



Bancos NOSQL

Ciência de Dados

Centro Universitário Senac

Prof. Celso Crivelaro

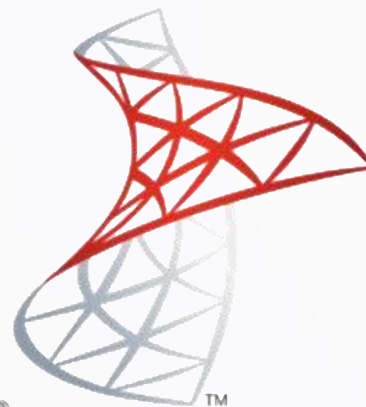
celso.vcrivelaro@sp.senac.br



Bancos RDBMS



ORACLE®



MySQL®

Microsoft®
SQL Server®

Senac





Bancos NoSQL



NoSQL = Not only SQL

Problemas de bancos RDBMS

Estrutura rígida

Problemas para armazenar para volumes gigantes de dados

Problemas para indexar para volumes gigantes de dados

Relacionamentos grandes são problemáticos

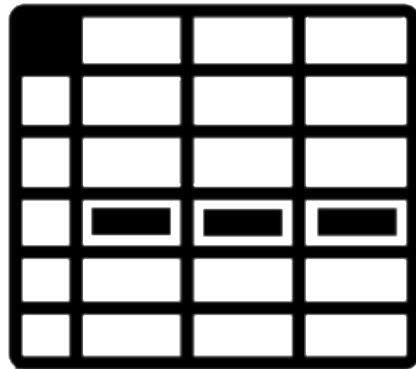
Dificuldade para escalar verticalmente

Tipos de Bancos NoSQL

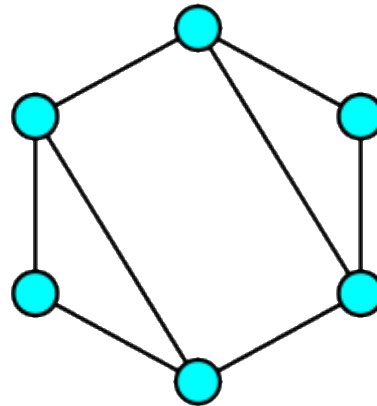
Documentos



Colunar



Grafos



**Chave -
Valor**



**Séries
Temporais**

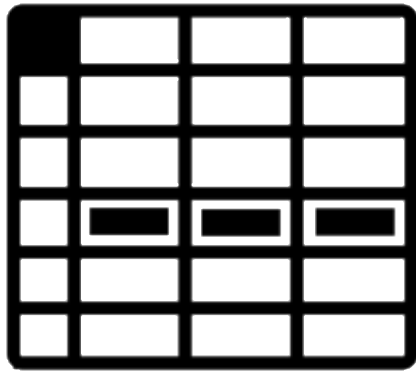


Tipos de Bancos NoSQL

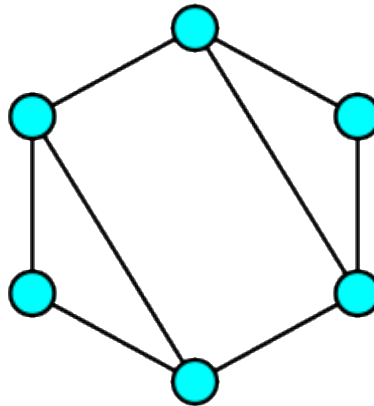
Documentos



Colunar



Grafos



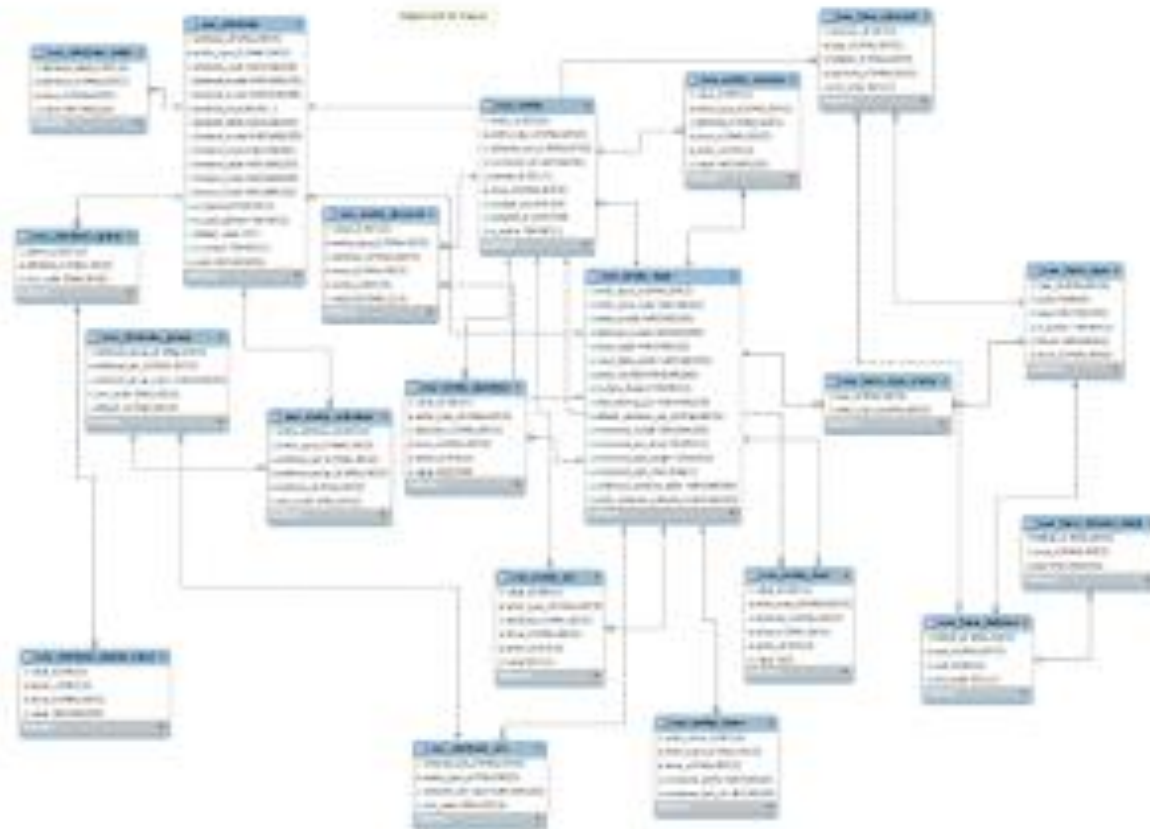
**Chave -
Valor**



**Séries
Temporais**



Relational Model



Document Model

```
{
  "_id" : "rp-prod132546",
  "name" : "Marvel T2 Athena",
  "brand" : "Pinarello",
  "category" : "bike",
  "type" : "Road Bike",
  "price" : 2949.99,

  "size" : "55cm",
  "wheel_size" : "700c",
  "frameset" : {
    "frame" : "Carbon Toryaca",
    "fork" : "Onda 2V C"
  },
  "groupset" : {
    "chainset" : "Camp. Athena 50/34",
    "brake" : "Camp."
  },
  "wheelset" : {
    "wheels" : "Camp. Zonda",
    "tyres" : "Vittoria Pro"
  }
}
```

Document Store

Similar a chave-valor

Fornece estrutura para dados semi-estruturados como XML, JSON, BSON

Não há relacionamentos entre os documentos, apenas estruturas embarcadas

Documentos podem ser armazenados em nós diferentes do banco

Documentos são a operação atômica neste banco



mongoDB®



Safe by Default, Optimized for Efficiency



CouchDB

Vantagens

Bom para leitura e gravação

Dados em que a estrutura muda

Permite sharding

Senac

Desvantagens

Repetição de dados

Complexidade de agregação

Senac



Indexadores / Buscadores



elasticsearch

Solr



Senac

Indexadores

Busca em memória de dados indexados

Tratamento de texto para busca de equivalência. Ex. Médico = Médica = Médicos

Usados em caixas de busca de sites

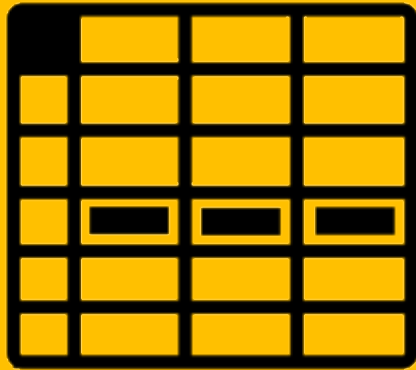
Senac

Tipos de Bancos NoSQL

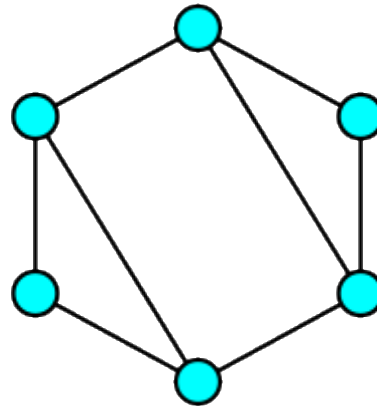
Documentos



Colunar



Grafos



**Chave -
Valor**



**Séries
Temporais**



Colunar

Formato parecido com o RDBMS

Formato de Armazenamento é feito em **COLONAS** em vez de Linhas

Virtualmente o número de colunas por ser ilimitado na definição do schema

Linhas diferentes podem ser armazenadas em lugares diferentes do disco

Recomendado para volumes gigantescos de dados

UserProfile

Bob

emailAddress

bob@example.com

1465676582

gender

male

1465676582

age

35

1465676582

Britney

emailAddress

brit@example.com

1465676432

gender

female

1465676432

Row →

Row Key

Column

Name

Value

Timestamp

Column

Name

Value

Timestamp

Column

Name

Value

Timestamp

Tori

emailAddress

tori@example.com

1435636158

country

Sweden

1435636158

hairColor

Blue

1465633654

Senac

Keyspace

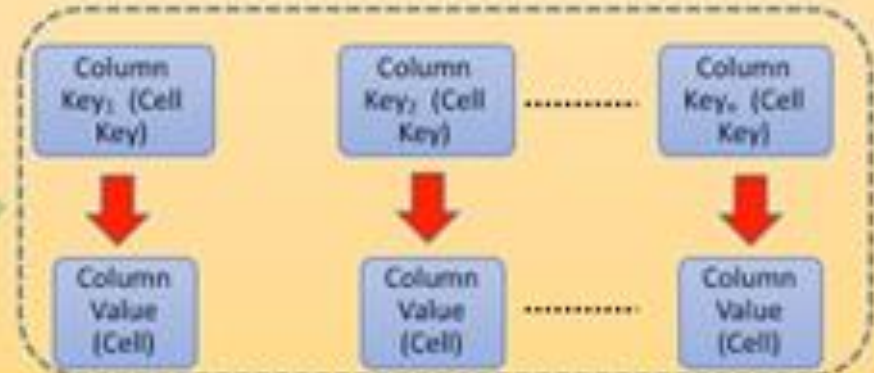
Settings

Column family

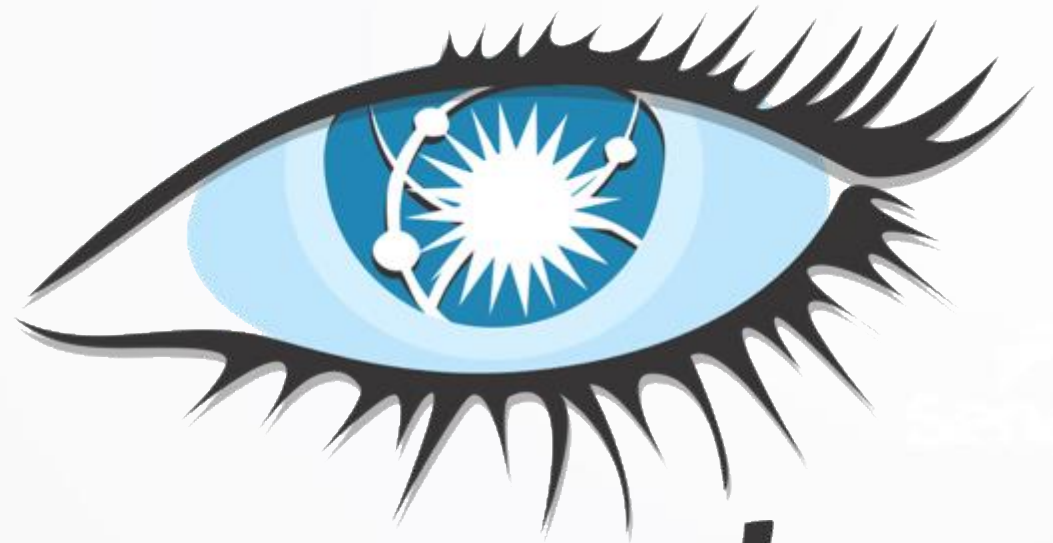
Settings

Row

Row Key



APACHE
HBASE



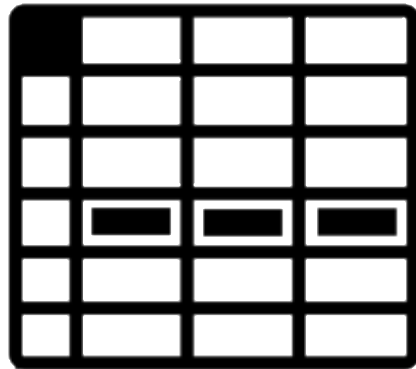
cassandra

Tipos de Bancos NoSQL

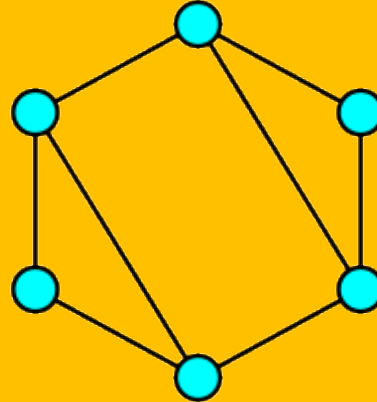
Documentos



Colunar



Grafos



**Chave -
Valor**



**Séries
Temporais**



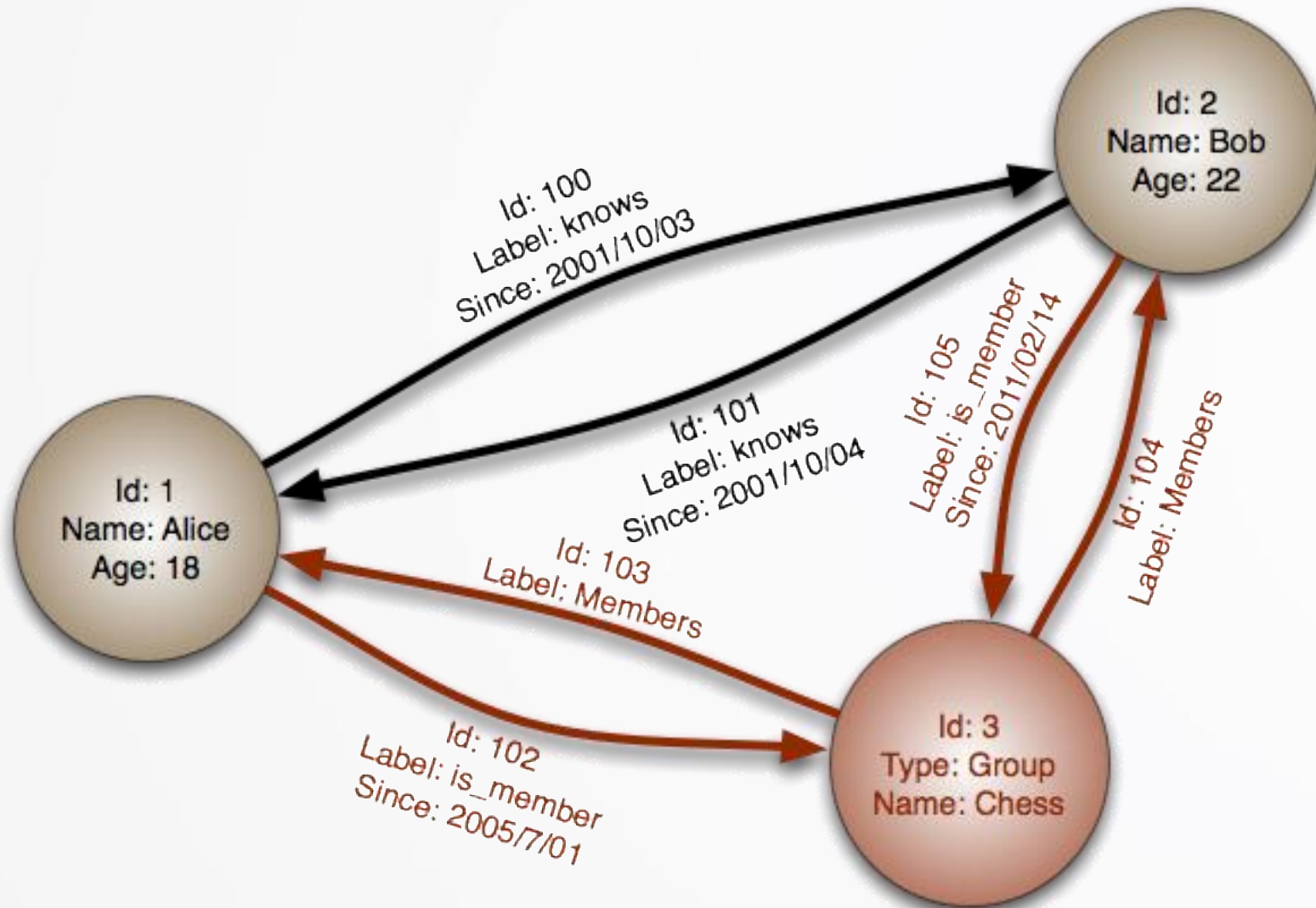
Orientado a Grafos

Conceito chave é grafo: nós e relacionamentos

Diferente de documentos, o ponto forte são os relacionamentos entre os dados

Pode encontrar relacionamentos em vários níveis e se pode ter múltiplos relacionamentos entre os nós

Senac



Senac



neo4j



ArangoDB



OrientDB®

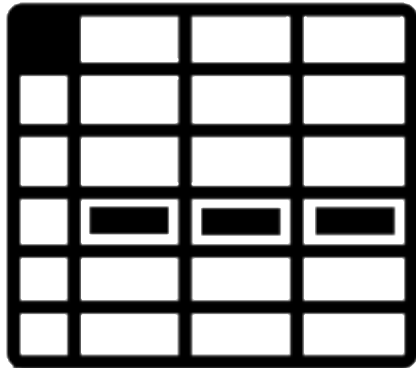
Senac

Tipos de Bancos NoSQL

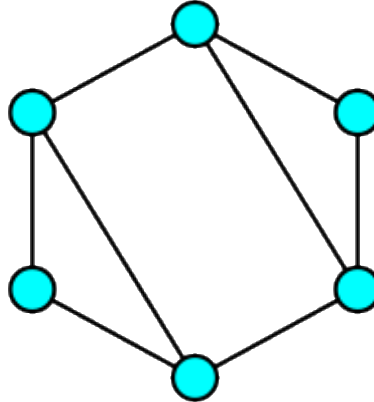
Documentos



Colunar



Grafos



**Chave -
Valor**



**Séries
Temporais**



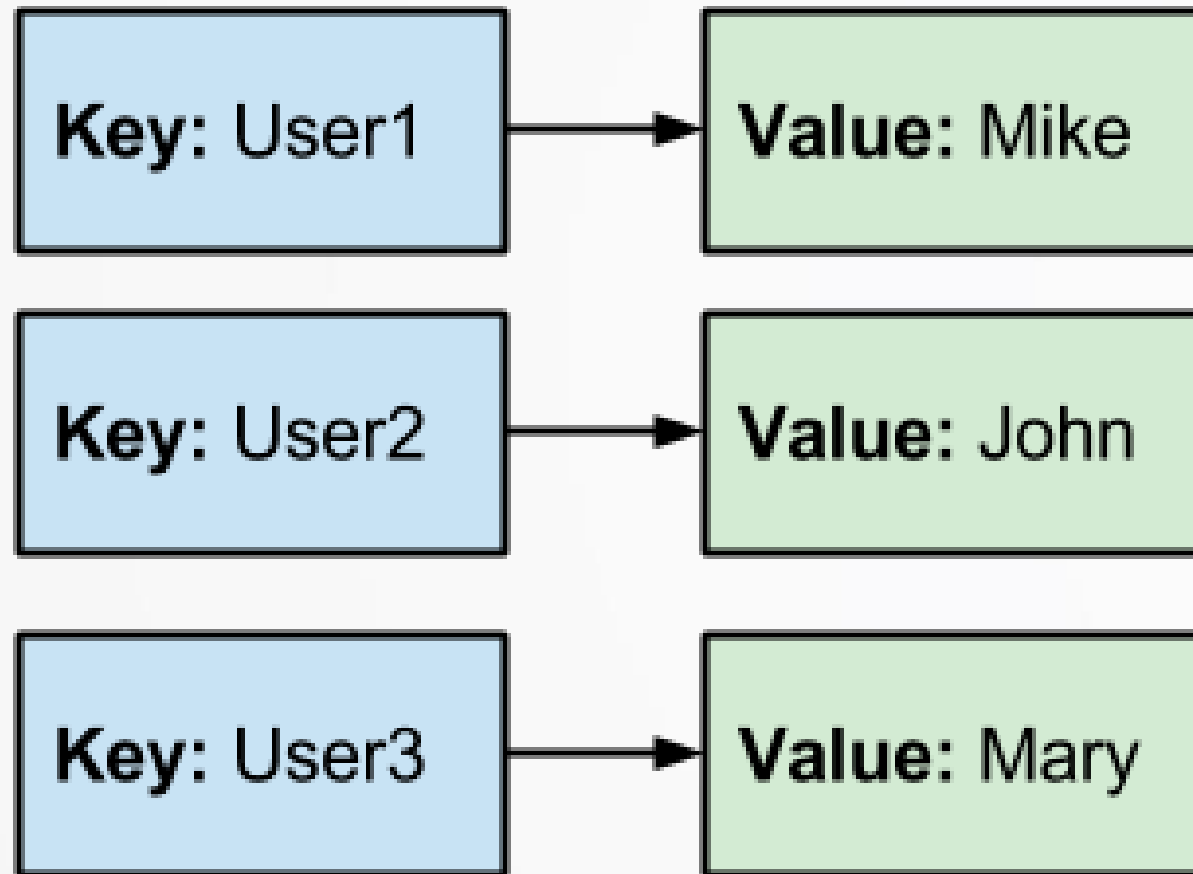
Chave-Valor

Formato mais simples: Chave única retorna um valor

Valor pode ter vários tipos de Dados diferentes: String, Inteiro, Hash, Lista...

Foco em performance – fácil fazer caching aos dados

Dificuldade para ações de atomicidade e transações



Senac



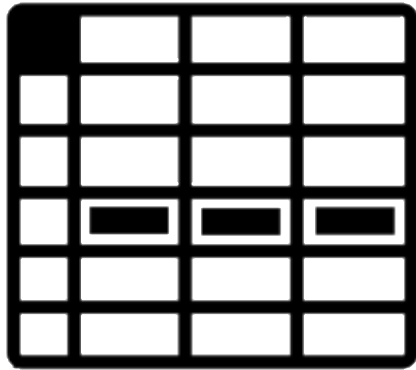
Senac

Tipos de Bancos NoSQL

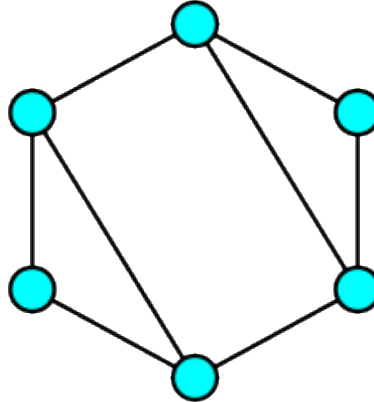
Documentos



Colunar



Grafos



**Chave -
Valor**



**Séries
Temporais**



Série Temporal

Tempo é o principal fator do dado

Permite uma rápida indexação e agrupamento por tempo

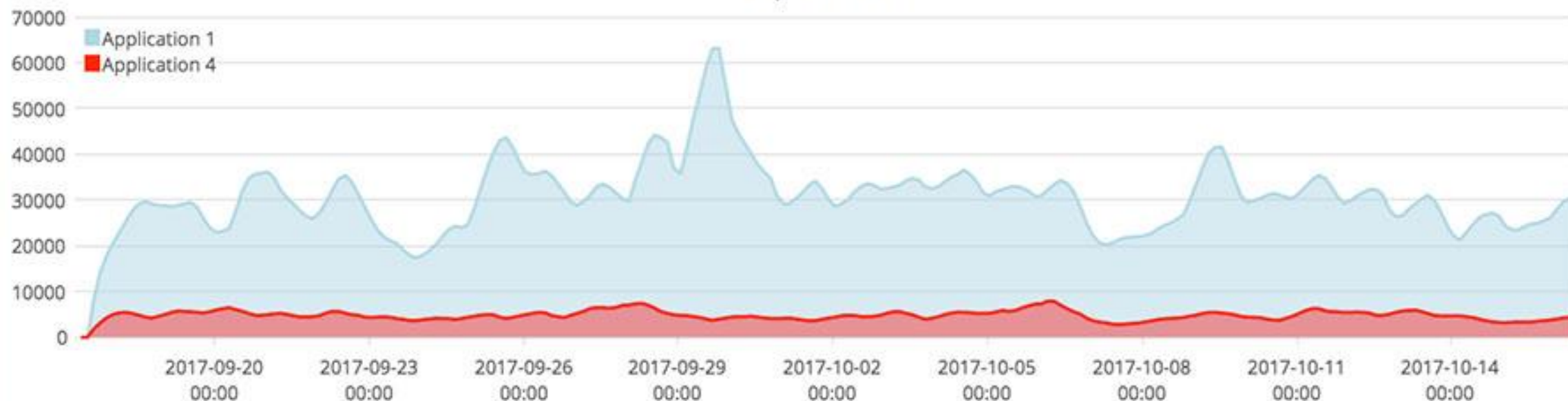
Têm operações padrão com data e agrupamentos

Permite a fácil construção de gráficos

Excelente para sensores e telemetria

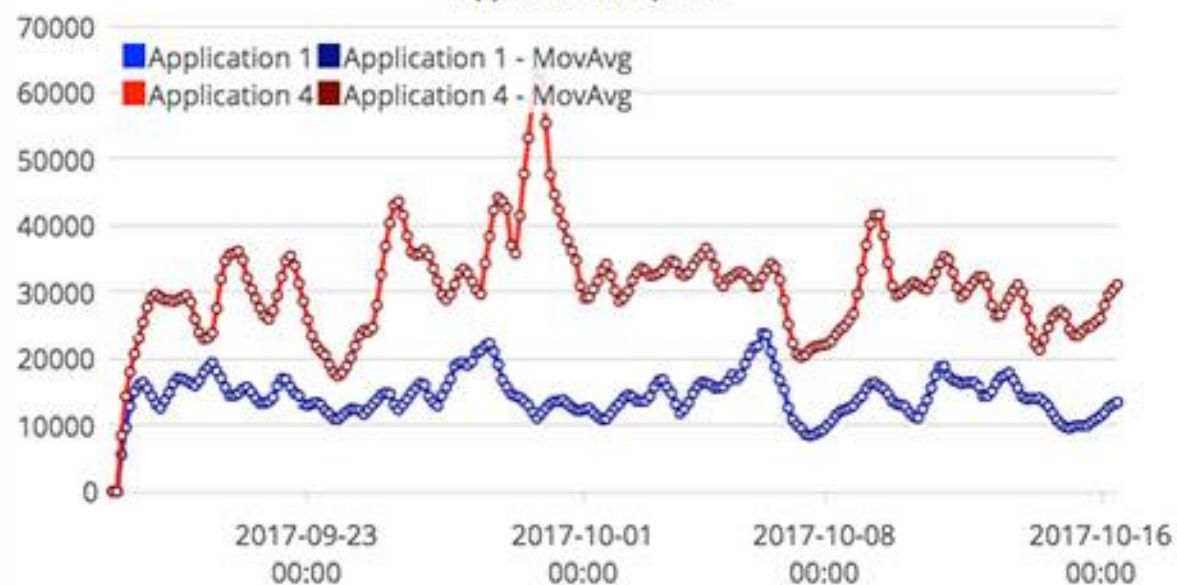
TS - Req v Time

Requests vs Time



TS - App Requests

Application Requests



TS - App4

TS - App1





elasticsearch



Assunto Extra: Teorema CAP

Teorema CAP

C - Consistência

Todos os clients têm a mesma visão sobre os dados ao mesmo tempo

A - Disponibilidade

Cada cliente pode ler e escrever dados em qualquer momento

P – Tolerância ao Particionamento

Sistema funciona apesar de falhas na comunicação entre os nós da rede



Teorema CAP: Apenas se consegue
fornecer 2 propriedades ao mesmo
tempo

A - Disponibilidade

C - Consistência

**P – Tolerância ao
Particionamento**

A - Disponibilidade

**Problemas de Rede
podem parar o sistema**

**Cliente pode ver um
dato inconsistente**

C - Consistência

**Risco do dato estar
indisponível**

**P – Tolerância ao
Particionamento**

A - Disponibilidade

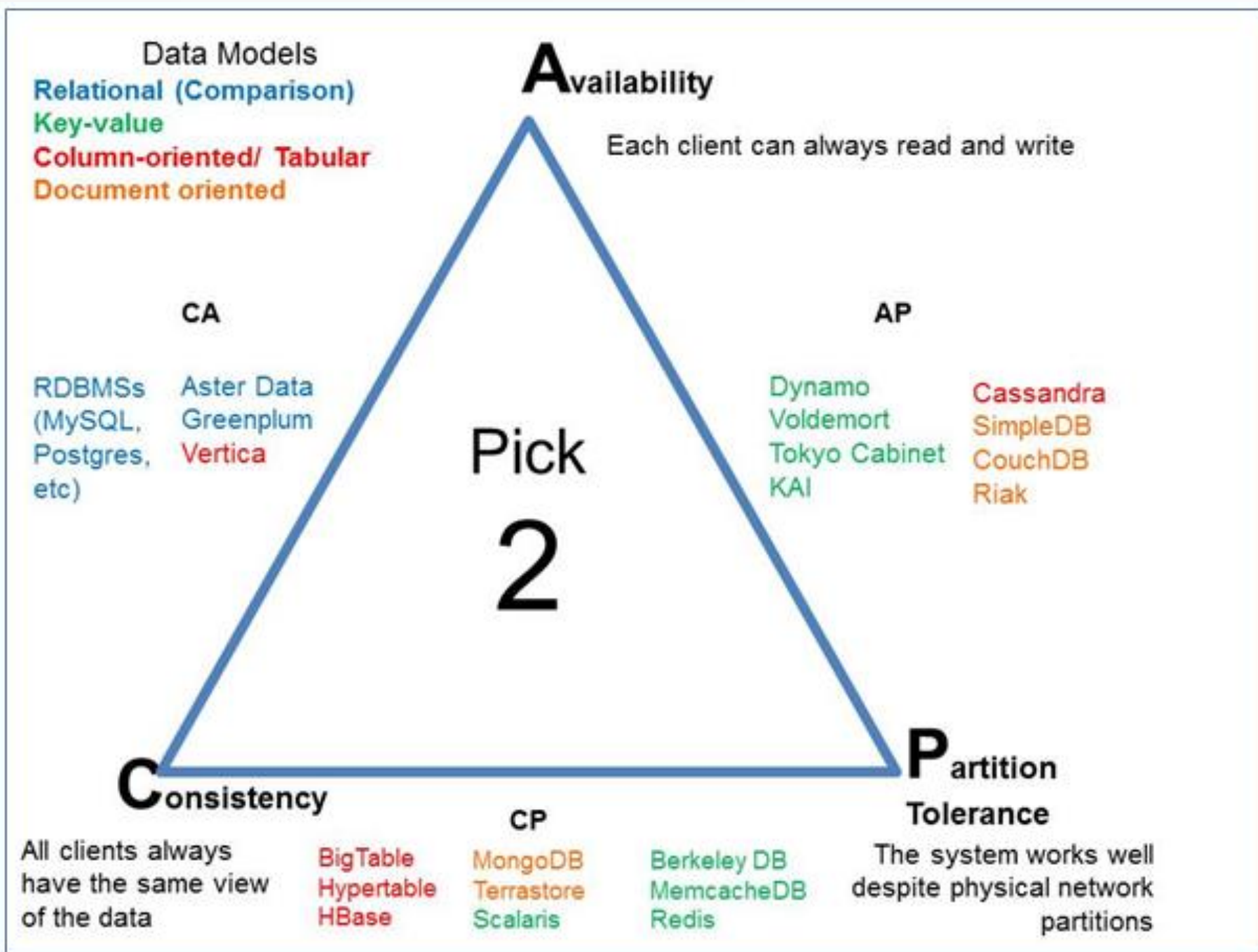
Bancos Relacionais

DNS, Cache

C - Consistência

Sistemas Financeiros

**P – Tolerância ao
Particionamento**





Muito Obrigado!

Diretoria de Pós-graduação e Pesquisa
Centro Universitário Senac

Prof. Celso Crivelaro
celsocrivelaro@gmail.com