

Bancos de Dados NoSQL

Processamento Analítico de Dados
Profa. Dra. Cristina Dutra de Aguiar Ciferri

Banco de Dados NoSQL

(Not Only SQL)

- Foco no armazenamento de gigantescos volumes de dados (**big data**)
- Características
 - não relacional
 - altamente escalável
 - eficientemente distribuído
 - em geral, de código aberto
 - replicação facilitada dos dados
 - SQL não é a linguagem padrão
 - alto desempenho
 - livre de esquema
 - ...

Relacional *versus* NoSQL

- Bancos de dados **relacionais**
 - dados estruturados
 - esquema bem definido
 - ... garantem as propriedades **ACID**
- Bancos de dados **NoSQL**
 - dados semiestruturados ou não estruturados
 - esquema flexível ou nenhum esquema
 - ... relaxam as propriedades ACID
 - teorema **CAP** e propriedades **BASE**

Propriedades ACID das transações

Atomicidade

Consistência

Isolação

Durabilidade

Propriedades ACID

- Atomicidade
 - *uma transação é uma unidade atômica*
 - todas as operações das transações são finalizadas e refletidas no BD ou nenhuma delas é finalizada e refletida
- Consistência
 - *transformações preservam a consistência*
 - a execução correta de uma transação leva o BD de um estado consistente a outro estado consistente

Propriedades ACID

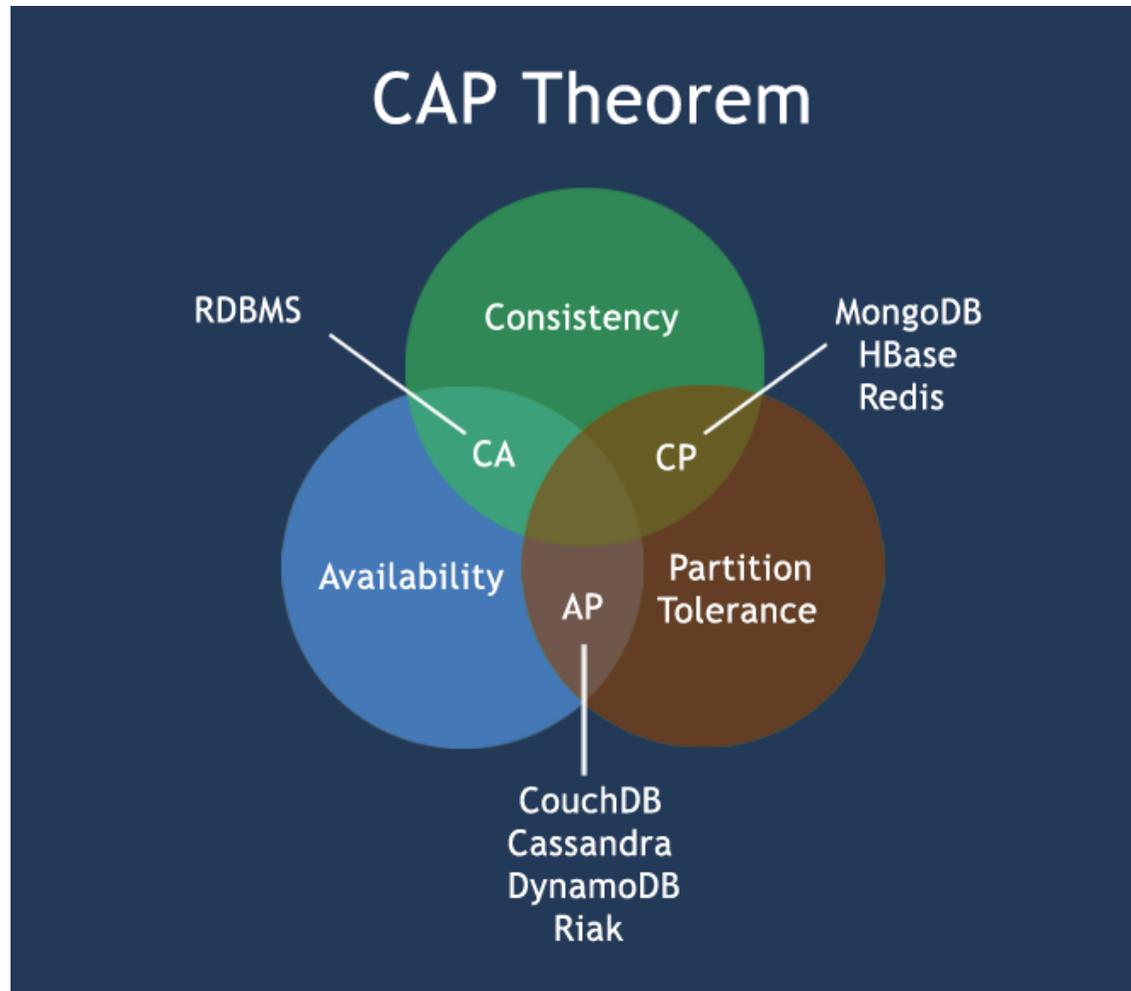
- Isolação
 - *transações são isoladas umas das outras*
 - cada transação assume que está sendo executada sozinha no sistema, e o SGBD garante que os resultados intermediários da transação são escondidos de outras transações executando concorrentemente
- Durabilidade
 - os valores dos dados alterados durante a execução de uma transação devem persistir após a finalização desta

Teorema CAP

- Propriedades
 - consistência (Consistency)
 - disponibilidade (Availability)
 - tolerância à partição (Partition Tolerance)
- Teorema
 - é impossível garantir essas três propriedades ao mesmo tempo
 - é possível garantir quaisquer duas dessas propriedades ao mesmo tempo

Gilbert, S.; Lynch, N. A. Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. SIGACT News, 33(2):51–59, 2002.

Teorema CAP



<http://blog.triscal.com.br/wp-content/uploads/2013/06/cap-theoram-image.png>

Propriedades BASE

- Propriedades
 - Basically Available
 - Soft state
 - Eventually concistent
- **Consistência eventual**
 - ponto fundamental
 - uma requisição de leitura logo após uma requisição de escrita pode ter como retorno o valor antigo

Pritchett, D. BASE: An ACID alternative. *ACM Queue*, 6(3):48–55, 2008.

Consistência Eventual

Considere uma operação de escrita op_1 no tempo t_1 que ocorreu no nó n_1 e que alterou o dado do valor x para y , e uma operação de leitura op_2 no tempo t_2 que foi atendida pelo nó n_2 .

Com base nas propriedades BASE, é possível que a operação op_2 tenha como resposta o valor x (ou seja, valor antigo) ao invés do valor y (ou seja, novo valor), dependendo do tempo decorrido entre t_1 e t_2 .

Entretanto, é garantido que se o tempo entre op_1 e op_2 for suficiente, op_2 terá como resposta o valor y .

Bancos de dados NoSQL

- Classificados pelo modelo de dados
 - chave-valor
 - documentos
 - famílias de colunas
 - grafos

Chave-Valor

- Características



- representa dados por conjuntos de chaves e valores associados

- proporciona bom desempenho para aplicações na nuvem

- oferece menor capacidade de busca, permitindo apenas busca por chave

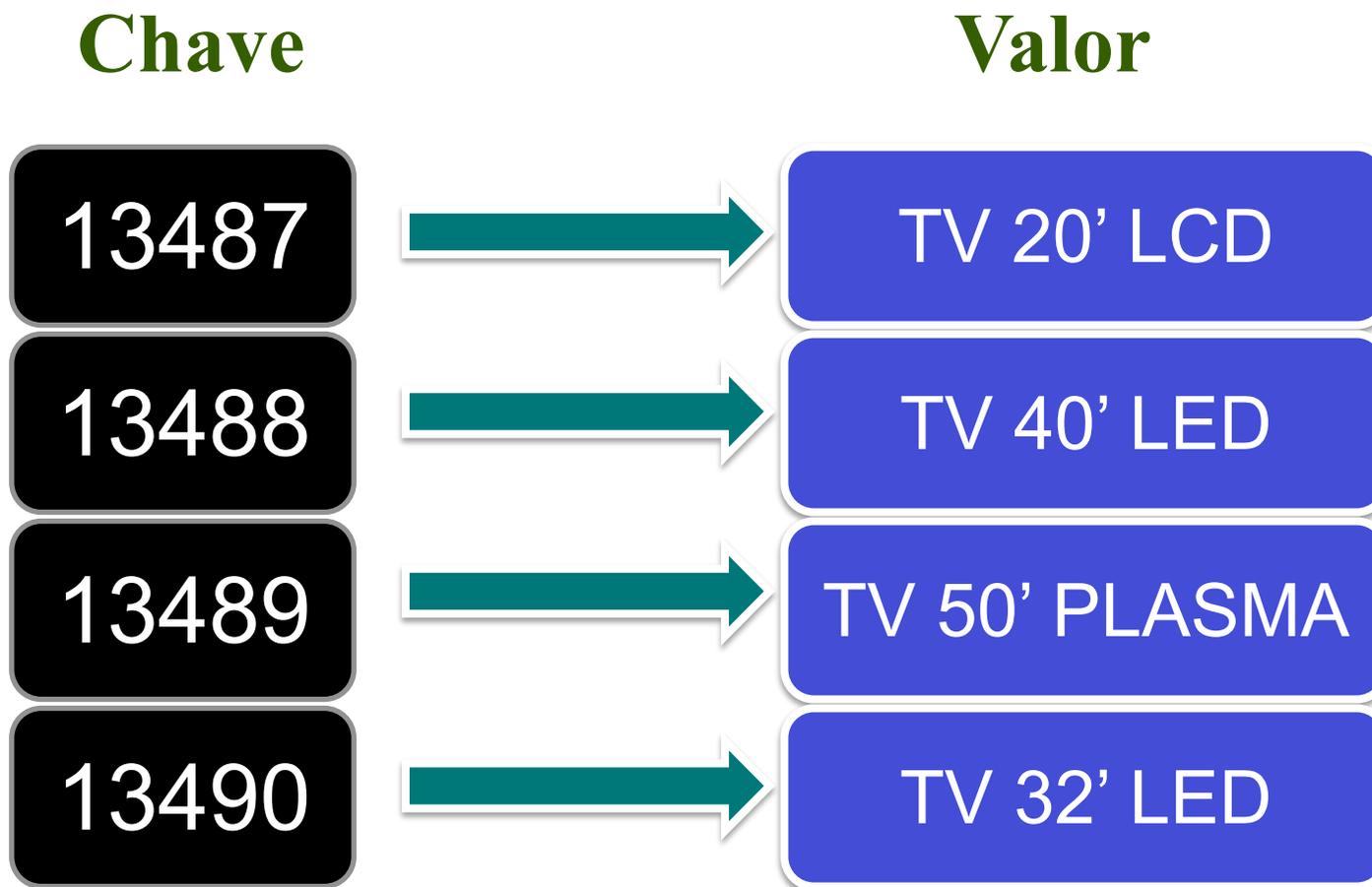


- Aplicações típicas

- gerenciamento de sessões



Chave-Valor: Exemplo



Documentos

- Características

- cada documento é uma coleção de pares chave-valor
- usualmente baseado nos formatos XML ou JSON
- permite a realização de consultas mais elaboradas, envolvendo filtros por atributos e a possibilidade de uso de índices



- Aplicações típicas

- aplicações Web



Documentos: Exemplo

```
{  
  "Id": 13487,  
  "Produto": "TV",  
  "Tamanho": "20",  
  "Tela": "LCD"  
}
```

```
{  
  "Id": 13488,  
  "Produto": "TV",  
  "Tamanho": "40",  
  "Tela": "LED"  
}
```

```
{  
  "Id": 13489,  
  "Produto": "TV",  
  "Tamanho": "50",  
  "Tela": "PLASMA"  
}
```

```
{  
  "Id": 13490,  
  "Produto": "TV",  
  "Tamanho": "32",  
  "Tela": "LED"  
}
```

Família de Colunas



- Características

- cada coluna é definida por uma tripla

- linha, coluna e *timestamp*

- definições adicionais

- supercoluna: coluna que contém outras colunas
- família de colunas: colunas de uma mesma família são armazenadas no mesmo conjunto de arquivos

- não implementa junções

- Aplicações típicas

- gerenciamento de grandes volumes de dados



Família de Colunas: Exemplo

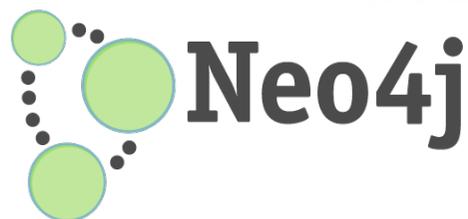
Famílias de Colunas

	Principais		Adicionais	
<i>Rowkey</i>	Produto	Tamanho	Tela	Tipo
13487				
13488		20'	LCD	
13489		40'	LED	
13490		50'	PLASMA	SMART
		32'	LED	SMART

Grafos

- Características
 - baseados na teoria dos grafos
 - dados são modelados por nós e arestas, os quais podem possuir propriedades
- Aplicações típicas
 - redes sociais
 - recomendações

InfiniteGraph
The Distributed Graph Database



Grafos: Exemplo

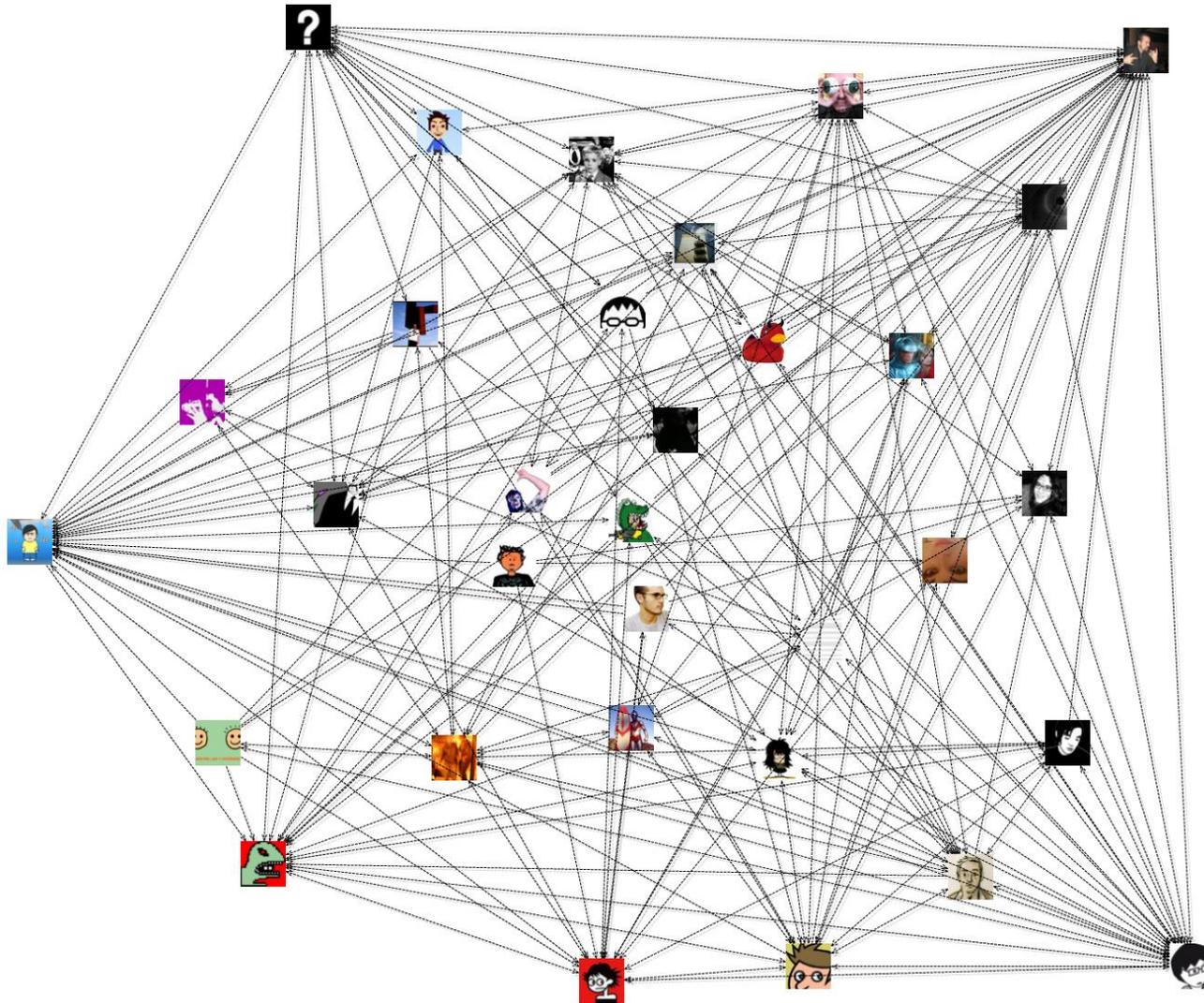


Imagen: <http://drzito.files.wordpress.com/2007/05/network.jpg>