

# MBA em IA e Big Data

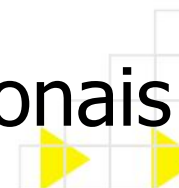


## Curso 01 - Linguagens e Ferramentas para Inteligência Artificial e Big Data (Python e SQL)

Introdução ao NoSQL

*Jose Fernando Rodrigues Junior*  
*ICMC-USP São Carlos*

Objetivo: apresentar conceitos sobre sistemas não relacionais



# Big Data: resumo em 6Vs



## Big Data

## Open Data

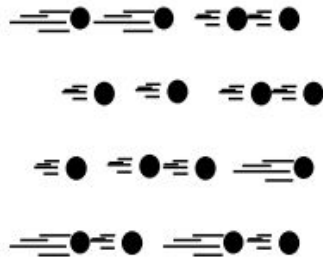
### Volume



#### Data at Rest

Terabytes to exabytes of existing data to process

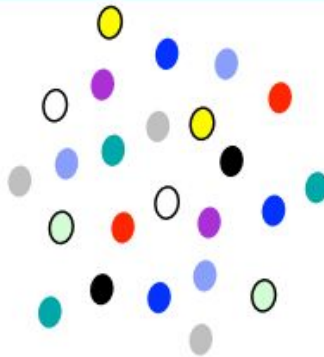
### Velocity



#### Data in Motion

Streaming data, milliseconds to seconds to respond

### Variety



#### Data in Many Forms

Structured, unstructured, text, multimedia

### Veracity



#### Data in Doubt

Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations

### Visibility



#### Data in the Open

Open data is generally open to anyone. Which raises issues of privacy. Security and provenance

### Value

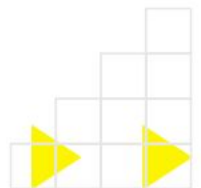


#### Data of Many Values

Large range of data values from free (data philanthropy to high value monetization)

# Motivação

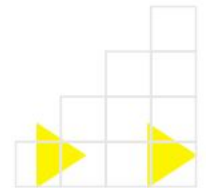
- ❑ Problemas com bases relacionais
  - ❑ Esquema rígido
  - ❑ Não é facilmente escalável
  - ❑ Depende de frequentes junções



# NoSQL



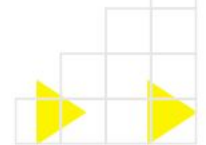
- **Not Only SQL**
- Classe de sistemas **não-relacionais** de armazenamento
- Usualmente, não requerem um **esquema** de dados, e/ou não priorizam operações de **junção**
- **Flexibilizam** uma ou mais das propriedades ACID
- São mais versáteis em termos de **escala horizontal**



# NoSQL

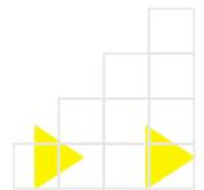
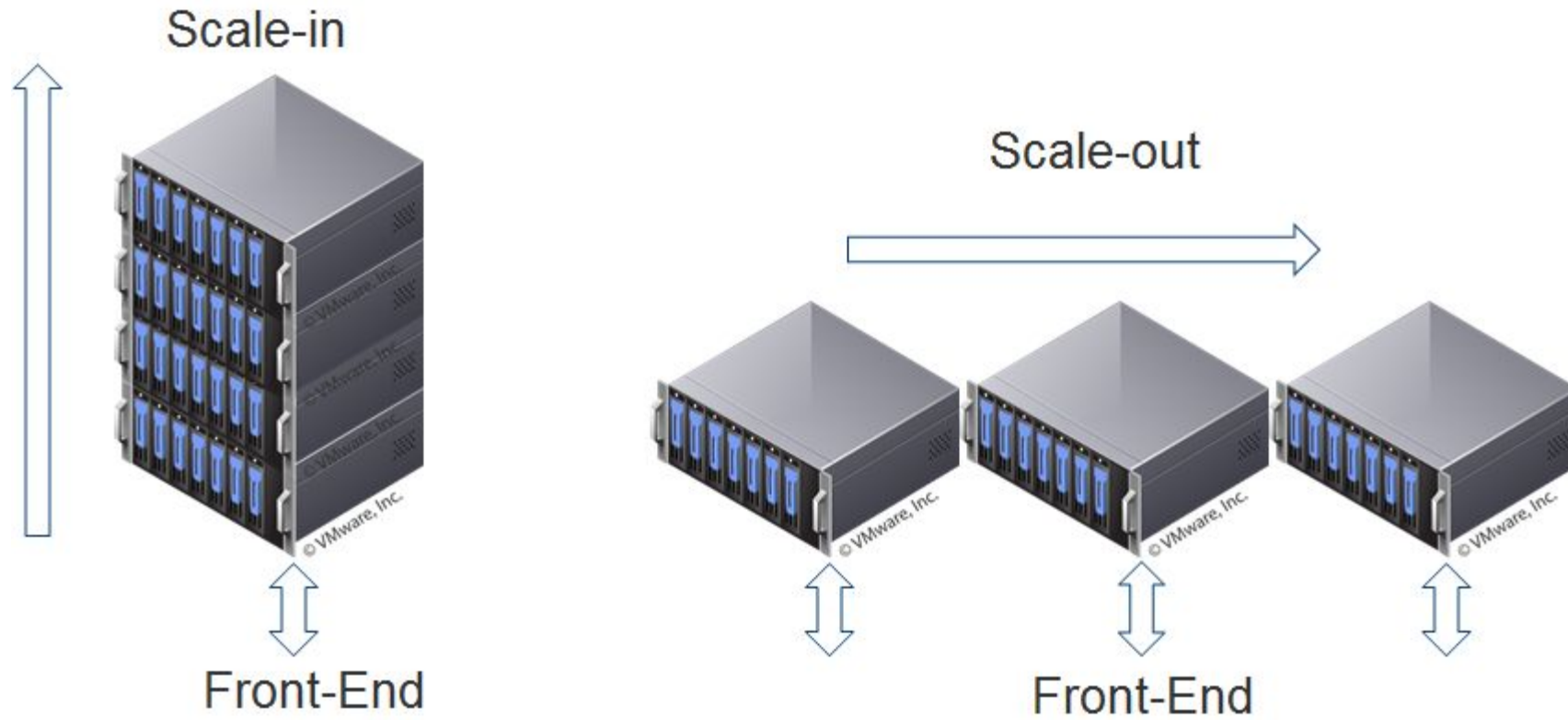


- São mais aderentes às **estruturas de dados usuais nas linguagens de programação** modernas
- Se **alicerçam nos recursos computacionais abundantes da atualidade**:
  - Por exemplo: um banco de dados com 2 GB em 1999 era grande, hoje cabe em **memória** → tendência que deve se intensificar ao longo das próximas décadas
  - Com **redes de dados** extremamente rápidas, podem enviar e receber grandes quantidades de dados em tempo razoável e escalar horizontalmente com facilidade





# NoSQL



# ACID x BASE

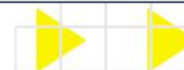


- **Basically Available:** **disponibilidade** em detrimento à consistência - maximização de leituras e escritas, sem a garantia de que estejam consistentes; alta disponibilidade por meio de replicação e distribuição em diversos nós;
  - **Soft State:** o **estado dos dados é “relaxado”**; por ex., a leitura de um dado não garante que ele se refere ao valor mais atualizado; ele pode mudar mesmo sem a ocorrência de leituras/escritas, pois os dados ainda podem estar em processamento - integridade é **responsabilidade do programador**;
  - **Eventual Consistency:** o sistema será eventualmente (em algum momento) consistente após a entrada/saída de dados; os dados serão replicados para nós diferentes e, **eventualmente, atingirão um estado consistente**; mas a consistência não é garantida no nível da transação.
- ⇒ **BASE** prevê outras propriedades pois, muitas vezes, as propriedades ACID não são necessárias, mas sempre são custosas;
- ⇒ **BASE** permite escalonamento horizontal facilitado, tolerância a falhas, e **alta disponibilidade** ao custo de menor consistência.

# ACID x BASE

BASE oferece menos garantias do que ACID, mas é mais escalável e se molda mais rapidamente a alterações no modelo de dados.

⇒ [Say Hello to NoSQL](#)





# Tipos de NoSQL

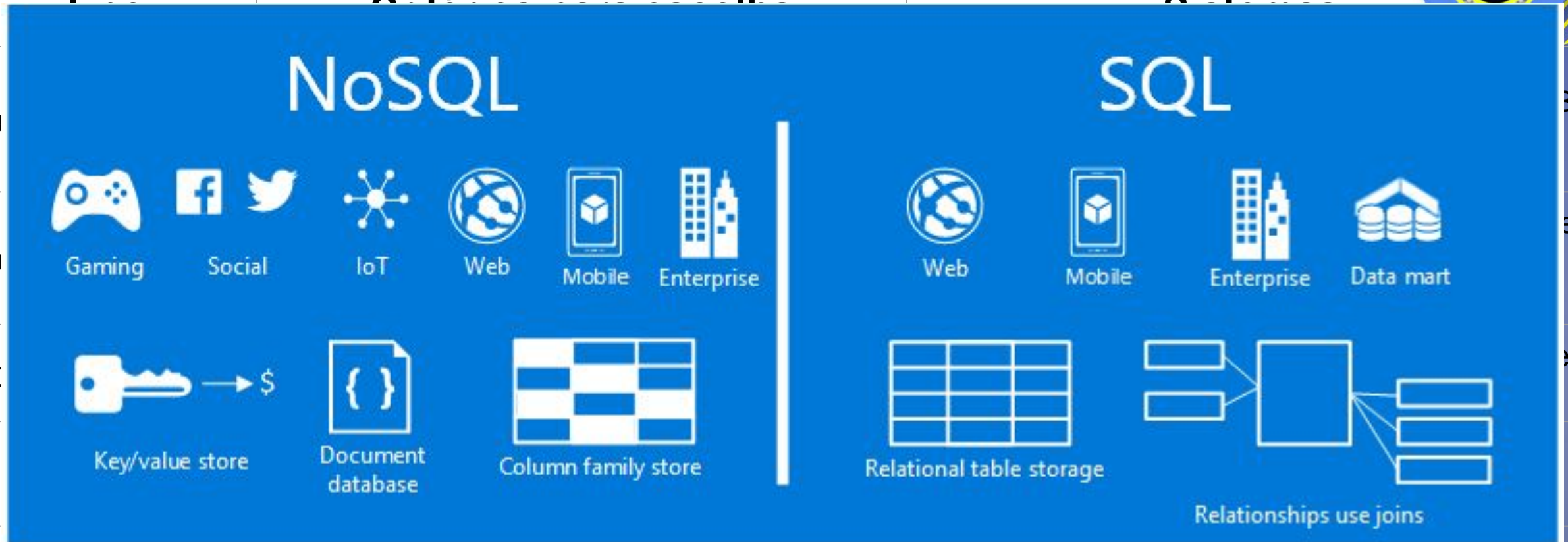


Tipo	Critérios para escolha	Sistemas
Key:Value Store	Esquema simples; poucos updates; consultas simples	Aerospike; Apache Cassandra; Berkeley DB; Redis; Riak
Document Store	Esquema flexível; consultas mais complexas; dados JSON/XML; indexação complexa (dados geoespaciais, texto)	MongoDB; Couchbase; CouchDB; Firebase; Google Firestone
Column-Oriented	Grande volume; escrita otimizada; recuperação de dados orientada a colunas	Cassandra; HBase; MS Azure Table Storage; Google Bigtable;
Graph	Operações de data traversal; consultas orientados aos relacionamentos	Neo4j; JanusGraph; TigerGraph; Giraph
Multi-modelo		Amazon DynamoDB; MS Azure Cosmos; Oracle NoSQL; Couchbase; ArangoDB; Virtuoso;

⇒ [Como escolher um Sistema NoSQL](#)



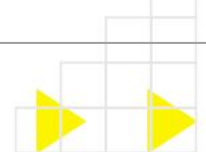
# Tipos de NoSQL



Multi-modelo

Amazon DynamoDB; MS Azure Cosmos; Oracle NoSQL; Couchbase; ArangoDB; Virtuoso;

⇒ [Como escolher um Sistema NoSQL](#)

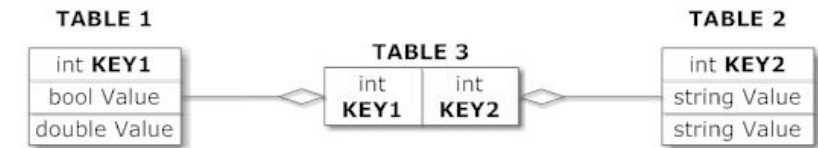


# MongoDB

- mongoDB = “Humongous DB”
  - Open-source
  - Document-based
  - “High performance, high availability”
  - Escalabilidade facilitada



## Relational Model



## Document Model

Collection ("Things")

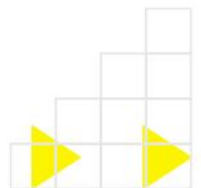


# MongoDB

- mongoDB = “Humongous DB”
  - Open-source
  - Document-based
  - “High performance, high availability”
  - Escalabilidade facilitada

⇒ Atualmente, a solução NoSQL mais usada no mercado

<https://db-engines.com/en/ranking>



**Mongodb: The Definitive Guide: Powerful and Scalable Data Storag,**

**By Kristina Chodorow and Mike Dirolf**

**Published: 9/24/2019**

**Pages: 514**

**Language: English**

**Publisher: O'Reilly Media, CA**

