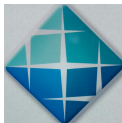


# Tópicos de Big Data em Python

## Aula 03

Evandro J.R. Silva<sup>1</sup>

<sup>1</sup> Bacharelado em Ciência da Computação  
Estácio Teresina



# Sumário

1 Introdução

2 Computação em Nuvem

3 Streaming

4 FIM

# Introdução

# Introdução

- Tema: Princípios de Big Data
- Tópicos:
  - 1.3 Plataformas em Nuvem para aplicações de Big Data
  - 1.4 Processamento e Streaming de Dados
- Objetivos
  - Apresentar os conceitos da computação em nuvem e de streaming de dados, descrevendo as principais plataformas em nuvem, para aplicações de Big Data e processamento de streaming de dados.

# Computação em Nuvem

# Computação em Nuvem

- "O paradigma **Computação em Nuvem** se refere ao uso, por meio da internet, de memória e de processamento distribuídos em servidores."
- A arquitetura distribuída fica *transparente* para o usuário final, ou seja, há uma redução de complexidade.
- Outra redução é a do custo, pois o usuário não precisa mais arcar com as despesas de manutenção do hardware.

# Computação em Nuvem

- A computação em nuvem viabilizou a execução de tarefas que exigem muito processamento.
- Por causa disso, acaba se apresentando como uma solução bastante adequada para o Big Data, seja com uma alocação maior de máquinas físicas ou virtuais de acordo com a demanda, seja com a capacidade de processamento paralelo.
- Porém, ainda existem muitos problemas como gargalo na transferência de dados, flutuação de desempenho dos servidores, etc.

# Computação em Nuvem

## ■ Principais Características

### ■ Modelo de Negócio

- *Software as Service* (SaaS): o software é disponibilizado para o uso livre ou comercial como um serviço sob demanda. Exemplos: Google Workspace, Dropbox.
- *Platform as Service* (PaaS): entrega de uma plataforma de programação como um serviço. Exemplos: AWS Elastic Beanstalk, Windows Azure.
- *Infrastructure as Service* (IaaS): o provedor oferece uma infraestrutura computacional ao usuário final. Exemplos: Amazon Web Services (AWS), Microsoft Azure.



# Computação em Nuvem

## ■ Principais Características

### ■ **Modelo de Negócio**

### ■ **Privacidade**

- *Nuvens Privadas.*
- *Nuvens Públicas.*
- *Nuvens Mistas.*

# Computação em Nuvem

## ■ Principais Características

### ■ **Modelo de Negócio**

### ■ **Privacidade**

### ■ **Precificação**

- *Gratuito.*
- *Pague-pelo-uso ou pay-per-use.*

# Computação em Nuvem

- Principais Características
  - **Características Arquiteturais**
    - *Heterogeneidade* [de hardware].
    - *Virtualização*.
    - *Elasticidade*.
    - *Disponibilidade*.

# Computação em Nuvem

- Principais Características
  - **Características Arquiteturais**
  
- **Métodos de Acesso**
  - *Navegadores.*
  - *Clientes Leves.*
  - *Clientes Móveis.*
  - *APIs*

# Computação em Nuvem

- Principais Características

- **Características Arquiteturais**

- **Métodos de Acesso**

- **Padrões:** padrões [de segurança, de dados, etc.] utilizados na transferência de dados

# Computação em Nuvem

## ■ Principais Características

### ■ Características Arquiteturais

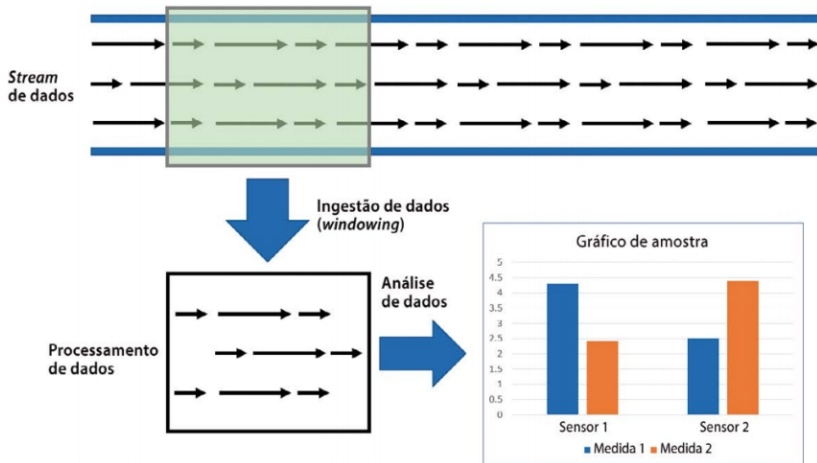
### ■ Métodos de Acesso

- **Padrões:** padrões [de segurança, de dados, etc.] utilizados na transferência de dados
- **Orientação:** pode ser centrada em tarefa ou centrada no usuário. Muda conforme o tipo de serviço disponibilizado.

# Streaming

# Streaming

- “É uma evolução do big data pela forma de prospectar e analisar os dados dinamicamente e de modo contínuo”.
- Metáfora comumente utilizada para *streaming*: **rio**.





# Streaming

- Nesse fluxo de dados contínuo, o processamento se dá pela captura de um trecho, conhecido como janela.
- Uma vez coletados, esses dados são processados sequencialmente e de forma incremental, registro por registro ou em uma janela que desliza com base no tempo.
- Esse processo de captura de dados na janela é conhecido como ingestão de dados, o que possibilita o processamento e a realização da análise desses dados.

# Streaming

- São diversos os profissionais que estão envolvidos nas questões de controle de fluxos de dados e monitoramento de seu uso.

# Streaming

- São diversos os profissionais que estão envolvidos nas questões de controle de fluxos de dados e monitoramento de seu uso.
  - **Administradores de rede:** examinam o fluxo de dados enquanto entram e saem dessas redes, bem como trabalham em melhorar as topologias existentes ou na criação de novas.

# Streaming

- São diversos os profissionais que estão envolvidos nas questões de controle de fluxos de dados e monitoramento de seu uso.
  - **Administradores de rede:** examinam o fluxo de dados enquanto entram e saem dessas redes, bem como trabalham em melhorar as topologias existentes ou na criação de novas.
  - **Profissionais envolvidos com tarefas de TI:** examinam como a rede global trata questões de transporte ao longo da transmissão.

# Streaming

- Tradicionalmente os dados são transferidos em lotes e armazenados, então algum algoritmo de Big Data é aplicado sobre eles.
- Este não é o caso do *streaming*, onde os dados são atualizados constantemente. Há uma diferença entre [por exemplo] levantar os dados dos últimos 3 meses e ser atualizado todo dia (isso apresenta algumas dificuldades ou desafios)

# Streaming

- Tradicionalmente os dados são transferidos em lotes e armazenados, então algum algoritmo de Big Data é aplicado sobre eles.
- Este não é o caso do *streaming*, onde os dados são atualizados constantemente. Há uma diferença entre [por exemplo] levantar os dados dos últimos 3 meses e ser atualizado todo dia (isso apresenta algumas dificuldades ou desafios)
  - O sistema não tem controle da ordem de chegada dos dados, nem do tamanho do fluxo.
  - A estratégia de armazenamento influencia no processamento, pois alguns dados podem ter sido armazenados somente temporariamente, ou descartados logo após seu processamento.

# Streaming

## ■ Fontes de streaming

- São variadas as fontes. As mais comuns são os dados coletados a todo momento enquanto usamos computadores, smartphones, acessamos algum site, fazemos alguma compra, utilizamos ou atualizamos nossas redes sociais, etc.

# Streaming

## ■ Fontes de streaming

- São variadas as fontes. As mais comuns são os dados coletados a todo momento enquanto usamos computadores, smartphones, acessamos algum site, fazemos alguma compra, utilizamos ou atualizamos nossas redes sociais, etc.
- Além disso, existem os dados que são coletados de sistemas, sejam de programas ou servidores, sejam de infraestruturas (rede elétrica, por exemplo).



# Streaming

## ■ Fontes de streaming

- São variadas as fontes. As mais comuns são os dados coletados a todo momento enquanto usamos computadores, smartphones, acessamos algum site, fazemos alguma compra, utilizamos ou atualizamos nossas redes sociais, etc.
- Além disso, existem os dados que são coletados de sistemas, sejam de programas ou servidores, sejam de infraestruturas (rede elétrica, por exemplo).
- Existem ainda os monitoramentos constantes de câmeras de vigilância (quantas câmeras existem só em São Paulo?), rastreamento de voo, e rastreamento em tempo real de condições atmosféricas

Terminamos por hoje!

Aula baseada no livro:

GOLDSCHMIDT, Ronaldo. **Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações**. Capítulo 8. Rio de Janeiro: Elsevier, 2015.

PEREIRA, Mariana Araújo; NEUMANN, Fabiano Berlink; MILANI, Alessandra M. Paz; et al. **Framework de Big Data**. Capítulo 7. Porto Alegre: SAGAH, 2019.