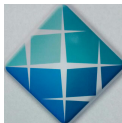


Tópicos de Big Data em Python

Aula 02

Evandro J.R. Silva¹

¹Bacharelado em Ciência da Computação
Estácio Teresina



Sumário

- 1 Big Data e IoT
 - Tecnologias para Big Data
 - Internet das Coisas
 - Arquitetura da IoT
- 2 Integrando Big Data e IoT
- 3 FIM

Big Data e IoT

Big Data e IoT

- Relembrando um pouco da aula passada:

Big Data e IoT

- Relembrando um pouco da aula passada:
 - **Big Data** "é a descoberta de informação baseada nos dados de instituições e empresas, o que pode revelar outros novos fatores".
 - Com a análise dos dados as empresas podem conhecer melhor seus clientes, introduzir novos produtos e serviços, gerenciar melhor os riscos na tomada de decisão, e até conseguir reduzir custos.
 - Essa análise é feita sobre um **grande volume de dados**.

Big Data e IoT

- Três tipos principais de análise de dados
 - **Análise exploratória:**
 - **Análise preditiva:**
 - **Análise de agrupamento:**

Big Data e IoT

- Três tipos principais de análise de dados
 - **Análise exploratória:** permite a compreensão de como os dados estão distribuídos e qual o formato em que se encontram. A informação adquirida nas análises exploratórias serve de apoio na tomada de decisão sobre o tipo de tarefa de mineração dos dados e na escolha do algoritmo.
 - **Análise preditiva:**
 - **Análise de agrupamento:**

Big Data e IoT

- Três tipos principais de análise de dados
 - **Análise exploratória:**
 - **Análise preditiva:** prevê resultados futuros com base em dados do passado. Pode ser entendida como um processo que permite descobrir o relacionamento existente entre os exemplares de um conjunto de dados, descritos por uma série de características e rótulos atribuídos a esses dados.
 - **Análise de agrupamento:**

Big Data e IoT

- Três tipos principais de análise de dados

- **Análise exploratória:**

- **Análise preditiva:**

- **Análise de agrupamento:** consiste no agrupamento de objetos de acordo com suas características. Também é conhecida como *clustering* e permite, com base em um conjunto de características, a descoberta de relações existentes entre exemplares de conjuntos de dados.

Big Data e IoT

Tecnologias para Big Data

Tecnologias para Big Data

- As tecnologias para Big Data podem ser organizadas em duas categorias: (1) ferramentas de análise (*analytics*) e (2) tecnologias de infraestrutura, que servem para processar e armazenar os grandes volumes de dados.
- Vejamos algumas delas:

Tecnologias para Big Data

- As tecnologias para Big Data podem ser organizadas em duas categorias: (1) ferramentas de análise (*analytics*) e (2) tecnologias de infraestrutura, que servem para processar e armazenar os grandes volumes de dados.
- Vejamos algumas delas:
 - **Hadoop:** plataforma de *software* em Java de computação distribuída voltada para *clusters* e processamento de grandes volumes de dados, com suporte a tolerância a falhas.

Tecnologias para Big Data

- As tecnologias para Big Data podem ser organizadas em duas categorias: (1) ferramentas de análise (*analytics*) e (2) tecnologias de infraestrutura, que servem para processar e armazenar os grandes volumes de dados.
- Vejamos algumas delas:
 - **Hadoop:** plataforma de *software* em Java de computação distribuída voltada para *clusters* e processamento de grandes volumes de dados, com suporte a tolerância a falhas.
 - **Map Reduce:** é um *framework* desenvolvido pela Google para suportar computações paralelas em grandes coleções de dados em *clusters* de computadores.

Tecnologias para Big Data

- As tecnologias para Big Data podem ser organizadas em duas categorias: (1) ferramentas de análise (*analytics*) e (2) tecnologias de infraestrutura, que servem para processar e armazenar os grandes volumes de dados.
- Vejamos algumas delas:
 - **Hadoop:** plataforma de *software* em Java de computação distribuída voltada para *clusters* e processamento de grandes volumes de dados, com suporte a tolerância a falhas.
 - **Map Reduce:** é um *framework* desenvolvido pela Google para suportar computações paralelas em grandes coleções de dados em *clusters* de computadores.
 - **Linguagens de script:** linguagens de programação adequadas a Big Data, como o Python.

Tecnologias para Big Data

- As tecnologias para Big Data podem ser organizadas em duas categorias: (1) ferramentas de análise (*analytics*) e (2) tecnologias de infraestrutura, que servem para processar e armazenar os grandes volumes de dados.
- Vejamos algumas delas:
 - **Hadoop:** plataforma de *software* em Java de computação distribuída voltada para *clusters* e processamento de grandes volumes de dados, com suporte a tolerância a falhas.
 - **Map Reduce:** é um *framework* desenvolvido pela Google para suportar computações paralelas em grandes coleções de dados em *clusters* de computadores.
 - **Linguagens de script:** linguagens de programação adequadas a Big Data, como o Python.
 - **Visual Analytics:** método de análise em grandes volumes de dados com saída em formato visual ou gráfico.

Tecnologias para Big Data

- As tecnologias para Big Data podem ser organizadas em duas categorias: (1) ferramentas de análise (*analytics*) e (2) tecnologias de infraestrutura, que servem para processar e armazenar os grandes volumes de dados.
- Vejamos algumas delas:
 - **Hadoop:** plataforma de *software* em Java de computação distribuída voltada para *clusters* e processamento de grandes volumes de dados, com suporte a tolerância a falhas.
 - **Map Reduce:** é um *framework* desenvolvido pela Google para suportar computações paralelas em grandes coleções de dados em *clusters* de computadores.
 - **Linguagens de script:** linguagens de programação adequadas a Big Data, como o Python.
 - **Visual Analytics:** método de análise em grandes volumes de dados com saída em formato visual ou gráfico.
 - **Processamento de Linguagem Natural (PLN):** área/conceito da Inteligência Artificial que permite a análise de textos.

Tecnologias para Big Data

- As tecnologias para Big Data podem ser organizadas em duas categorias: (1) ferramentas de análise (*analytics*) e (2) tecnologias de infraestrutura, que servem para processar e armazenar os grandes volumes de dados.
- Vejamos algumas delas:
 - **Hadoop:** plataforma de *software* em Java de computação distribuída voltada para *clusters* e processamento de grandes volumes de dados, com suporte a tolerância a falhas.
 - **Map Reduce:** é um *framework* desenvolvido pela Google para suportar computações paralelas em grandes coleções de dados em *clusters* de computadores.
 - **Linguagens de script:** linguagens de programação adequadas a Big Data, como o Python.
 - **Visual Analytics:** método de análise em grandes volumes de dados com saída em formato visual ou gráfico.
 - **Processamento de Linguagem Natural (PLN):** área/conceito da Inteligência Artificial que permite a análise de textos.
 - **In-memory analytics:** processamento de Big Data realizado na memória do computador com o objetivo de aumentar a velocidade das análises.

Big Data e IoT

IoT - Internet das Coisas

IoT

- IoT (*Internet of Things*) — comumente traduzido como Internet das Coisas — é utilizado para transmitir a ideia de que a Internet pode estar presente em todas as coisas.
- Seu conceito é baseado na ideia de fusão do mundo real com o mundo digital, fazendo com que indivíduos estejam em constante comunicação e interação com outras pessoas e objetos.

Arquitetura da IoT

- Ainda não existe uma arquitetura universal.
- A integração das aplicações de IoT ainda é um grande desafio (como garantir a interoperabilidade de todas as coisas?).
- Diversas empresas se uniram para formar uma arquitetura genérica para IoT, a *oneM2M* (*one machine-to-machine*), cujo principal objetivo é desenvolver especificações para garantir a funcionalidade global dos sistemas.
- Sua arquitetura básica possui três tipos de camadas:
 - **camada de aplicação:** engloba os elementos de ponta, como sensores, geolocalizadores, smartphones, etc.
 - **camada de serviços comuns:** conjunto de funções comuns utilizando uma arquitetura funcional. Pode ser acessada pela camada de aplicação via API.
 - **camada de serviços de rede.**

Integrando Big Data e IoT

Integrando Big Data e IoT

- A relação entre IoT e Big Data pode ser dividida em três etapas para permitir o gerenciamento de dados da IoT:

Integrando Big Data e IoT

- A relação entre IoT e Big Data pode ser dividida em três etapas para permitir o gerenciamento de dados da IoT:
 - 1 Consiste em gerenciar fontes de dados da IoT nas quais dispositivos sensores conectados usam aplicativos para interagir uns com os outros. Por exemplo, a interação de dispositivos como câmeras de vigilância, semáforos inteligentes e dispositivos domésticos inteligentes gera grandes quantidades de fontes de dados com diferentes formatos.

Integrando Big Data e IoT

- A relação entre IoT e Big Data pode ser dividida em três etapas para permitir o gerenciamento de dados da IoT:
 - 2 Os dados gerados são chamados de Big Data, que são baseados em seu volume, velocidade e variedade. Essas enormes quantidades de dados são armazenadas em arquivos de Big Data em bancos de dados distribuídos, compartilhados e tolerantes a falhas.

Integrando Big Data e IoT

- A relação entre IoT e Big Data pode ser dividida em três etapas para permitir o gerenciamento de dados da IoT:
- 3 Aplicação de ferramentas de análise, as quais podem analisar os grandes conjuntos de dados de IoT armazenados.

Integrando Big Data e IoT

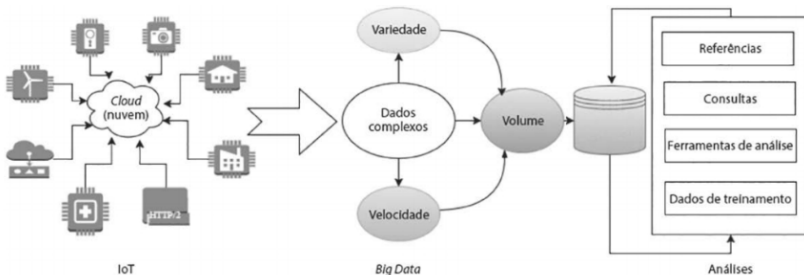


Figura 1. Relação entre IoT e análise de *Big Data*.

Fonte: Adaptada de Marjani et al. (2017).

Integrando Big Data e IoT

Crescimento da Internet das Coisas — IoT

O número de equipamentos conectados vão exceder 50 bilhões

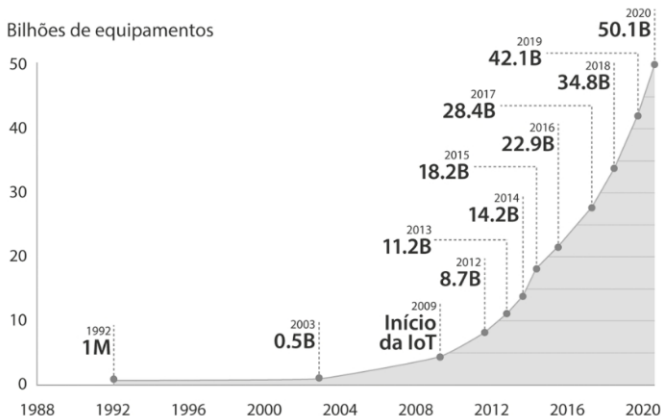


Figura 2. Estimativa, projeção e crescimento.

Fonte: Adaptada de Santos (2018).

Integrando Big Data e IoT

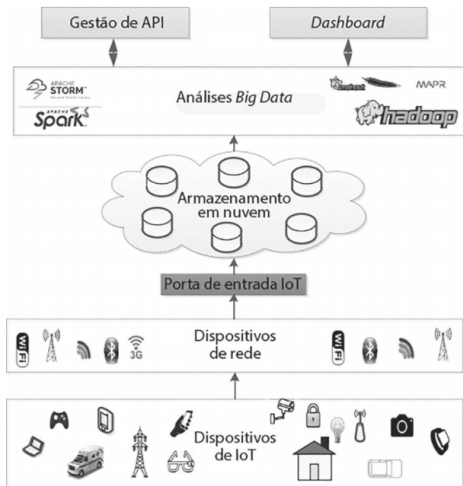


Figura 3. Arquitetura de IoT e análise de *Big Data*.

Fonte: Adaptada de Marjani et al. (2017).

Terminamos por hoje!

Aula baseada no livro:

MORAIS, Izabelly. **Introdução a Big Data e Internet das Coisas (IOT)**. Porto Alegre: SAGAH, 2018. Capítulos 1 e 2.