



## UNIDADE I

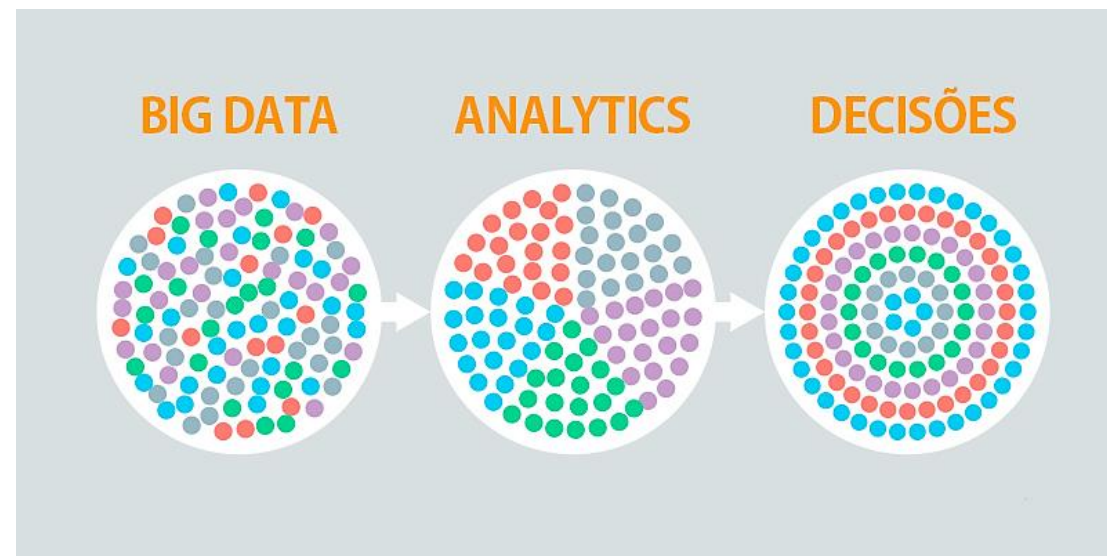
---

Estudos Disciplinares  
*Big Data Analytics*

Prof. Dr. Kauê Rosalem

# O que é *Big Data*?

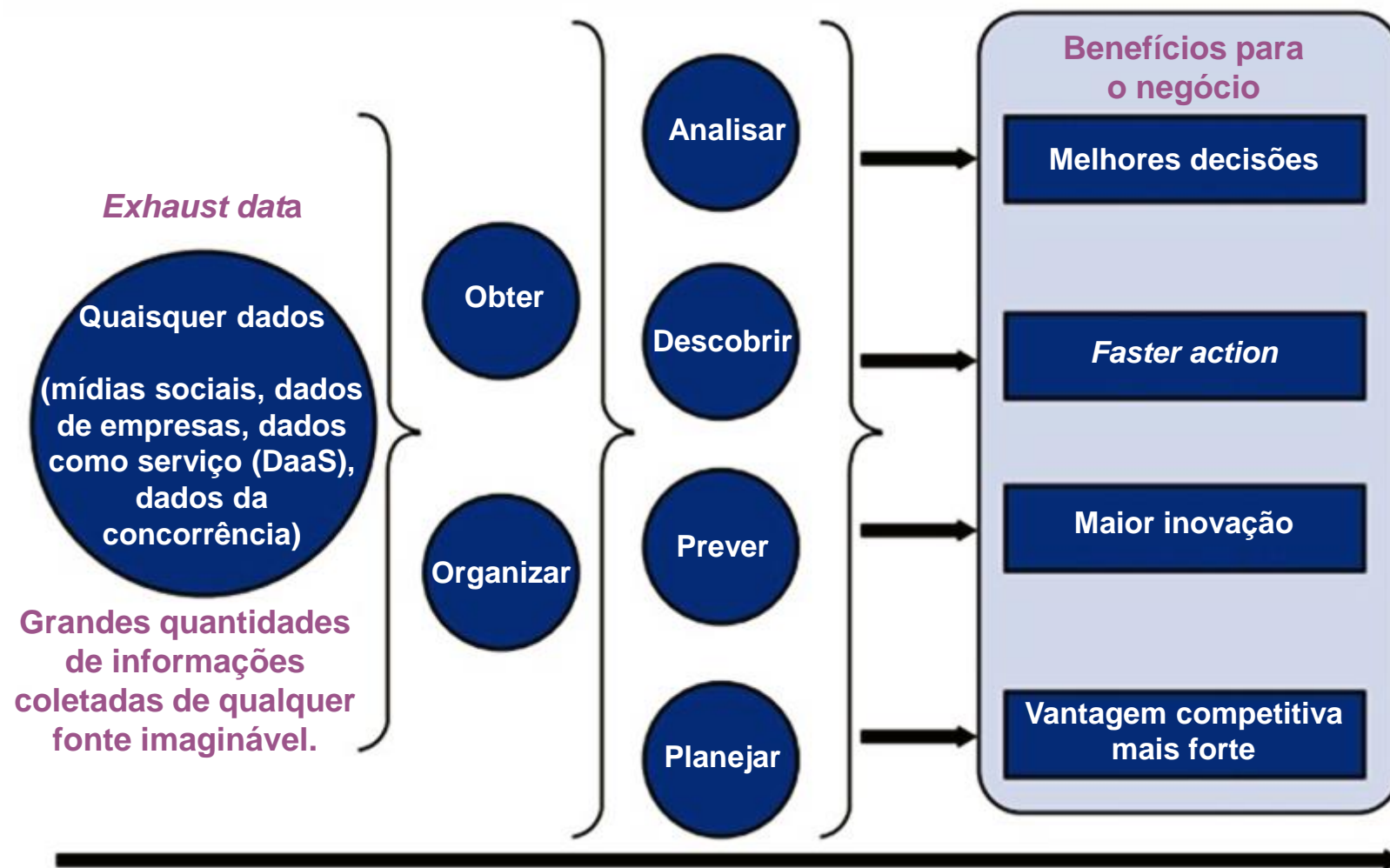
- *Big Data* é uma coleção de conjuntos de dados, grandes e complexos, que **não podem ser processados por bancos de dados ou aplicações de processamento tradicionais**.
- Capacidade de uma sociedade de obter informações de maneiras novas a fim de gerar **ideias úteis, bens e serviços de valor significativo**.
- Podemos definir o conceito de *Big Data* como sendo conjuntos de **dados extremamente amplos** e que, por este motivo, necessitam de **ferramentas especialmente** preparadas para lidar com grandes volumes, de forma que toda e qualquer informação nestes meios possa ser encontrada, analisada e aproveitada em tempo hábil.



Fonte:  
<https://www.idealmarketing.com.br/>

# Importância de *Big Data*

- Surgiram tecnologias que permitem processar esta grande quantidade de dados de forma **eficiente e com baixo custo**.
- Dados podem ser transformados em *insights* de **negócios**.



# Desafios de Big Data

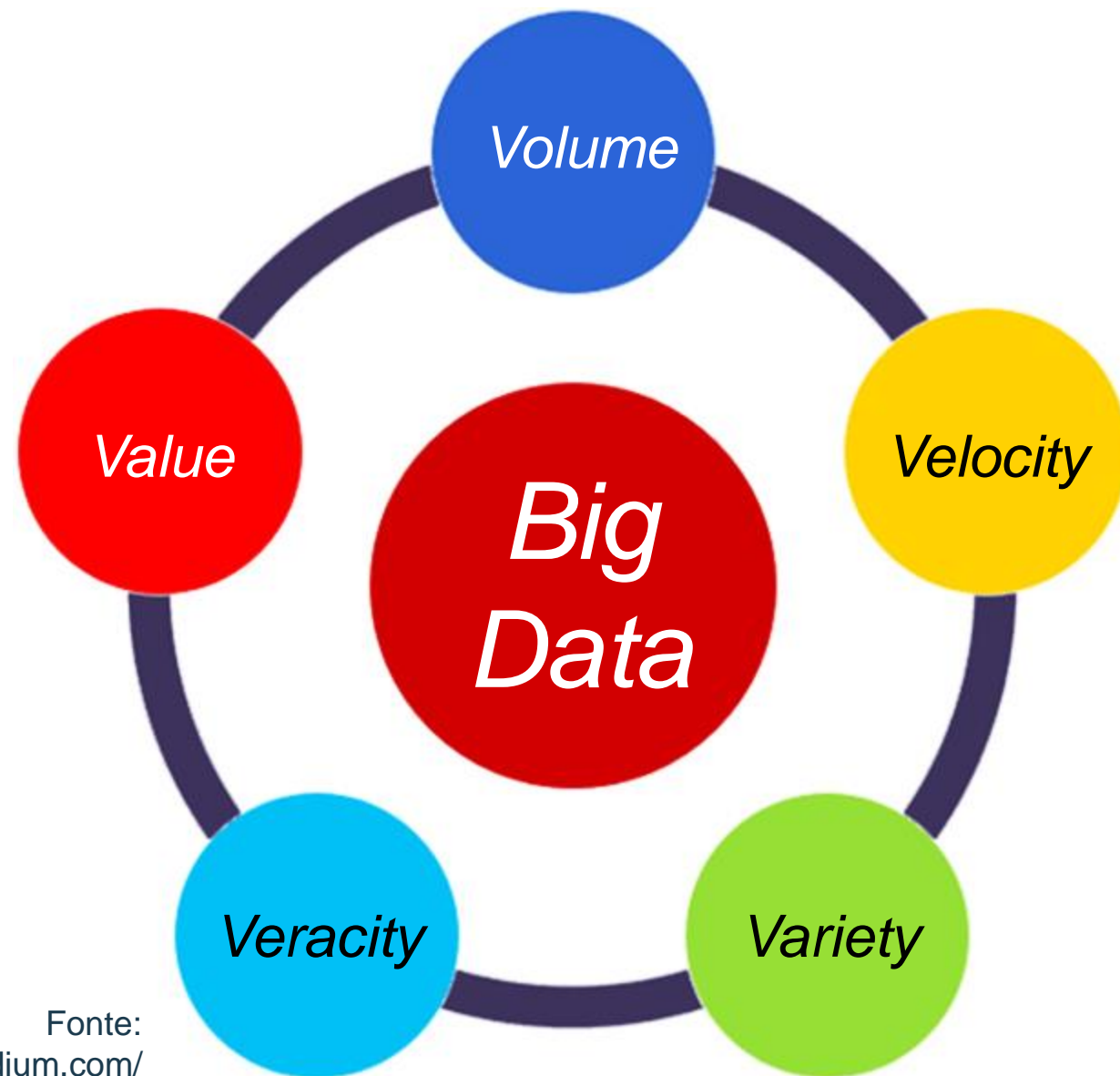
- Encontrar profissionais habilitados em *Big Data*, como **analistas e gerentes** que saibam usar *Big Data* de forma efetiva para tomada de decisões.
- Compreender as **plataformas e ferramentas** para *Big Data*.
- **Coletar, armazenar e analisar dados** de diferentes fontes, em diferentes formatos e gerados em diferentes velocidades.
- Migrar do sistema tradicional de coleta e armazenamento de dados para uma **estrutura de *Big Data***.



Fonte: <https://bleuwire.com/5-biggest-big-data-challenges/>

# Os 5 Vs do *Big Data*

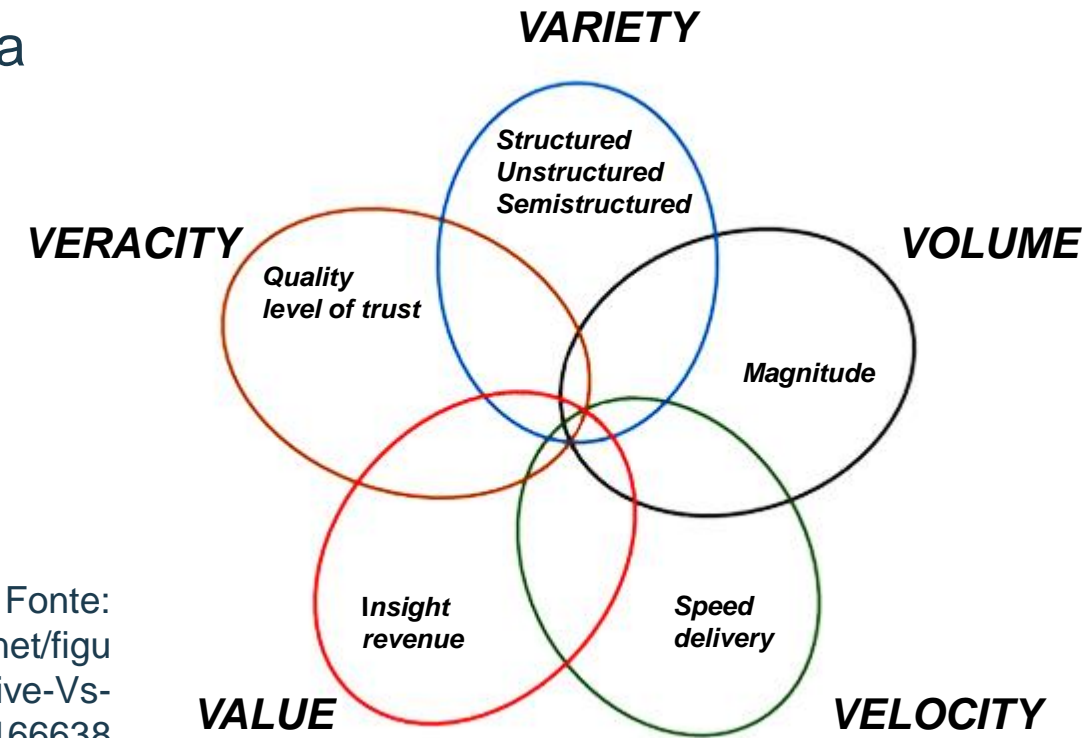
- Volume: tamanho dos dados.
- Variedade: formato dos dados.
- Velocidade: geração dos dados.
- Veracidade: confiabilidade dos dados.
- Valor: qualidade dos dados.



Fonte:  
<https://medium.com/>

# Exemplos dos 5 Vs do *Big Data*

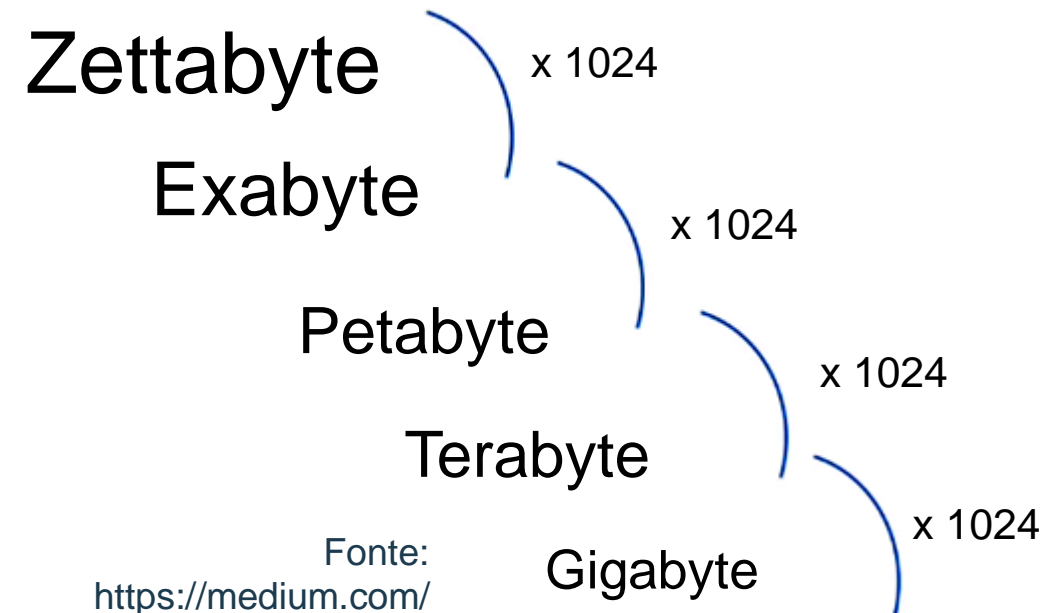
- **Volume:** existem atualmente cerca de 6 bilhões de telefones móveis no planeta.
- **Variedade:** mais de 4 bilhões de horas por mês são usadas para assistir a vídeos no YouTube.
- **Velocidade:** aproximadamente 100 sensores estão instalados nos carros modernos para monitorar nível de combustível, temperatura dos motores, pressão dos pneus e muitos outros aspectos do veículo.
- **Veracidade:** trilhões de dólares por ano são desperdiçados devido aos problemas de qualidade dos dados.
- **Valor:** dados sendo transformados em informação para valores diagnósticos e preventivos para a tomada de decisão em departamentos de recursos humanos.



Fonte:  
[https://www.researchgate.net/figure/Representation-of-the-five-Vs-of-big-data\\_fig2\\_338166638](https://www.researchgate.net/figure/Representation-of-the-five-Vs-of-big-data_fig2_338166638)

# Qual o Volume de Armazenamento do Facebook e Google?

- 1000 Kilobytes = 1 Megabyte
- 1000 Megabytes = 1 Gigabyte
- 1000 Gigabytes = 1 Terabyte
- 1000 Terabytes = 1 Petabyte
- 1000 Petabytes = 1 Exabyte
- 1000 Exabytes = 1 Zettabyte
- **1000 Zettabytes = 1 Yottabyte** → Facebook e Google estão aproximadamente aqui.
- 1000 Yottabytes = 1 Brontobyte
- 1000 Brontobytes = 1 Geobyte





# Construção da Informação

- DADO => INFORMAÇÃO => CONHECIMENTO

## Dado:

- Não tem significado por si só;
- Fatos ou figuras a serem processados.

## Informação:

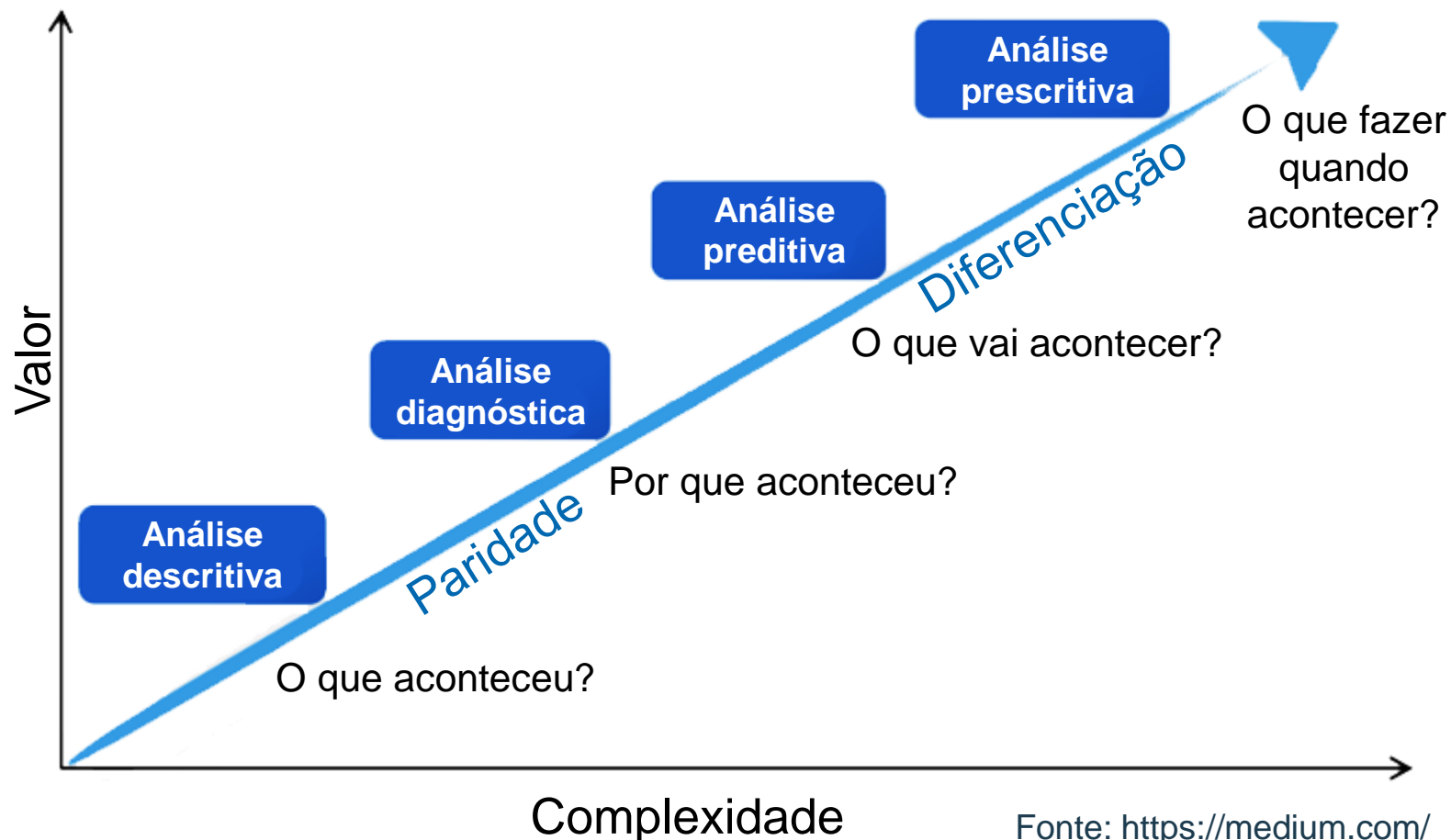
- Dado organizado de forma a ter significado para a pessoa que vai recebê-la.





# Análise de Dados (Analytics)

- **Análise descritiva:** encarrega-se de analisar o que aconteceu;
- **Análise diagnóstica:** analisa porque determinado evento aconteceu;
- **Análise preditiva:** aponta o que irá (ou poderá) acontecer (predição);
- **Análise prescritiva:** identifica o que se deseja que aconteça e o que pode ser feito para alcançar esse objetivo (reforçar, modificar ou evitar uma predição).



# Interatividade

Considere as seguintes afirmações:

- I. Processar de forma eficiente e com baixo custo grandes volume de dados.
- II. Responder ao aumento da velocidade de geração dos dados.
- III. Coletar e analisar dados de diferentes formatos e fontes.

Qual das afirmações está correta com relação às oportunidades do *Big Data*?

- a) Apenas a afirmação I.
- b) Apenas a afirmação II.
- c) Apenas a afirmação III.
- d) Afirmações I e II.
- e) Afirmações I, II e III.

# Resposta

Considere as seguintes afirmações:

- I. Processar de forma eficiente e com baixo custo grandes volume de dados.
- II. Responder ao aumento da velocidade de geração dos dados.
- III. Coletar e analisar dados de diferentes formatos e fontes.

Qual das afirmações está correta com relação às oportunidades do *Big Data*?

- a) Apenas a afirmação I.
- b) Apenas a afirmação II.
- c) Apenas a afirmação III.
- d) Afirmações I e II.
- e) Afirmações I, II e III.

## Aspectos do mundo de *Big Data*

- Cerca de 90% de todos os dados gerados no planeta foram gerados nos últimos 2 anos.
- Aproximadamente 80% dos dados são **não estruturados** ou estão em diferentes formatos, o que dificulta as análises.
- Modelos de análise de **dados estruturados** possuem **limitações** quando precisam tratar grandes volumes de dados.
- Muitas empresas não sabem quais dados precisam ser analisados.
- Muitas empresas nem mesmo sabem que os dados estão disponíveis.
- Dados preciosos são **descartados** por falta de conhecimento ou ferramentas de tratamento.
- É **caro** manter e organizar grandes volumes de dados não estruturados.
  - Dados são a matéria-prima dos **negócios**.

## Aspectos do mundo de *Big Data*

- Neste exato momento, uma verdadeira enxurrada de dados, ou 2,5 quintilhões de bytes por dia é gerada para nortear indivíduos, empresas e governos, e está dobrando a cada dois anos.
- Toda vez que fazemos uma compra, uma ligação ou interagimos nas **redes sociais**, estamos produzindo esses dados.
- Com a recente conectividade em objetos (Internet das Coisas), tais como relógios, carros e até geladeiras, as informações capturadas se tornam massivas e podem ser cruzadas para criar *roadmaps* cada vez mais elaborados, apontando e até prevendo o comportamento de empresas e clientes.
  - Entre 2005 e 2020, o universo digital irá crescer de 130 exabytes para 40.000 exabytes ou 40 trilhões de gigabytes.
  - A revolução não está nas máquinas que calculam os dados, e sim nos dados em si e na maneira como são utilizados.

# Armazenamentos de *Big Data*

- À medida que o mundo do *Big Data* se expandiu, um número crescente de opções de armazenamento surgiu na mesma proporção.
- Quando comparados, ***Data Lakes***, ***Data Hubs*** e ***Data Warehouses*** têm funções principais semelhantes, que são frequentemente confundidas com termos intercambiáveis.
- Embora as três estruturas sejam usadas para hospedar dados para análise e relatórios de negócios, elas geralmente armazenam diferentes tipos de dados e funcionam com diferentes sistemas de geração de dados e padrões de acesso.
- A diferença entre *Data Lake* x *Data Warehouse* nem sempre é bem clara, com o termo “*Data Lake*” frequentemente usado quando algo não se encaixa na arquitetura tradicional do *Data Warehouse*.
  - É importante entender as diferenças entre as opções de armazenamento de dados para encontrar uma solução que funcione para as necessidades de cada empresa.
  - Segundo o Gartner, “as consultas de clientes referentes a *Data Hubs* aumentaram 20% de 2018 a 2019”.
  - Curiosamente, o Gartner observou que “mais de 25% dessas pesquisas eram realmente sobre conceitos de *Data Lake*”.

# *Data Warehouse*

- O *Data Warehouse* é um repositório central de dados integrados e estruturados de duas ou mais fontes diferentes.
- Esse sistema é usado principalmente para relatórios e análise de dados e é considerado um componente principal da inteligência de negócios (*Business Intelligence*).
- Os *Data Warehouses* implementam padrões analíticos predefinidos e distribuídos para um grande número de usuários na empresa.
- Uma característica do *Data Warehouse* é que ele possui *schema* bem definido.
- Isso significa que os dados devem ser limpos, tratados e organizados antes de efetuar a carga no *Data Warehouse*, o que é normalmente feito durante o processo ETL (Extração, Transformação e Carga).



# Data Lake

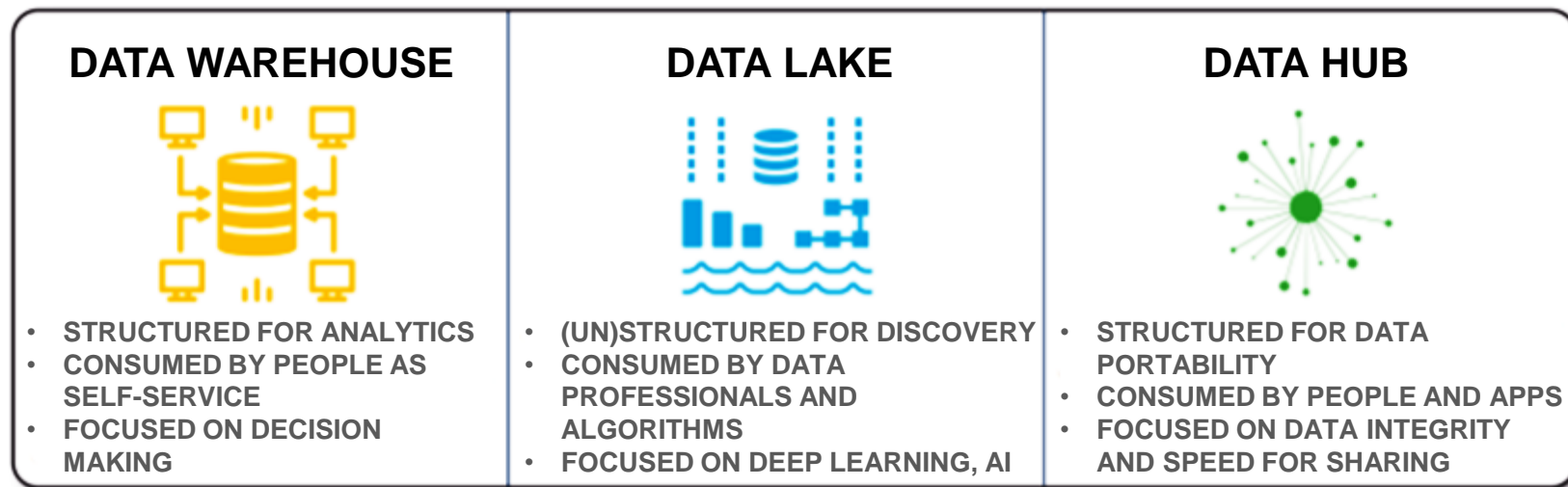
- O *Data Lake* é um repositório único de todos os dados corporativos estruturados e não estruturados.
- Ele hospeda dados não refinados com garantia de qualidade limitada e exige que o consumidor (analista) processe e adicione manualmente valor aos dados.
- Os *Data Lakes* são, em geral, uma boa base para a preparação de dados, geração de relatórios, visualização, análise avançada, *Data Science* e *Machine Learning*.
- O *Data Lake* tem um conceito inverso do DW, ou seja, não possui *schema* definido.
- Isso significa que os dados podem ser armazenados sem limpeza, tratamento ou organização, em seu estado bruto.
  - O processo de limpeza e tratamento dos dados será feito mais tarde pelo Cientista de Dados quando for usar os dados para suas análises e construção de modelos de *Machine Learning*.

# Data Hub

- O *Data Hub* é o local ideal para os dados principais de uma empresa.
- Ele centraliza os dados da empresa que são críticos entre aplicativos e permite o compartilhamento contínuo de dados entre diversos setores, enquanto é a principal fonte de dados confiáveis para a iniciativa de governança de dados.
- Os *Data Hubs* fornecem dados mestres para aplicativos e processos corporativos.
- Eles também são usados para conectar aplicativos de negócios a estruturas de análise, como *Data Warehouses* e *Data Lakes*.
- Um *Data Hub* é uma coleção simples de objetos de dados organizados de várias fontes, sendo útil quando as empresas desejam compartilhar e distribuir dados eficientemente em um ou mais formatos desejados.
  - Embora um *Data Hub* compartilhe muitas semelhanças com um *Data Warehouse*, ele não se limita aos dados operacionais e permite navegar por diferentes níveis de granularidade dos dados.

# Diferenças entre os tipos de Armazenamentos

- Em alguns casos, os *Data Warehouses* e os *Data Lakes* oferecem controles de governança, mas apenas de maneira reativa, enquanto os *Data Hubs* aplicam proativamente a governança aos dados que fluem pela infraestrutura.
- *Data Lakes*, *Data Warehouses* e *Data Hubs* não são alternativas intercambiáveis.
- No entanto, são complementares e, juntos, podem apoiar iniciativas baseadas em dados e transformação digital.



# Profissionais que trabalham com Estruturas de Armazenamento

- **Cientistas de dados** não são os responsáveis por construir *Data Warehouses* ou *Data Lakes*, sendo, na verdade, possíveis consumidores dos dados armazenados.
- **Engenheiros de dados** são os profissionais que normalmente criam e integram essas estruturas, em especial os *Data Lakes*.
- **Arquitetos de dados** são os responsáveis pela definição, *design* e integração dessas estruturas de armazenamento.
- No dia a dia, **Administradores de Bancos de Dados ou Administradores de Sistemas** podem ser os responsáveis pela administração e manutenção.
- Empresa que estão mais maduras na gestão de dados empregam equipes de Data Ops para a gestão completa de suas soluções de armazenamento e análise de dados, sendo o **Engenheiro Data Ops** o responsável pelo trabalho.



Fonte:

<https://www.sbcoaching.com.br/blog/big-data/>

# Interatividade

Qual dos armazenamentos é a principal fonte de dados confiáveis para a iniciativa de governança de dados?

- a) *Data Hub.*
- b) *Data Lake.*
- c) *Data Warehouse.*
- d) *Data Driven.*
- e) *Data Ops.*

# Resposta

Qual dos armazenamentos é a principal fonte de dados confiáveis para a iniciativa de governança de dados?

- a) *Data Hub.*
- b) *Data Lake.*
- c) *Data Warehouse.*
- d) *Data Driven.*
- e) *Data Ops.*

# Projeto de *Big Data*



Fonte: IDC

- Embora a forma como você gerencia o seu projeto de *Big Data* irá variar dependendo do caso de uso específico e perfil da empresa, existem 4 etapas principais para implementar com sucesso um projeto de *Big Data*:

1. **Definição do *Business Case*** – com objetivos claramente definidos que geram valor comercial para o negócio da empresa.
2. **Planejamento do Projeto** – um plano e um escopo bem geridos levarão ao sucesso.
3. **Definição dos Requisitos Técnicos** – os requisitos detalhados assegurarão que você construa o que precisa para alcançar seus objetivos.
4. **Criação de um “*Total Business Value Assessment*”** – uma visão holística que deve desconsiderar decisões políticas e emocionais.



# 1. Definição do Business Case

Para explorar as expectativas da organização em projetos de *Big Data*, recomendamos responder a estas perguntas primeiro:

- Qual é o objetivo do projeto?
- Em que direção o negócio está indo?
- Quais são os obstáculos para chegar lá?
- Quem são os principais interessados e quais são seus papéis?
- Qual é o caso de uso de *Big Data* mais importante determinado pelas principais partes interessadas?

Fonte:  
<https://www.knowledgetrain.co.uk/business-analysis/how-to-write-a-business-case>



# 1. Definição do Business Case

Quanto mais específicas e conectadas aos seus objetivos de negócio forem suas respostas, maior a chance do projeto ter sucesso:

- Determine os objetivos de alto nível da empresa e como *Big Data* pode suportar esses objetivos.
- Identifique a área problemática, como *marketing*, atendimento ao cliente ou desenvolvimento de negócios e as motivações por trás do projeto.
- Descreva o problema e os obstáculos em termos não técnicos.
- Faça um inventário de quaisquer soluções e ferramentas usadas atualmente para solucionar o problema comercial.
- Considere as vantagens e desvantagens das soluções atuais.
- Navegue no processo para iniciar novos projetos e implementar soluções.
  - Identifique as partes interessadas que se beneficiarão do projeto de *Big Data*.
  - Entreviste as partes interessadas individualmente para determinar os objetivos e preocupações do projeto.
  - Documente os objetivos comerciais decididos pelos principais tomadores de decisão.
  - Atribua prioridades aos objetivos de negócio.
  - Crie o *Business Case*, detalhando todos os itens anteriores.

## 2. Planejamento do Projeto

Como resultado de suas pesquisas e reuniões, você provavelmente terá um objetivo nebuloso, como “reduzir o *churn* do cliente”. Esta etapa pretende construir um objetivo concreto e específico acordado pelas partes interessadas do projeto. É nesta etapa que você deve:

- Especificar metas esperadas em termos comerciais mensuráveis.
- Identificar todas as questões comerciais com a maior precisão possível.
- Determinar quaisquer outros requisitos de negócio quantificáveis.
- Definir como seria uma implementação bem sucedida de *Big Data*.

## 2. Planejamento do Projeto

As tarefas seguintes são considerações que você pode usar para garantir que tenha capturado corretamente os critérios de sucesso:

- Com a maior precisão possível, documente os critérios de sucesso para este projeto.
- Certifique-se de que cada objetivo comercial identificado tenha um critério mensurável que determinará se esse objetivo foi cumprido com sucesso.
- Compartilhe e obtenha aprovação de seus critérios de sucesso entre os principais interessados.
- Determine o escopo adequado, especificamente o que está incluído e o que não está incluído.
  - Desenvolva um orçamento aproximado.
  - Defina uma linha de tempo e marcos de sucesso em 3 meses, 6 meses e um ano.

### 3. Definição dos Requisitos Técnicos

- Esta etapa permitirá que você determine a qualidade de seus dados e descreva os resultados dessas etapas na documentação do projeto.

Considere quais fontes de dados você vai aproveitar:

- Fontes de dados existentes. Isso inclui uma grande variedade de dados, como dados transacionais, dados de pesquisa, *logs* de servidores, dados de redes sociais, arquivos pdf, etc. Considere se suas fontes de dados existentes são suficientes para atender às suas necessidades.
  - Compra de fontes de dados. Sua organização usa dados suplementares, como dados demográficos? Isso poderia ser relevante para o processo de análise e complementar os dados existentes. A aquisição de fontes de dados confiáveis pode trazer benefícios para o resultado final do projeto.
  - Fontes de dados adicionais. Se as fontes acima não atendem às suas necessidades, talvez seja necessário realizar pesquisas ou iniciar o rastreamento adicional para complementar os dados existentes.

### 3. Definição dos Requisitos Técnicos

Ao examinar suas fontes de dados, pergunte:

- Quais atributos do(s) banco(s) de dados parecem mais promissores?
- Quais atributos parecem irrelevantes e podem ser excluídos?
- Há dados suficientes para tirar conclusões generalizáveis ou fazer previsões precisas?
- Existem muitos atributos para o seu método analítico escolhido?
- Você está mesclando várias fontes de dados? Em caso afirmativo, existem áreas que podem representar um problema na fusão?
  - Já considerou como os valores em falta (*missing data*) são tratados em cada uma das suas fontes de dados?
  - Você vai usar *Streaming* de Dados? Como esses dados serão coletados e armazenados?

### 3. Definição dos Requisitos Técnicos

Há muitas maneiras de descrever os dados, mas a maioria das descrições se concentra na quantidade e qualidade dos dados. Abaixo são apresentadas algumas características-chave para descrever os dados:

- **Volume de dados:** para a maioria das técnicas analíticas existem *trade-offs* associados ao tamanho dos dados. Grandes conjuntos de dados podem produzir modelos mais precisos, mas também podem aumentar o tempo de processamento.
- **Velocidade dos dados:** há também *trade-offs* associados com os dados em repouso ou em movimento (estático ou em tempo real). A velocidade se traduz em quão rápido os dados são criados dentro de um determinado período de tempo.



### 3. Definição dos Requisitos Técnicos

- **Variedade de dados:** os dados podem ter uma variedade de formatos, como numérico, categórico (*string*) ou Booleano (verdadeiro/falso). Prestar atenção ao tipo de dado pode evitar problemas durante análises posteriores. Frequentemente, os valores no banco de dados são representações de características como gênero ou tipo de produto. Por exemplo, um conjunto de dados pode usar M e F para representar masculino e feminino, enquanto outro pode usar os valores numéricos 1 e 2. Observe qualquer esquema conflitante nos dados.
- **Tempo para a ação:** os dados podem ser usados para tomar medidas imediatas, além de serem armazenados para futuras análises sem tempo crítico. É importante identificar quais dados provavelmente serão usados para ações em tempo real (<150ms), ações próximas em tempo real (segundos) ou ações críticas sem tempo (minutos a horas).

### 3. Definição dos Requisitos Técnicos

- Considere quais interfaces e ferramentas são necessárias para que sua empresa trabalhe com suas fontes de dados.
- Além da implementação da infraestrutura, seu projeto de *Big Data* deve oferecer a capacidade de criar aplicativos e análises personalizadas usando API e ferramentas nativas como parte do *Hadoop*, bancos de dados e processamento de fluxo – além de interfaces abstraídas e unificadas para melhorar a experiência do usuário.
- Os usuários devem ter a capacidade de produzir tabelas, gráficos e outros elementos de visualização usando ferramentas de BI, como: Business Objects, Microstrategy, Cognos, Tableau, Datameer ou outras ferramentas similares.
- Tais análises visuais podem ajudar a abordar os objetivos do projeto de *Big Data* definidos durante a fase de compreensão do negócio.
  - Outras vezes, é mais apropriado utilizar ferramentas que suportem análises estatísticas e construção de modelos de *Machine Learning*, tais como R, Python, Java, Scala, SAS, Matlab etc.
  - E claro, ainda oferecer suporte a aplicações de Inteligência Artificial.

### 3. Definição dos Requisitos Técnicos

Isso tudo nos leva a mais um grupo de questões que devem ser respondidas:

- Quem precisa trabalhar com os dados?
- Quais são suas habilidades técnicas?
- Treinamento será necessário?
- Quais ferramentas você atualmente possui em sua empresa que você gostaria de aproveitar?
- Essas ferramentas possuem conectores de *Big Data* ou métodos de interface?
- Quais novas ferramentas podem ajudar com sua mineração de dados, análise, visualização, relatórios, etc.?
  - Como e onde os dados serão armazenados?
  - Serão utilizados *Data Lakes*? Em nuvem ou *on-premises*?
  - Arquiteturas de *Enterprise Data Hubs* serão consideradas?
  - Quais são as ferramentas de relatórios e visualização necessárias para alcançar o sucesso aos olhos dos usuários finais?

# Interatividade

Em qual das etapas do projeto de *Big Data* devemos identificar todas as questões comerciais com a maior precisão possível?

- a) Definição dos Requisitos Técnicos.
- b) Planejamento do Projeto.
- c) Definição do *Business Case*.
- d) Criação de um “*Total Business Value Assessment*”.
- e) Apresentação de Resultados.

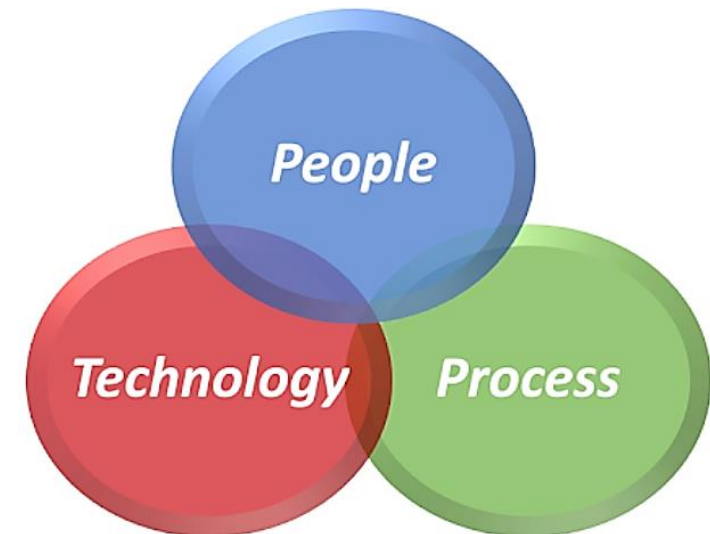
## Resposta

Em qual das etapas do projeto de *Big Data* devemos identificar todas as questões comerciais com a maior precisão possível?

- a) Definição dos Requisitos Técnicos.
- b) **Planejamento do Projeto.**
- c) Definição do *Business Case*.
- d) Criação de um “*Total Business Value Assessment*”.
- e) Apresentação de Resultados.

## 4. Criação de um “*Total Business Value Assessment*”

- Avalie suas opções com um “*Total Business Value Assessment*”.
- Isso significa que você realiza pelo menos uma análise de custo total de propriedade de 3 anos, mas você também inclui itens como o “*time-to-business value*”, facilidade de uso, escalabilidade, base em padrões e o nível de maturidade da empresa.
- No entanto, antes de começar a avaliar as opções da sua solução, é importante conhecer sua “equipe de compras”.
- As equipes de compras geralmente consistem em partes interessadas de múltiplos níveis organizacionais e às vezes divisões múltiplas fora da TI.
- No mínimo, deve haver um patrocinador executivo, líder do projeto ou líder da equipe do projeto, tomador de decisões técnicas e um decisor econômico.



# Time-to-Business Value

- Como os projetos exigem um investimento inicial significativo, é importante entender o tempo antes da solução começar a gerar valor.
- Como muitos desses projetos podem ampliar demais os prazos acordados pelo fornecedor, é recomendável solicitar informações sobre projetos de escopo semelhantes que já foram concluídos para outros clientes a fim de determinar melhor se o fornecedor entrega os marcos estabelecidos.





## Facilidade de uso

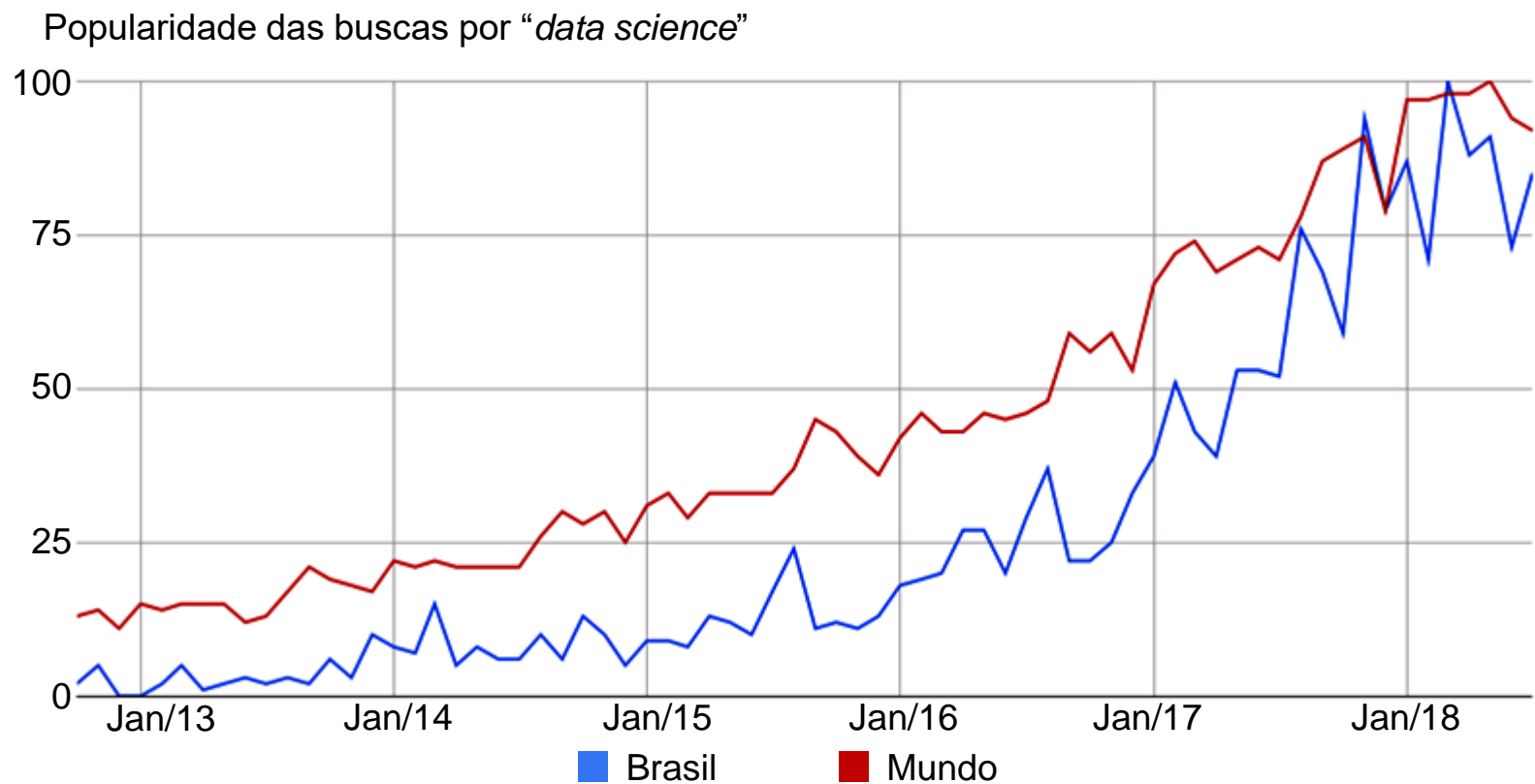
- Ao considerar as novas soluções de *Big Data*, é imperativo considerar como a solução afetará sua necessidade de aumentar os recursos internos para implementação e manutenção contínua.
- Especificamente, novos recursos, como Cientistas de Dados e Engenheiros de Dados, podem ser necessários para gerenciar um projeto internamente.
- Além disso, serão necessários recursos de TI corporativos para manter o projeto uma vez implementado.
  - A facilidade de uso também se traduz em facilidade de integração dentro de sua infraestrutura técnica interna.
  - Isso pode ter um grande impacto no valor do tempo para o negócio.

# Escalabilidade

- As soluções ideais se expandirão com as necessidades de negócio em constante evolução.
- Enquanto um caso de uso pode ser o *driver* inicial, a solução ideal irá suportar casos de usos futuros.
- Além disso, considere os custos associados à escala.
- Algumas soluções oferecem capacidade inicial a baixo custo, mas o custo aumenta rapidamente à medida que a expansão ocorre.

# Baseado em padrões

- No mundo de hoje, as tecnologias populares vão e vêm.
- Considere se suas soluções propostas utilizam tecnologias baseadas em padrões abertos e melhores em sua classe para impulsionar inovação e receita para as necessidades da empresa.
- Além disso, explore os riscos associados com a capacidade de adaptar a sua solução e sua capacidade de agir rapidamente quando precisar fazê-lo.



## Maturidade da empresa

- Um aspecto importante da seleção é se a solução pode operar dentro de uma configuração corporativa.
- As soluções ideais possuem alta disponibilidade e recuperação de desastres.
- Além disso, elas irão apoiar as suas necessidades de segurança e conformidade de negócios e atender seus SLAs definidos atualmente.
- Para entender melhor como as soluções funcionam em um ambiente empresarial, entre em contato com as referências dos clientes empresariais do provedor de soluções.

# Interatividade

■ Considere as seguintes afirmações para explicar “Ciência de Dados”:

- I. É o termo usado para definir a extração de *insights* de dados que são coletados de várias fontes.
- II. Inclui modelagem preditiva e ajuda a analisar e interpretar grandes quantidades de dados.
- III. É o processo para extrair informações valiosas a partir de dados.
- IV. Seus profissionais são chamados Cientistas de Dados.

Estão corretas as afirmações:

- a) I e II.
- b) III e IV.
- c) II e IV.
- d) I, III e IV.
- e) I, II, III e IV.

# Resposta

▪ Considere as seguintes afirmações para explicar “Ciência de Dados”:

- I. É o termo usado para definir a extração de *insights* de dados que são coletados de várias fontes.
- II. Inclui modelagem preditiva e ajuda a analisar e interpretar grandes quantidades de dados.
- III. É o processo para extrair informações valiosas a partir de dados.
- IV. Seus profissionais são chamados Cientistas de Dados.

Estão corretas as afirmações:

- a) I e II.
- b) III e IV.
- c) II e IV.
- d) I, III e IV.
- e) I, II, III e IV.

**ATÉ A PRÓXIMA!**