

BIG DATA

Rodrigo Macedo - <http://www.itnerante.com.br/profile/RodrigoMacedo>

Escopo do Curso

- Conceituação Geral.
- Pilares do Big Data.
- Questões de Concursos.
- Dados Estruturados x Não Estruturados.
- Questões de Concursos.
- Armazenamento e Processamento Distribuído.
- Questões de Concurso.
- Analytics e Machine Learning.
- Questões de Concurso.





Definições de Big Data

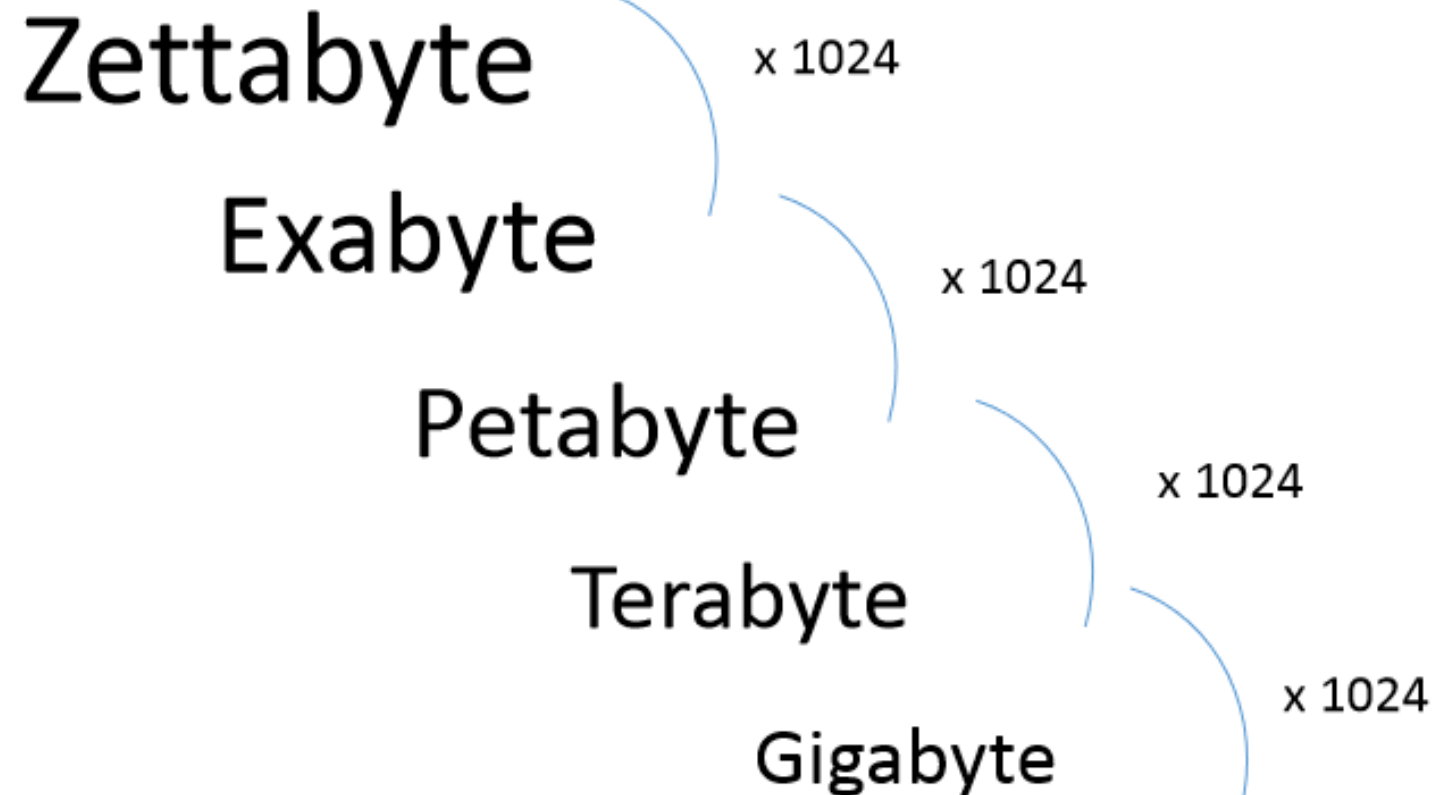
"Big Data é a capacidade de uma empresa para armazenar, processar e acessar todos os dados de que necessita para funcionar, tomar decisões, reduzir riscos, e servir os clientes." - Forrester

"Big Data, em geral, é definido como alto volume, velocidade e variedade ativos de informação que exigem formas eficazes e inovadoras de custos de processamento de informações para uma melhor percepção e tomada de decisão." - Gartner

"Big data são os dados caracterizados por três atributos: volume, variedade e velocidade." - IBM

"Big data são os dados caracterizados por quatro atributos principais: volume, variedade, velocidade e valor." - Oracle

Tamanho do Big Data





Movimento sem volta



Indivíduos com smartphones e sites de redes sociais continuarão a impulsionar o crescimento de dados exponencialmente

Os pilares do Big Data



Volume



- ▶ Sem dúvidas, o maior pilar do Big Data.
- ▶ É necessário ter grande conjunto de dados, pois uma empresa nunca vai analisar um dado isolado. Precisa-se dos dados do setor de vendas, recrutamento, fornecimento, etc.

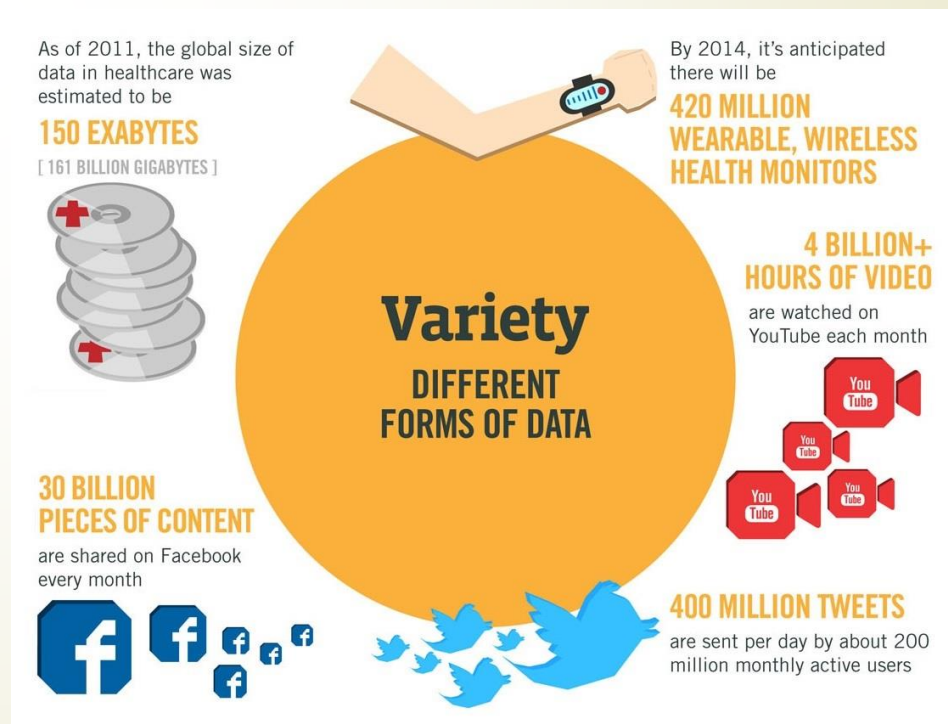
Velocidade

- ▶ Tem a ver com a qualidade temporal em que esses dados são acessados.
- ▶ Quando possível, solicitar a análise desses dados em tempo real.
- ▶ Em outros casos, minimizar a latência da informação.
- ▶ 1 terabyte de informação é criada durante uma única sessão da bolsa de valores Americana, a New York Stock Exchange (NYSE);



Variedade

- Os dados que alimentam um Big Data, podem ser desde dados textuais, até análises de sentimentos, imagens, sons, vídeos, etc.



Veracidade

- Acima de tudo, as informações que forem extraídas das análises de dados, devem ser confiáveis e íntegras.
- Não adianta nada, prover de uma grande quantidade de armazenamento para diferentes tipos de dados que poderão ser acessados o mais rápido possível, se as informações que forem geradas, apresentarem uma alta taxa de erros.





Q1- [ESAF ANAC 2016]


Big Data é:

- A) volume + variedade + agilidade + efetividade, tudo agregando + valor + atualidade.
- B) volume + oportunidade + segurança + veracidade, tudo agregando + valor.
- C) dimensão + variedade + otimização + veracidade, tudo agregando + agilidade.
- D) volume + variedade + velocidade + veracidade, tudo agregando + valor.
- E) volume + disponibilidade + velocidade + portabilidade, tudo requerendo - valor.

Q1- [ESAF ANAC 2016]

Big Data é:


- A) volume + variedade + agilidade + efetividade, tudo agregando + valor + atualidade.
- B) volume + oportunidade + segurança + veracidade, tudo agregando + valor.
- C) dimensão + variedade + otimização + veracidade, tudo agregando + agilidade.
- D) volume + variedade + velocidade + veracidade, tudo agregando + valor.**
- E) volume + disponibilidade + velocidade + portabilidade, tudo requerendo - valor.



Q2- [CESGRANRIO PETROBRAS 2018] A principal definição de Big Data parte de três características, conhecidas como 3 V do Big Data, a saber: velocidade, variedade e volume.

O termo velocidade refere-se, principalmente, à

- A) necessidade das aplicações de gerar respostas rapidamente, a partir de grandes massas de dados.
- B) existência de um alto fluxo de dados na entrada.
- C) necessidade de gerar aplicações rapidamente, em função da demanda do negócio.
- D) importância da facilidade de manipular cubos de visualização de dados, rapidamente.
- E) rapidez com que os dados se tornam inválidos com o tempo.



Q2- [CESGRANRIO PETROBRAS 2018] A principal definição de Big Data parte de três características, conhecidas como 3 V do Big Data, a saber: velocidade, variedade e volume.

O termo velocidade refere-se, principalmente, à


A) necessidade das aplicações de gerar respostas rapidamente, a partir de grandes massas de dados.

B) existência de um alto fluxo de dados na entrada.

C) necessidade de gerar aplicações rapidamente, em função da demanda do negócio.

D) importância da facilidade de manipular cubos de visualização de dados, rapidamente.

E) rapidez com que os dados se tornam inválidos com o tempo.




Q3- [FCC SEFAZ-SC 2018] No âmbito da ciência de dados na definição de *Big Data*, utilizam-se características ou atributos que alguns pesquisadores adotam como sendo os cinco Vs. Porém, a base necessária para o reconhecimento de *Big Data* é formada por três propriedades:

- A) valor, velocidade e volume.
- B) valor, veracidade e volume.
- C) variedade, velocidade e volume.
- D) variedade, valor e volume.
- E) velocidade, veracidade e volume.

Q4 – [CESPE PF 2018] Julgue o item que segue, relativo a noções de mineração de dados, *big data* e aprendizado de máquina.

Big data refere-se a uma nova geração de tecnologias e arquiteturas projetadas para processar volumes muito grandes e com grande variedade de dados, permitindo alta velocidade de captura, descoberta e análise.



Q3- [FCC SEFAZ-SC 2018] No âmbito da ciência de dados na definição de *Big Data*, utilizam-se características ou atributos que alguns pesquisadores adotam como sendo os cinco Vs. Porém, a base necessária para o reconhecimento de *Big Data* é formada por três propriedades:

- A) valor, velocidade e volume.
- B) valor, veracidade e volume.
- C) variedade, velocidade e volume.**
- D) variedade, valor e volume.
- E) velocidade, veracidade e volume.

Q4 – [CESPE PF 2018] Julgue o item que segue, relativo a noções de mineração de dados, *big data* e aprendizado de máquina.

***Big data* refere-se a uma nova geração de tecnologias e arquiteturas projetadas para processar volumes muito grandes e com grande variedade de dados, permitindo alta velocidade de captura, descoberta e análise. CERTO.**




Q5- [IADES APEX Brasil 2018] Assinale a alternativa que apresenta o conceito de *Big Data*.

- A) Conjuntos de dados de grande volume que se utilizam de ferramentas especiais de processamento, pesquisa e análise, e que podem ser aproveitados no tempo necessário, com precisão e grande velocidade. valor, veracidade e volume.
- B) São bancos de dados de fácil acesso e rápida velocidade, operados como computadores pessoais.
- C) Manuseio de informações necessárias às empresas e aos negócios do mundo moderno, que podem ser armazenadas em computadores pessoais, utilizando-se a técnica de nuvem de dados.
- D) São apenas grandes volumes de dados que precisam ainda ser mais bem aproveitados pelo mundo corporativo.
- E) Refere-se a um grande número de computadores pessoais (PC) interligados entre si em uma grande rede de informação.


Q5- [IADES APEX Brasil 2018] Assinale a alternativa que apresenta o conceito de *Big Data*.

- A) Conjuntos de dados de grande volume que se utilizam de ferramentas especiais de processamento, pesquisa e análise, e que podem ser aproveitados no tempo necessário, com precisão e grande velocidade.valor, veracidade e volume.**
- B) São bancos de dados de fácil acesso e rápida velocidade, operados como computadores pessoais.
- C) Manuseio de informações necessárias às empresas e aos negócios do mundo moderno, que podem ser armazenadas em computadores pessoais, utilizando-se a técnica de nuvem de dados.
- D) São apenas grandes volumes de dados que precisam ainda ser mais bem aproveitados pelo mundo corporativo.
- E) Refere-se a um grande número de computadores pessoais (PC) interligados entre si em uma grande rede de informação.



Q6- [FCC DPE-RS 2017] Os sistemas de *Big Data* costumam ser caracterizados pelos chamados 3 Vs, sendo que o V de

- A) Veracidade corresponde à rapidez na geração e obtenção de dados.
- B) Valor corresponde à grande quantidade de dados acumulada.
- C) Volume corresponde à rapidez na geração e obtenção de dados.
- D) Velocidade corresponde à confiança na geração e obtenção dos dados.
- E) Variedade corresponde ao grande número de tipos ou formas de dados.



Q6- [FCC DPE-RS 2017] Os sistemas de *Big Data* costumam ser caracterizados pelos chamados 3 Vs, sendo que o V de

- A) Veracidade corresponde à rapidez na geração e obtenção de dados.
- B) Valor corresponde à grande quantidade de dados acumulada.
- C) Volume corresponde à rapidez na geração e obtenção de dados.
- D) Velocidade corresponde à confiança na geração e obtenção dos dados.
- E) Variedade corresponde ao grande número de tipos ou formas de dados.**



Aproximadamente 80% dos dados são não-estruturados ou estão em diferentes formatos, o que dificulta a análise.

Dados estruturados x não estruturados

DADOS ESTRUTURADOS



bases de dados

DADOS NÃO ESTRUTURADOS



Dados estruturados x não estruturados

Formato dos Dados

Fonte dos Dados

Interno



Estruturado



- Resultados de pesquisas
- Registros de vendas
- Medidas de controle de processos
- Bancos de dados de sistemas internos (ERP, CRM)

Não-estruturado



- E-mails, cartas, mensagens de texto
- Legendas de vídeos
- Comentários de clientes
- Mensagens de voz
- Imagens / ilustrações
- Avaliação de funcionários

Externo



- Likes do Facebook, retweets
- Horário de publicação de posts, tweets, updates
- Pontuação em sites de classificação

- Conteúdo publicado em redes sociais
- Comentários em fóruns online
- Imagens
- Vídeos de câmeras de segurança




Q7- [INSTITUTO AOCP CASAN 2016] *Em relação à Big Data e NoSQL, é correto afirmar que:*

- A) são conceitos concorrentes, portanto não podem ser implementados juntos.*
- B) são conceitos que se complementam e com características eficientes para trabalhar com pequenas quantidades de informações.*
- C) são duas ferramentas de empresas concorrentes.*
- D) são conceitos que se complementam.*
- E) os SGBDs Oracle e MySQL são implementações desses conceitos.*

Q7- [INSTITUTO AOCP CASAN 2016] *Em relação à Big Data e NoSQL, é correto afirmar que:*

- A) são conceitos concorrentes, portanto não podem ser implementados juntos.*
- B) são conceitos que se complementam e com características eficientes para trabalhar com pequenas quantidades de informações.*
- C) são duas ferramentas de empresas concorrentes.*
- D) são conceitos que se complementam.**
- E) os SGBDs Oracle e MySQL são implementações desses conceitos.*



Q8- [FCC SEFAZ-SC 2018] As soluções em Big Data Analytics, usadas, por exemplo, pela Fazenda Pública principalmente para evitar sonegações de tributos, trabalham com algoritmos complexos, agregando dados de origens diversas, relacionando-os e gerando conclusões fundamentais para a tomada de decisões. Na execução dessas análises pelos auditores, considere:

I. Dados estruturados.

II. Dados semiestruturados.

III. Dados não estruturados.

IV. Dados brutos, não processados.

V. Esquemas de dados gerados no momento da gravação.

Sobre um repositório de armazenamento, que contenha uma grande quantidade de dados a ser examinada, deverão ser utilizados APENAS os que constam de:


A) I, III e IV.

B) I, II, III e V.

C) III, IV e V.

D) I, II, III e IV.

E) I, II, IV e V.



Q8- [FCC SEFAZ-SC 2018] As soluções em Big Data Analytics, usadas, por exemplo, pela Fazenda Pública principalmente para evitar sonegações de tributos, trabalham com algoritmos complexos, agregando dados de origens diversas, relacionando-os e gerando conclusões fundamentais para a tomada de decisões. Na execução dessas análises pelos auditores, considere:

I. Dados estruturados.

II. Dados semiestruturados.

III. Dados não estruturados.

IV. Dados brutos, não processados.

V. Esquemas de dados gerados no momento da gravação.

Sobre um repositório de armazenamento, que contenha uma grande quantidade de dados a ser examinada, deverão ser utilizados APENAS os que constam de:

A) I, III e IV.

B) I, II, III e V.

C) III, IV e V.

D) I, II, III e IV.

E) I, II, IV e V.

Como armazenar e processar todos esses dados, se o volume aumenta de forma exponencial?



X



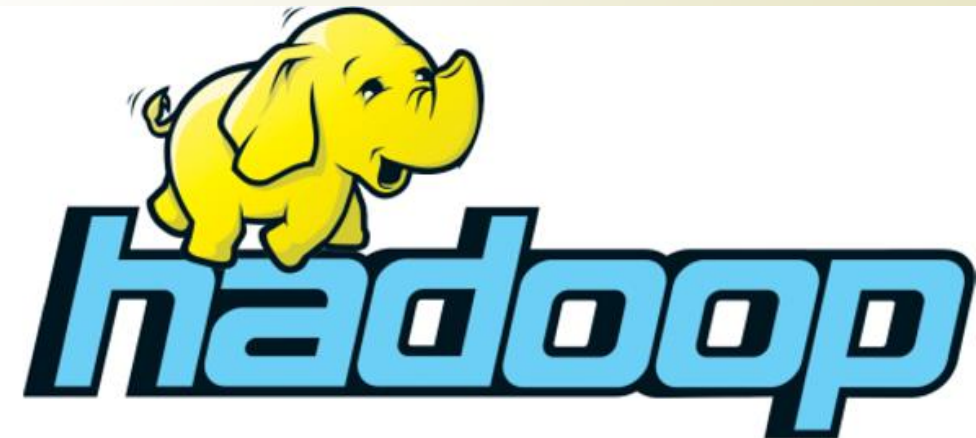
Computação Distribuída

- ▶ Executa aplicações através de máquinas diferentes, como se estas fossem uma só.
- ▶ Tornou-se possível com a popularização das redes de computadores.
- ▶ As máquinas podem estar interligadas por redes intranets, internet, redes públicas e privadas



Welcome to Hadoop

- Software open source para armazenamento e processamento em larga escala de grandes conjuntos de dados (Big Data), em clusters de hardware de baixo custo.
- O projeto Apache hadoop é composto de 3 módulos principais:
 1. Hadoop Distributed File System (HDFS).
 2. Hadoop Yarn.
 3. Hadoop MapReduce.





Apache Hadoop Ecosystem



Ambari

Provisioning, Managing and Monitoring Hadoop Clusters



Sqoop
Data Exchange



Zookeeper
Coordination



Oozie
Workflow



Pig
Scripting



Mahout
Machine Learning

R Connectors
Statistics



Hive
SQL Query



Hbase
Columnar Store



Flume
Log Collector



HDFS

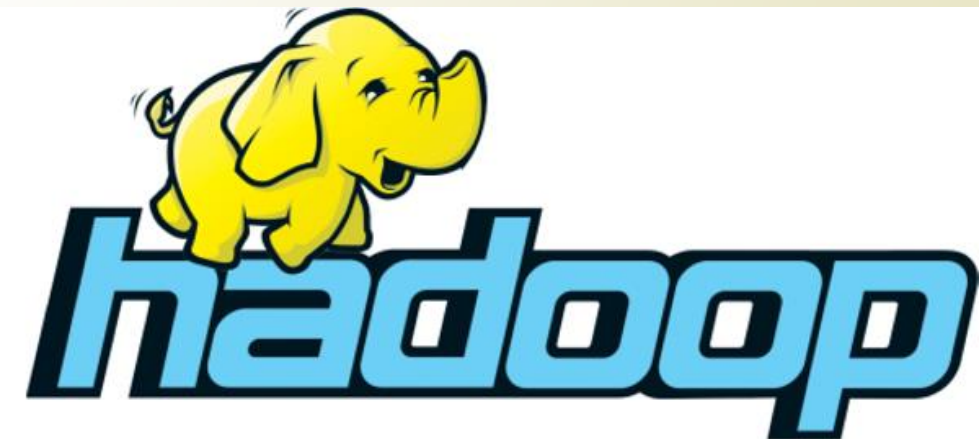
Hadoop Distributed File System

YARN Map Reduce v2
Distributed Processing Framework

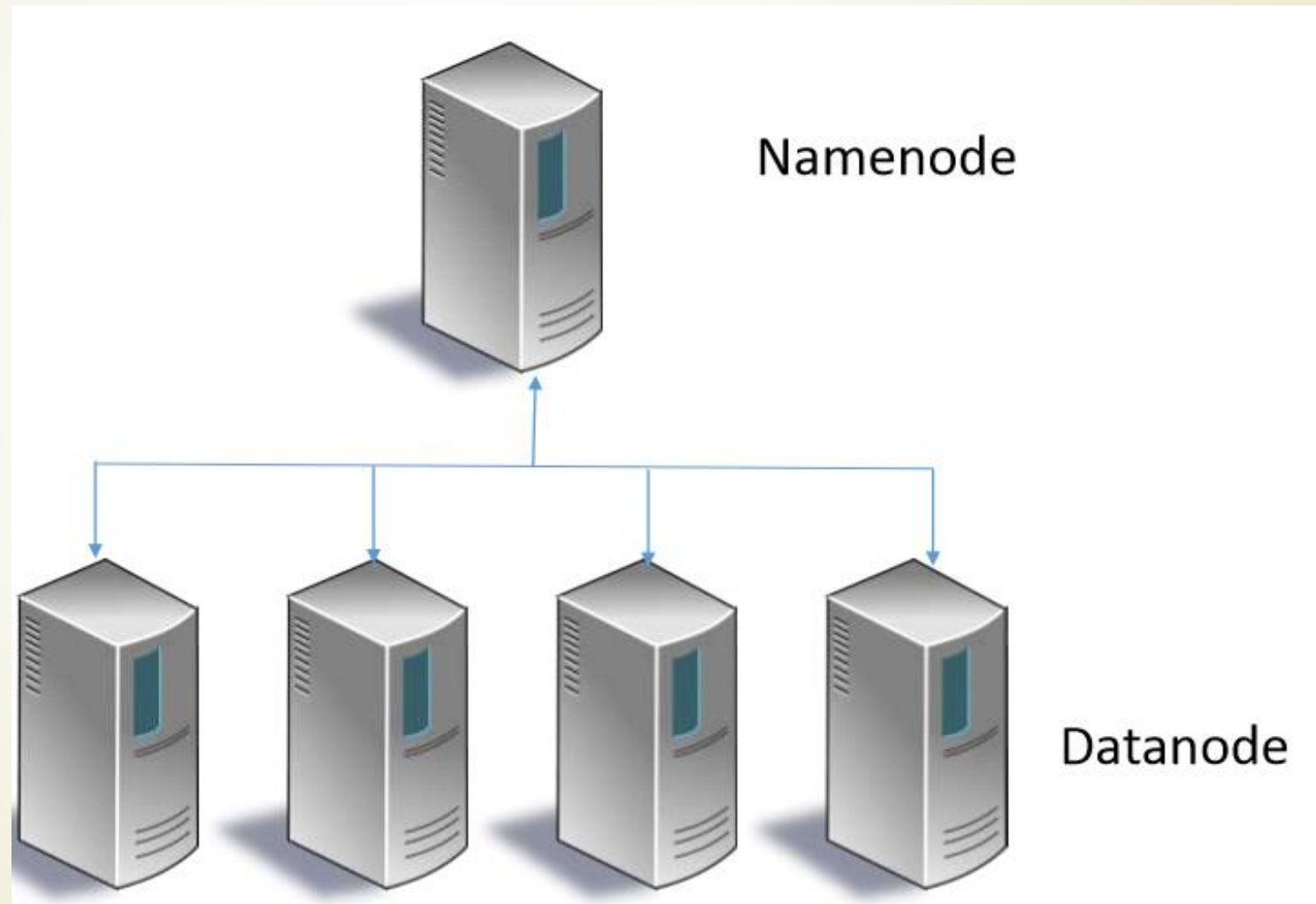


HDFS

- É otimizado para armazenar grandes arquivos.
- Pensado para executar em clusters de computadores de baixo custo.
- Foi pensado para ser ótimo em performance do tipo WORM (Write Once, Read Many Times), que é um eficiente padrão de processamento de dados.



HDFS



MapReduce

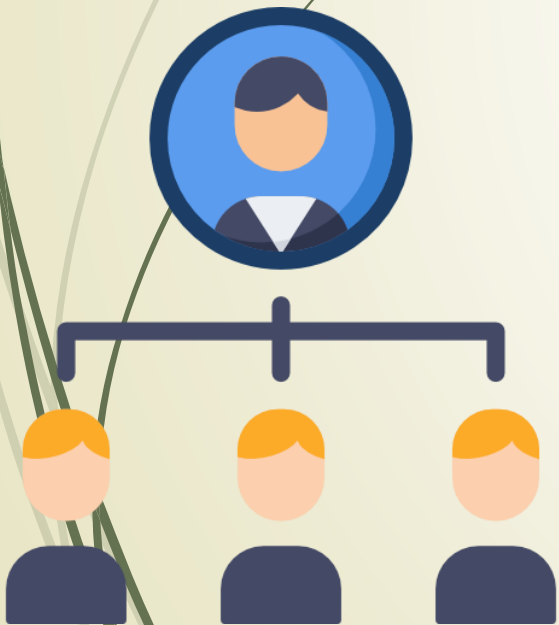
- Modelo de programação criado pelo Google para trabalhar com grande quantidade de dados.
- Consiste de duas fases: **Mapeamento** e **Redução**.



MapReduce – Etapa de Mapeamento

- ▶ Que tal aumentar o número de cestas, e consequentemente, colocar menos laranja em cada uma das cestas?
- ▶ Cada pessoa realizará a contagem em cada cesta.

Master

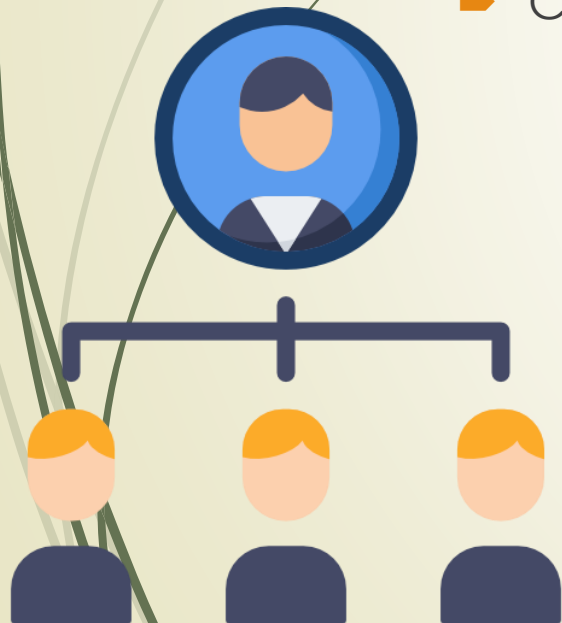


Slaves



MapReduce – Etapa de Redução

- Cesta 1: três laranjas, duas maduras.
- Cesta 2: três laranjas, três maduras.
- Cesta 3: três laranjas, nenhuma madura.
- Cesta 4: três laranjas, uma madura.



Master

Nodes





Q9- [CESPE TCE-PB 2018 ADAP] Julgue o item que se segue sobre Big Data.

A definição mais ampla de *big data* restringe o termo a duas partes — o volume absoluto e a velocidade —, o que facilita a extração das informações e dos *insights* de negócios.

Q10- [CESPE TCE-PB 2018 ADAP] Julgue o item que se segue sobre Big Data.

Para armazenar e recuperar grande volume de dados, o *big data* utiliza bancos SQL nativos, que são bancos de dados que podem estar configurados em quatro tipos diferentes de armazenamentos: valor chave, colunar, gráfico ou documento.

Q11- [CESPE TCE-PB 2018 ADAP] Julgue o item que se segue sobre Big Data.

O MapReduce é considerado um modelo de programação que permite o processamento de dados massivos em um algoritmo paralelo e distribuído.

Q9- [CESPE TCE-PB 2018 ADAP] Julgue o item que se segue sobre Big Data.

A definição mais ampla de *big data* restringe o termo a duas partes — o volume absoluto e a velocidade —, o que facilita a extração das informações e dos *insights* de negócios. ERRADO.

Q10- [CESPE TCE-PB 2018 ADAP] Julgue o item que se segue sobre Big Data.

Para armazenar e recuperar grande volume de dados, o *big data* utiliza bancos SQL nativos, que são bancos de dados que podem estar configurados em quatro tipos diferentes de armazenamentos: valor chave, colunar, gráfico ou documento. ERRADO.

Q11- [CESPE TCE-PB 2018 ADAP] Julgue o item que se segue sobre Big Data.

O MapReduce é considerado um modelo de programação que permite o processamento de dados massivos em um algoritmo paralelo e distribuído. CERTO.

Q12- [CESPE TCE PE 2017] Com relação a *Big Data*, julgue o item subsequente.

Além de estar relacionado à grande quantidade de informações a serem analisadas, o *Big Data* considera o volume, a velocidade e a variedade dos dados estruturados — dos quais se conhece a estrutura de armazenamento — bem como dos não estruturados, como imagens, vídeos, áudios e documentos.

Q13- [CESPE PF 2018] Em um *big data*, alimentado com os dados de um sítio de comércio eletrônico, são armazenadas informações diversificadas, que consideram a navegação dos usuários, os produtos comprados e outras preferências que o usuário demonstre nos seus acessos.

Tendo como referência as informações apresentadas, julgue o item seguinte.

O *big data* consiste de um grande depósito de dados estruturados, ao passo que os dados não estruturados são considerados *data files*.

Q14 – [CESPE TCE PE 2017] No que se refere a *Big Data*, julgue o item subsecutivo.

O termo *Big Data Analytics* refere-se aos poderosos softwares que tratam dados estruturados e não estruturados para transformá-los em informações úteis às organizações, permitindo-lhes analisar dados, como registros de call center, postagens de redes sociais, de blogs, dados de CRM e demonstrativos de resultados.

Q12- [CESPE TCE PE 2017] Com relação a *Big Data*, julgue o item subsequente.

Além de estar relacionado à grande quantidade de informações a serem analisadas, o *Big Data* considera o volume, a velocidade e a variedade dos dados estruturados — dos quais se conhece a estrutura de armazenamento — bem como dos não estruturados, como imagens, vídeos, áudios e documentos. CERTO.

Q13- [CESPE PF 2018] Em um *big data*, alimentado com os dados de um sítio de comércio eletrônico, são armazenadas informações diversificadas, que consideram a navegação dos usuários, os produtos comprados e outras preferências que o usuário demonstre nos seus acessos.
Tendo como referência as informações apresentadas, julgue o item seguinte.

O *big data* consiste de um grande depósito de dados estruturados, ao passo que os dados não estruturados são considerados *data files*.
ERRADO.

Q14 – [CESPE TCE PE 2017] No que se refere a *Big Data*, julgue o item subsequente.

O termo *Big Data Analytics* refere-se aos poderosos softwares que tratam dados estruturados e não estruturados para transformá-los em informações úteis às organizações, permitindo-lhes analisar dados, como registros de call center, postagens de redes sociais, de blogs, dados de CRM e demonstrativos de resultados. CERTO.



Q15 – [CESPE EBSERH 2018] Com relação a banco de dados, julgue o item seguinte.

As soluções de big data focalizam dados que já existem, descartam dados não estruturados e disponibilizam os dados estruturados.

Q16 – [CESPE PF 2018] Julgue o item seguinte, a respeito de *big data* e tecnologias relacionadas a esse conceito.

MapReduce permite o processamento de dados massivos usando um algoritmo paralelo mas não distribuído.





Q15 – [CESPE EBSE RH 2018] Com relação a banco de dados, julgue o item seguinte.

As soluções de big data focalizam dados que já existem, descartam dados não estruturados e disponibilizam os dados estruturados. ERRADO.

Q16 – [CESPE PF 2018] Julgue o item seguinte, a respeito de *big data* e tecnologias relacionadas a esse conceito.

MapReduce permite o processamento de dados massivos usando um algoritmo paralelo mas não distribuído. ERRADO.






Q17- [ESAF ANAC 2016] Para o processamento de grandes massas de dados, no contexto de Big Data, é muito utilizada uma plataforma de software em Java, de computação distribuída, voltada para clusters, inspirada no MapReduce e no GoogleFS. Esta plataforma é o(a)

- A) Yam Common.
- B) GoogleCrush.
- C) EMRx.
- D) Hadoop.
- E) MapFix.

Q18- [CESPE PF 2018] Julgue o item seguinte, a respeito de *big data* e tecnologias relacionadas a esse conceito.

MapReduce oferece um modelo de programação com processamento por meio de uma combinação entre chaves e valores.



Q17- [ESAF ANAC 2016] Para o processamento de grandes massas de dados, no contexto de Big Data, é muito utilizada uma plataforma de software em Java, de computação distribuída, voltada para clusters, inspirada no MapReduce e no GoogleFS. Esta plataforma é o(a)

A) Yam Common.

B) GoogleCrush.

C) EMRx.

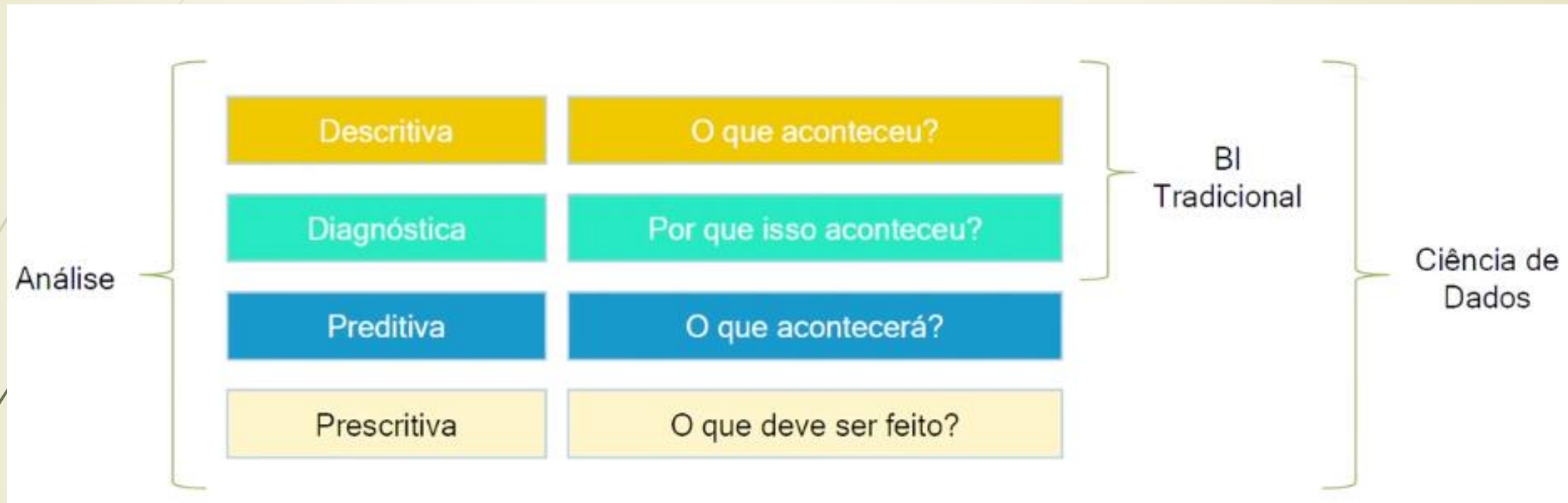
D) Hadoop.

E) MapFix.

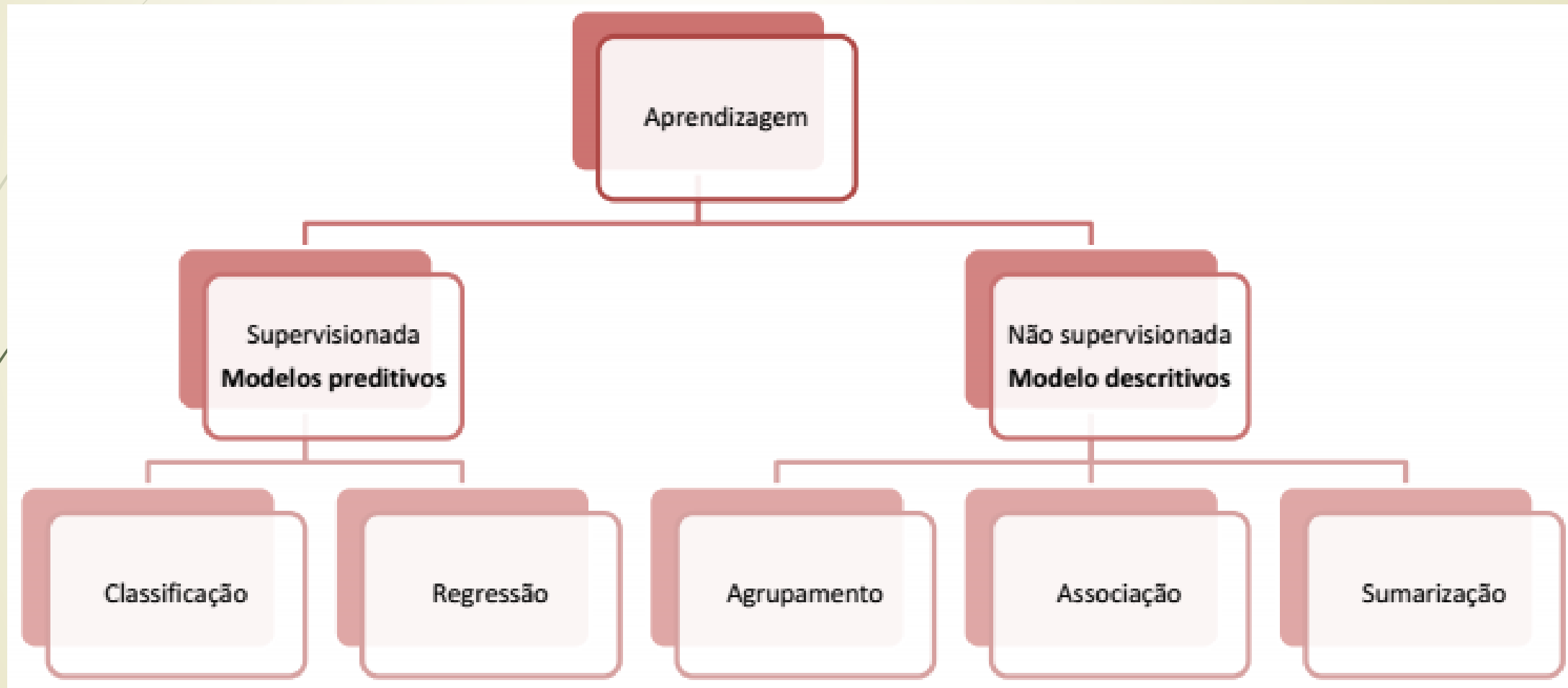
Q18- [CESPE PF 2018] Julgue o item seguinte, a respeito de *big data* e tecnologias relacionadas a esse conceito.

MapReduce oferece um modelo de programação com processamento por meio de uma combinação entre chaves e valores. CERTO.

O que fazer com esses dados?



Machine Learning



O que fazer com todos esses dados?



Após o atentado do World Trade Center, as autoridades de segurança norte-americanas descobriram, após uma análise minuciosa de toda a base de dados, estruturados ou não, armazenados nos sistemas de controle de acesso nos aeroportos e em escolas de pilotagem, que teria sido possível prever um plano de ataque que estava sendo desenvolvido.

Coletando Insights



Coletando Insights

- Série House of Cards
 - A partir da análise de navegação e reviews, Netflix percebeu que poderia criar uma série de sucesso;
 - Viu que Kevin Spacey tinha grande aceitação a partir da análise de dados;
 - Entendeu que thrillers políticos tinham grande apelo com o seu público;
- Recomendação de filmes
 - Utiliza machine learning com técnicas de filtro colaborativo e *k-nearest neighbor*.
 - Recomenda filmes com precisão com um catálogo gigantesco. No Brasil o catálogo é menor, mas ainda assim a precisão é muito boa.




Coletando Insights

Sistemas de recomendação de compras de produtos em seu e-commerce.


amazon





Q19- [CESPE TCE-MG 2018] Um dos desdobramentos de big data é o big data analytics, que se refere aos softwares capazes de tratar dados para transformá-los em informações úteis às organizações. O big data analytics difere do business intelligence por:

- A) priorizar o ambiente de negócios em detrimento de outras áreas.
- B) analisar dúvidas já conhecidas para as quais se deseje obter resposta.
- C) analisar o que já existe e o que está por vir, apontando novos caminhos.
- D) dar enfoque à coleta, à transformação e à disponibilização dos dados.
- E) analisar o que já existe, definindo as melhores hipóteses.



Q19- [CESPE TCE-MG 2018] Um dos desdobramentos de big data é o big data analytics, que se refere aos softwares capazes de tratar dados para transformá-los em informações úteis às organizações. O big data analytics difere do business intelligence por:

- A) priorizar o ambiente de negócios em detrimento de outras áreas.
- B) analisar dúvidas já conhecidas para as quais se deseje obter resposta.
- C) analisar o que já existe e o que está por vir, apontando novos caminhos.**
- D) dar enfoque à coleta, à transformação e à disponibilização dos dados.
- E) analisar o que já existe, definindo as melhores hipóteses.



Q20- [CESPE POLICIA FEDERAL 2018] Julgue o item seguinte, a respeito de big data e tecnologias relacionadas a esse conceito.

De maneira geral, big data não se refere apenas aos dados, mas também às soluções tecnológicas criadas para lidar com dados em volume, variedade e velocidade significativos.

Q21- [CESPE TCE-MG 2018] Uma empresa, ao implementar técnicas e softwares de big data, deu enfoque diferenciado à análise que tem como objetivo mostrar as consequências de determinado evento. Essa análise é do tipo:

- A) preemptiva.
- B) perceptiva.
- C) prescritiva.
- D) preditiva.
- E) evolutiva.



Q20- [CESPE POLICIA FEDERAL 2018] Julgue o item seguinte, a respeito de big data e tecnologias relacionadas a esse conceito.

De maneira geral, big data não se refere apenas aos dados, mas também às soluções tecnológicas criadas para lidar com dados em volume, variedade e velocidade significativos.

CERTO

Q21- [CESPE TCE-MG 2018] Uma empresa, ao implementar técnicas e softwares de big data, deu enfoque diferenciado à análise que tem como objetivo mostrar as consequências de determinado evento. Essa análise é do tipo:

A) preemptiva.

B) perceptiva.

C) prescritiva.

D) preditiva.

E) evolutiva.

Q22- [CESPE FUNPRESP-JUD 2016] A respeito de banco de dados, julgue o próximo item.

Uma big data não engloba dados não estruturados, mas inclui um imenso volume de dados estruturados suportado por tecnologias como o DataMining e o DataWarehouse para a obtenção de conhecimento a partir da manipulação desses dados.

Q23- [CESPE PF 2018] Em um *big data*, alimentado com os dados de um sítio de comércio eletrônico, são armazenadas informações diversificadas, que consideram a navegação dos usuários, os produtos comprados e outras preferências que o usuário demonstre nos seus acessos. Tendo como referência as informações apresentadas, julgue o item seguinte.


Dados coletados de redes sociais podem ser armazenados, correlacionados e expostos com o uso de análises preditivas.

Q22 - [CESPE FUNPESP-JUD 2016] A respeito de banco de dados, julgue o próximo item.

Uma big data não engloba dados não estruturados, mas inclui um imenso volume de dados estruturados suportado por tecnologias como o DataMining e o DataWarehouse para a obtenção de conhecimento a partir da manipulação desses dados. ERRADO.

Q23- [CESPE PF 2018] Em um *big data*, alimentado com os dados de um sítio de comércio eletrônico, são armazenadas informações diversificadas, que consideram a navegação dos usuários, os produtos comprados e outras preferências que o usuário demonstre nos seus acessos. Tendo como referência as informações apresentadas, julgue o item seguinte.

Dados coletados de redes sociais podem ser armazenados, correlacionados e expostos com o uso de análises preditivas. CERTO.




Q24- [CESPE PF 2018] Em um *big data*, alimentado com os dados de um sítio de comércio eletrônico, são armazenadas informações diversificadas, que consideram a navegação dos usuários, os produtos comprados e outras preferências que o usuário demonstre nos seus acessos. Tendo como referência as informações apresentadas, julgue o item seguinte.

Uma aplicação que reconheça o acesso de um usuário e forneça sugestões diferentes para cada tipo de usuário pode ser considerada uma aplicação que usa machine learning.

Q25 – [CESPE PF 2018] Julgue o item que segue, relativo a noções de mineração de dados, big data e aprendizado de máquina.

Situação hipotética: Na ação de obtenção de informações por meio de aprendizado de máquina, verificou-se que o processo que estava sendo realizado consistia em examinar as características de determinado objeto e atribuir-lhe uma ou mais classes; verificou-se também que os algoritmos utilizados eram embasados em algoritmos de aprendizagem supervisionados.

Assertiva: Nessa situação, a ação em realização está relacionada ao processo de classificação.



Q24- [CESPE PF 2018] Em um *big data*, alimentado com os dados de um sítio de comércio eletrônico, são armazenadas informações diversificadas, que consideram a navegação dos usuários, os produtos comprados e outras preferências que o usuário demonstre nos seus acessos. Tendo como referência as informações apresentadas, julgue o item seguinte.


Uma aplicação que reconheça o acesso de um usuário e forneça sugestões diferentes para cada tipo de usuário pode ser considerada uma aplicação que usa machine learning. CERTO.

Q25 – [CESPE PF 2018] Julgue o item que segue, relativo a noções de mineração de dados, big data e aprendizado de máquina.

Situação hipotética: Na ação de obtenção de informações por meio de aprendizado de máquina, verificou-se que o processo que estava sendo realizado consistia em examinar as características de determinado objeto e atribuir-lhe uma ou mais classes; verificou-se também que os algoritmos utilizados eram embasados em algoritmos de aprendizagem supervisionados.


Assertiva: Nessa situação, a ação em realização está relacionada ao processo de classificação. CERTO.





Q26- [FCC MPE PE 2018] Uma organização que lida com um grande volume de dados estruturados e não estruturados objetiva organizar esses dados para encontrar insights necessários para o negócio usando técnicas de aprendizagem de máquina. Terá maiores chances de sucesso para atingir seus objetivos investindo na área de

- A) Data Science.
- B) Business Intelligence.
- C) Big Data.
- D) Governança de Dados.
- E) Qualidade de Dados.



Q26- [FCC MPE PE 2018] Uma organização que lida com um grande volume de dados estruturados e não estruturados objetiva organizar esses dados para encontrar insights necessários para o negócio usando técnicas de aprendizagem de máquina. Terá maiores chances de sucesso para atingir seus objetivos investindo na área de

A) Data Science.

B) Business Intelligence.

C) Big Data.

D) Governança de Dados.

E) Qualidade de Dados.



Q27- [CESPE TCE MG 2018] Em machine learning, a categoria de aprendizagem por reforço identifica as tarefas em que


- A) um *software* interage com um ambiente dinâmico, como, por exemplo, veículos autônomos.
- B) as etiquetas de classificação não sejam fornecidas ao algoritmo, de modo a deixá-lo livre para entender as entradas recebidas.
- C) o aprendizado pode ser um objetivo em si mesmo ou um meio para se atingir um fim.
- D) o objetivo seja aprender um conjunto de regras generalistas para converter as entradas em saídas predefinidas.
- E) são apresentados ao computador exemplos de entradas e saídas desejadas, fornecidas por um orientador.



Q27- [CESPE TCE MG 2018] Em machine learning, a categoria de aprendizagem por reforço identifica as tarefas em que

- A) um software interage com um ambiente dinâmico, como, por exemplo, veículos autônomos.**
- B) as etiquetas de classificação não sejam fornecidas ao algoritmo, de modo a deixá-lo livre para entender as entradas recebidas.
- C) o aprendizado pode ser um objetivo em si mesmo ou um meio para se atingir um fim.
- D) o objetivo seja aprender um conjunto de regras generalistas para converter as entradas em saídas predefinidas.
- E) são apresentados ao computador exemplos de entradas e saídas desejadas, fornecidas por um orientador.

GABARITO



Q1 – LETRA D.
Q2 - LETRA B.
Q3 – LETRA C.
Q4 - CERTO.
Q5 - LETRA A.
Q6 - LETRA E.
Q7 – LETRA D.
Q8 – LETRA D.
Q9 - ERRADO.
Q10 - ERRADO.
Q11 – CERTO.
Q12 - CERTO.
Q13 - ERRADO.
Q27 – LETRA A.

Q14 – CERTO.
Q15 – ERRADO.
Q16 - ERRADO.
Q17 – LETRA D.
Q18 – CERTO.
Q19 - LETRA C.
Q20 – CERTO.
Q21 – LETRA C.
Q22 – ERRADO.
Q23 – CERTO.
Q24 – CERTO.
Q25 – CERTO.
Q26 – LETRA A.