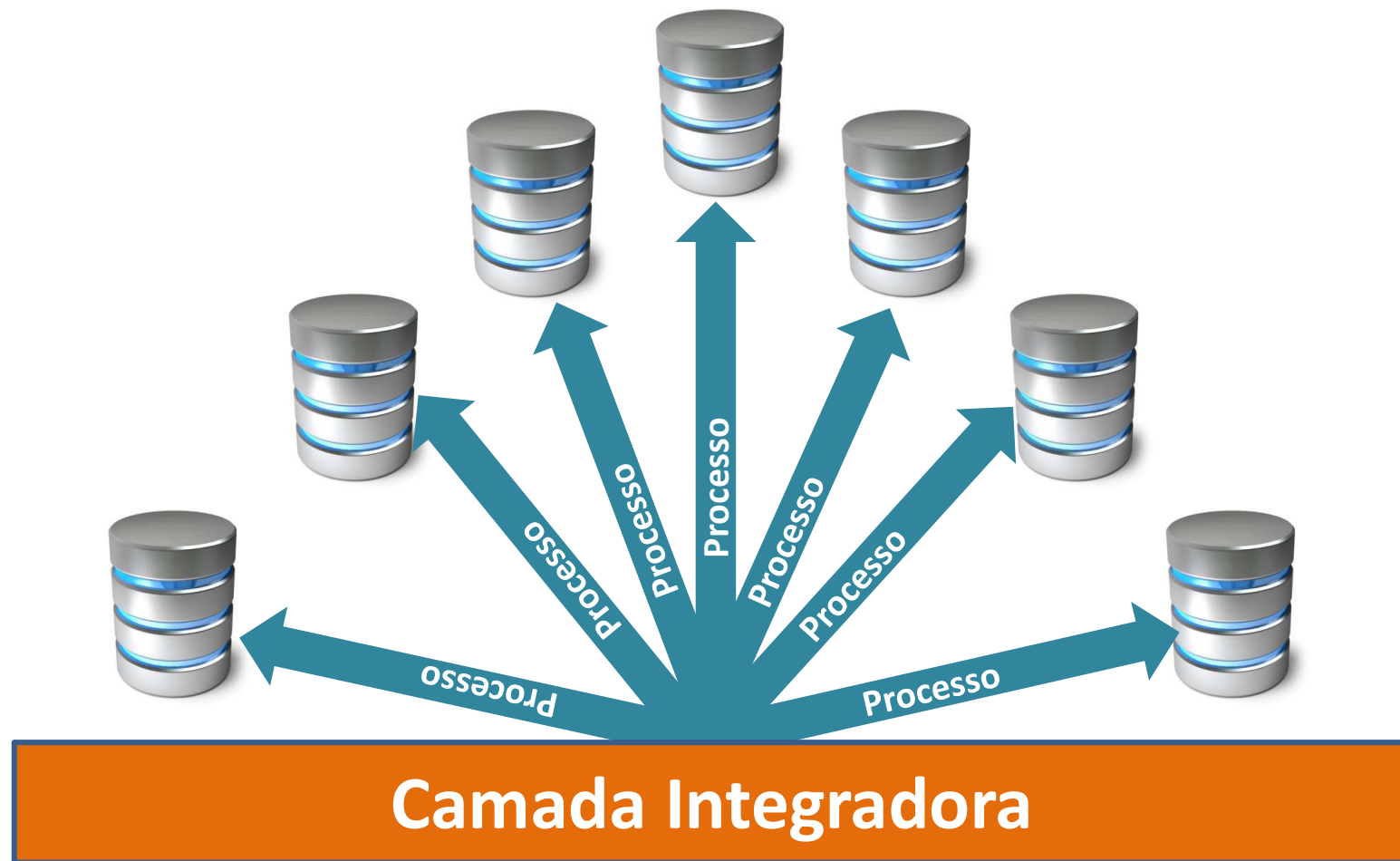
- ❑ processar o volume crescente de dados;
- ❑ esquemas flexíveis;
- ❑ maior distribuição do processamento;
- ❑ flexibilização das estruturas de armazenamento desses dados;
- ❑ escalabilidade horizontal;
- ❑ minimizar a incompatibilidade de impedância;
- ❑ lidar com documentos XML e JSON (JavaScript Object Notation).

Banco de Dados Convencional



Banco de Dados Relacionais

Banco de Dados Atual (NoSQL)



Bancos de Dados NoSQL

- ❑ As principais características dos SGBDs NoSQL são:
 - ❑ não utilizam o SQL (possuem linguagens semelhantes);
 - ❑ geralmente são projetos de código aberto;
 - ❑ a maioria dos BDs é orientada pela necessidade de execução distribuída;
 - ❑ oferecem uma gama de opções para consistência e distribuição; e
 - ❑ não possuem esquemas rígidos.

Bancos de Dados NoSQL

- ❑ Para atender a essas novas demandas surgiu uma nova geração de bancos de dados sem esquemas rígidos baseados em modelos não-relacionais, chamados bancos de dados NoSQL.

Bancos de Dados NoSQL

- ❑ O termo "NoSQL" foi utilizado pela primeira vez em 1990 por Carlo Strozzi para designar um BD relacional de código aberto que armazenava suas tabelas sob a forma de arquivos ASCII, e cada tupla era representada por uma linha com os campos separados por tabulação.
- ❑ O nome vem do fato de que o BD não utilizava SQL como linguagem de consulta. Em vez disso, ele era manipulado por shell scripts.
- ❑ Apenas em 2011 esse termo passou a representar uma família de bancos de dados que podem ser categorizados nos seguintes modelos: grafos, documentos, família de colunas e par chave/valor.

Bancos de Dados NoSQL

- ❑ O uso do termo “NoSQL” como conhecemos hoje é resultado de uma reunião realizada no dia 11 de julho de 2009, em São Francisco, organizada por Johan Oskarsson.
- ❑ Surgiu como um hashtag para o Twitter: curto, fácil de lembrar e sem muitos semelhantes no Google.
- ❑ O autor não esperava que se tornasse o nome da tendência tecnológica, pois o nome não descreve verdadeiramente esses sistemas.

Bancos de Dados NoSQL

❑ Obs:

Alguns autores alegam que “NoSQL” significa “Not Only SQL”. No entanto Fowler alega que, nesse caso, a escrita correta seria “NOSQL” . Além disso, essa definição causaria uma grande confusão, pois o Oracle e o Postgres se enquadrariam nela.

Bancos de Dados NoSQL

❑ Teorema CAP:

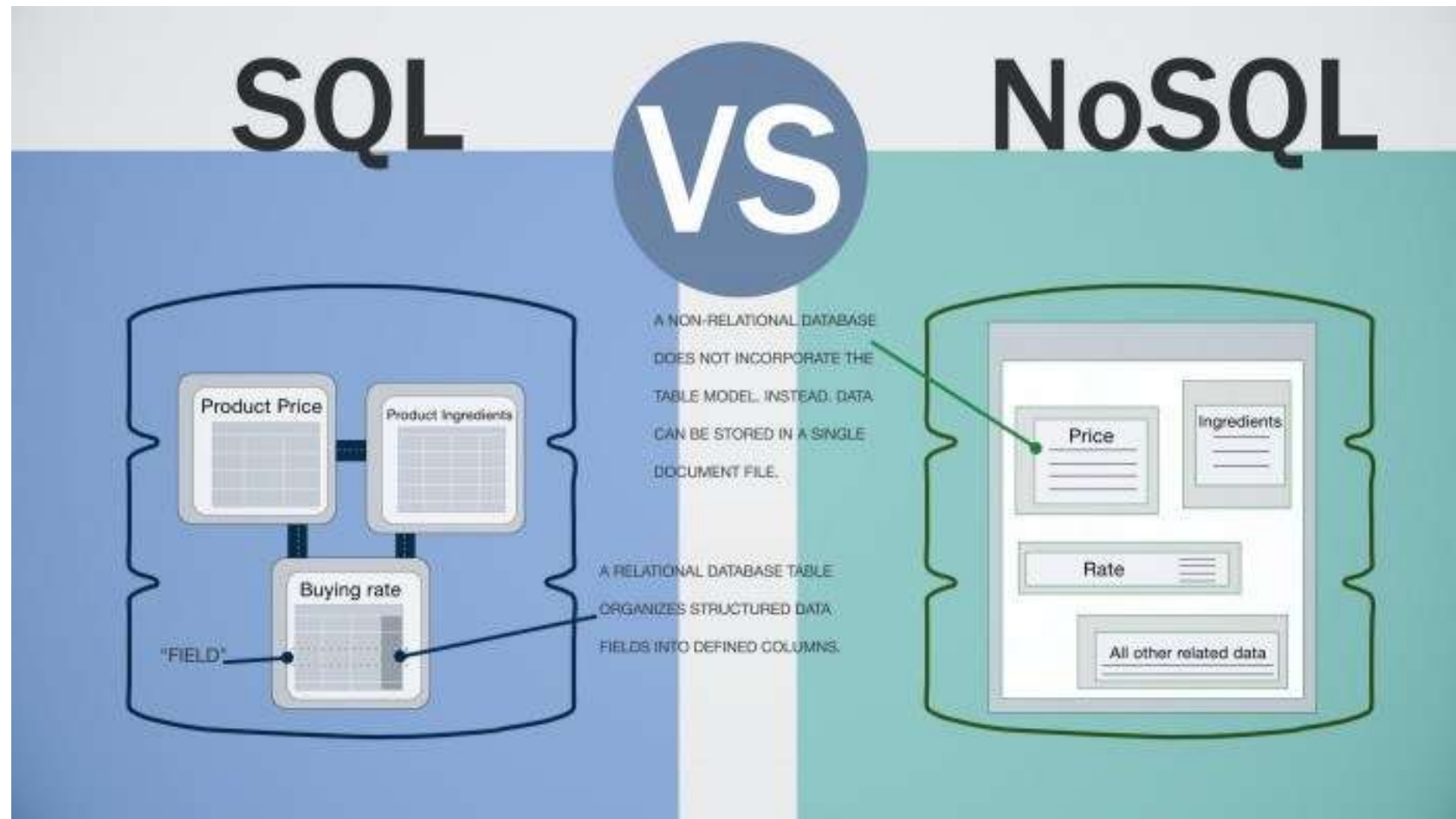
- ❑ Dadas três propriedades, Consistência, Disponibilidade (Availability) e de Particionamento, somente é possível obter duas delas.
- ❑ Servidor único: possui Consistência e Disponibilidade, mas não possui Particionamento.
- ❑ Embora o teorema CAP seja, muitas vezes, declarado como “só pode ter dois dos três”, na prática ele diz que, em um sistema que possa sofrer partições, como os sistemas distribuídos, você tem que balancear a consistência com disponibilidade. Nesse caso, o sistema resultante não seria nem perfeitamente consistente nem perfeitamente disponível, mas teria uma combinação razoável para suas necessidades específicas.

Bancos de Dados NoSQL

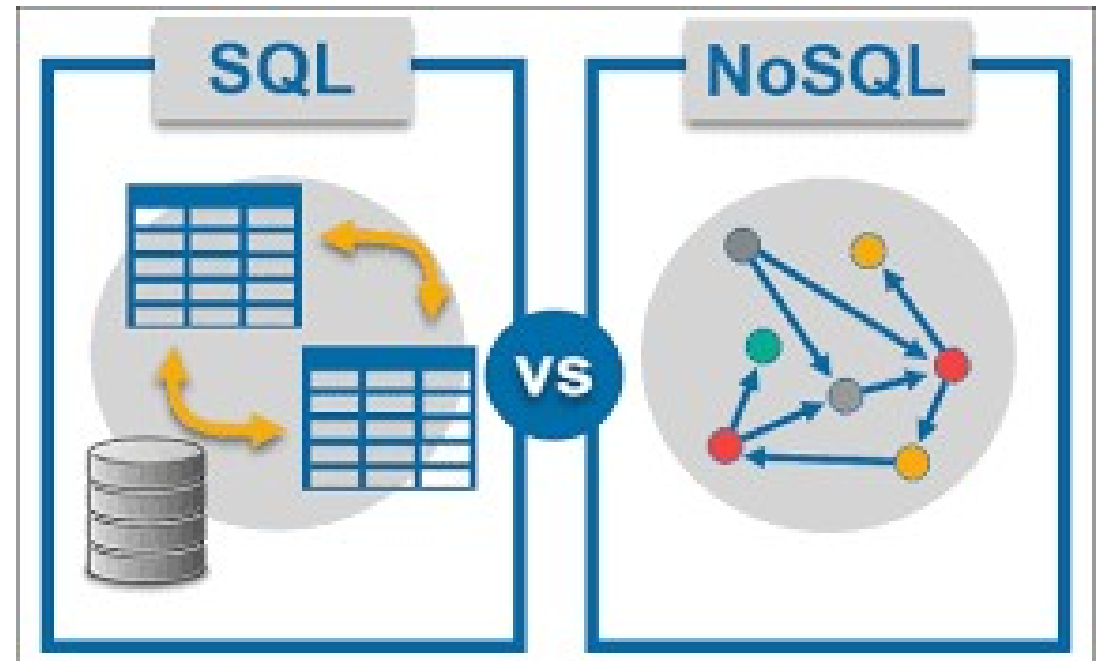
❑ ACID x BASE:

- ❑ BASE - Basically Available, Soft state, Eventual consistency.
- ❑ SGBDs NoSQL seguem as propriedades BASE.
- ❑ Não há definição consensual para “basicamente disponível” nem para “estado soft”.
- ❑ Consistência eventual significa que, em algum momento, todos os nodos em um sistema NoSQL podem ter inconsistências de replicação, mas, se não houver mais atualizações, todos os nodos acabarão atualizados com o mesmo valor.

Banco de Dados Atual (NoSQL)



Banco de Dados Atual (NoSQL)

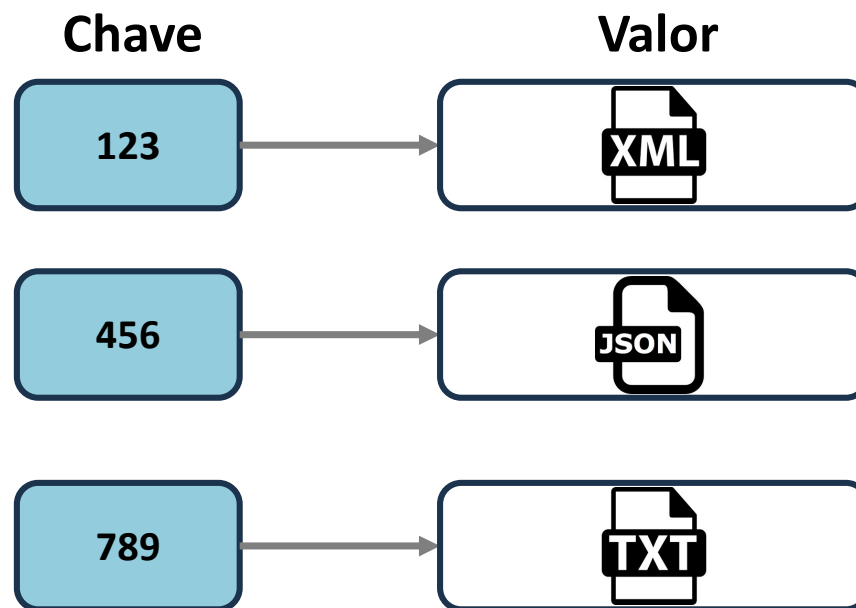


Bancos de Dados NoSQL

❑ Tipos de SGBDs NoSQL:

- ❑ Chave-Valor;
- ❑ Documentos;
- ❑ Família de Colunas; e
- ❑ Grafos.

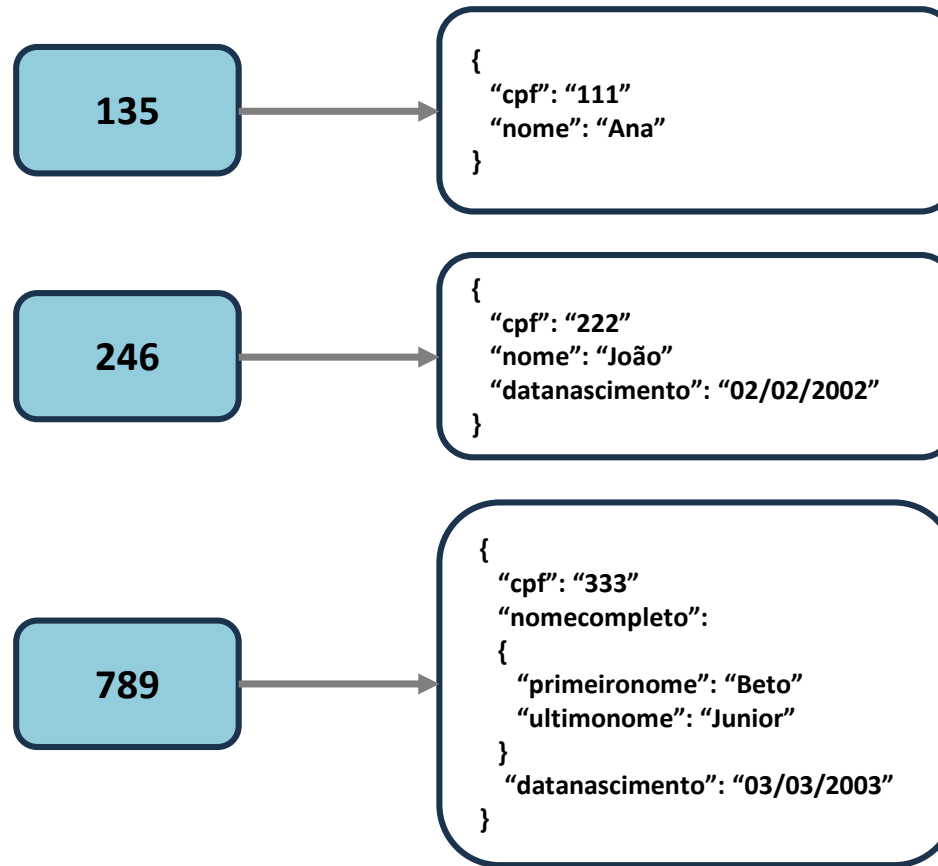
Modelo Chave-Valor



Modelo Documentos

Identificador

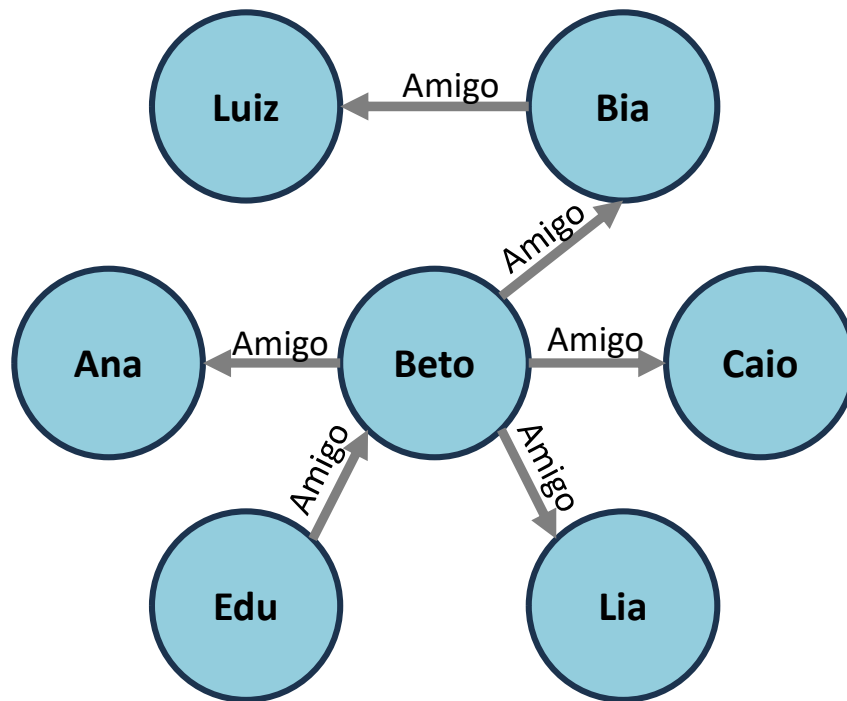
Documento

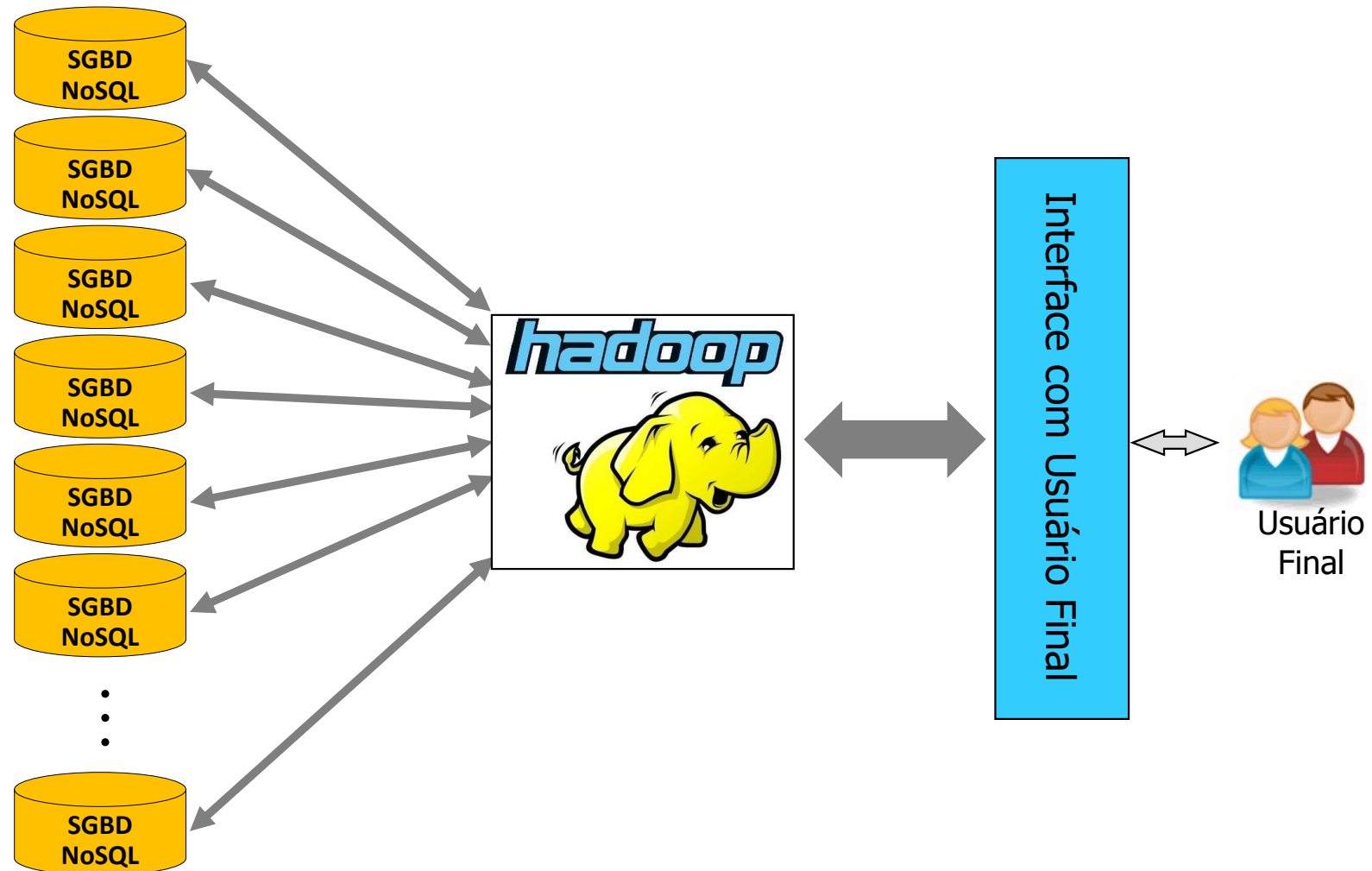


Modelo Família de Colunas

CPF	Nome	Sobrenome	Nascimento	Pais
111	Ana	Silva	01/01/2001	Brasil
222	Beto	Lima		
333	Caio	Abreu	03/02/2003	
444	Lia	Correia		

Modelo Grafos





Fim