

# Introdução a Big Data – 3V

Profa. Leticia T. M. Zoby

([leticia.zoby@udf.edu.br](mailto:leticia.zoby@udf.edu.br))

# Introdução a Big Data

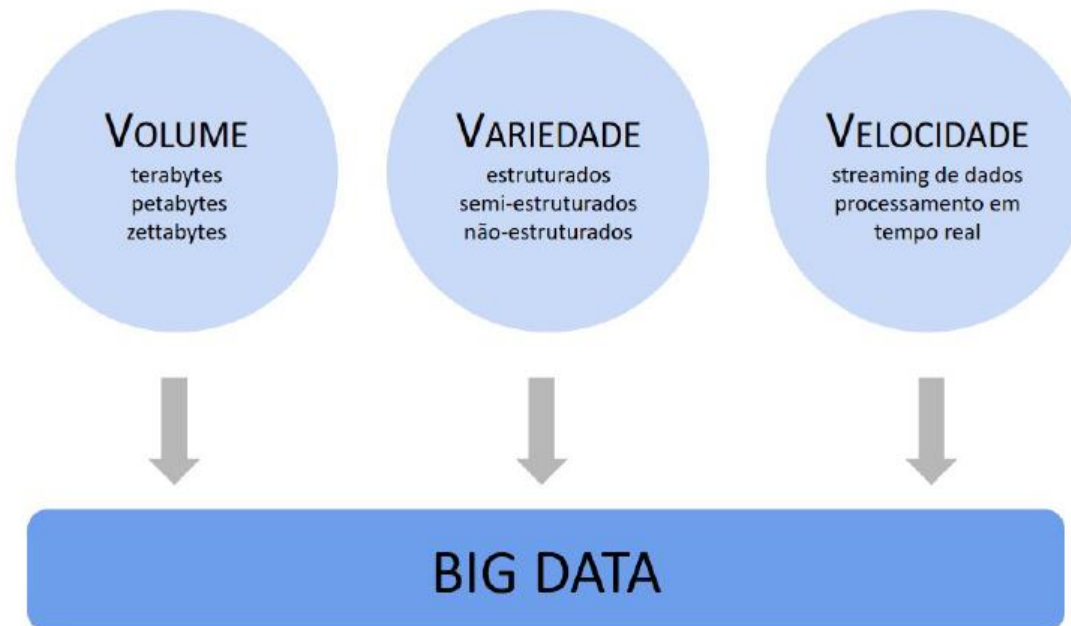
- Novos paradigmas da Internet: IoT(Internet of Things) e IoE(Internet of Everything);
- Necessidades de armazenar, processar, gerenciar e analisar tantos dados, em maneira de garantir os reais benefícios, é necessário um enorme esforço das organizações.
- O termo **Big Data** refere-se a um grande conjunto de dados armazenados. É um tipo de tecnologia que permite analisar dados em tempo real e podem ser provenientes de diferentes fontes e formas.

# Introdução a Big Data



# Introdução a Big Data

- Big Data
  - oferecer uma abordagem consistente no tratamento do constante crescimento e da complexidade dos dados.
  - 3 V's do *Big Data*: o Volume, a Variedade ,a Velocidade
    - Surgiu em 2001, com Doug Laney, da empresa Gartner Group



# Introdução a Big Data

- VOLUME

- Característica mais significativas no conceito de BD
- Faz referência à dimensão sem precedentes do volume de dados
  - 2024 ????



- ▶ A cada segundo, cerca de 40.000 buscas são realizadas no Google.
- ▶ A empresa Walmart manipula mais de 1 milhão de transações dos clientes por hora.
- ▶ A rede social Facebook contabilizou em junho de 2016 uma média de 1.13 bilhão de usuários, 2.5 bilhões de compartilhamentos e 2.7 bilhões de “curtidas” diariamente.
- ▶ A rede social de compartilhamento de fotos Instagram recebe atualmente cerca de 80 milhões de fotos por dia.
- ▶ Em 2013, a plataforma de blogs WordPress relatou a quantidade de 42 milhões de comentários por mês, entre os 3.6 bilhões de páginas existentes na plataforma.

# Introdução a Big Data



- VARIEDADE

- BD relacional é o modelo é o modelo de armazenamento de dados mais usado nos últimos 40 anos pelas corporações.
  - Dados são armazenados em formato de tabelas, estrutura previamente definida.
  - variedade de dados em e-mails, redes sociais, fotografias, áudios, telefones e cartões de crédito
- Variedade de informações:



- Dentre essas 3 categorias, estima-se que até 90% de todos os dados no mundo estão a forma de dados não estruturados.



# Introdução a Big Data

- VELOCIDADE

- Velocidade com que os dados são coletados, analisados e utilizados
- Muitas informações só se tornam realmente úteis se analisadas em tempo real.



# Introdução a Big Data

- **Definição de Big Data criada pela Gartner:**

“Big Data faz referência não somente ao volume, mas também à variedade e à velocidade de dados, necessitando de estratégias inovadoras e rentáveis para extração de valor dos dados e aumento da percepção”.

- **Definição de Big Data - SAS**

“Big data é um termo que descreve o grande volume de dados — tanto estruturados quanto não-estruturados — que sobrecarrega as empresas diariamente. Mas não é a quantidade de dados disponíveis que importa; é o que as organizações fazem com eles. O big data pode ser analisado para obter insights que levam a decisões melhores e ações estratégicas de negócio”.



# Introdução a Big Data

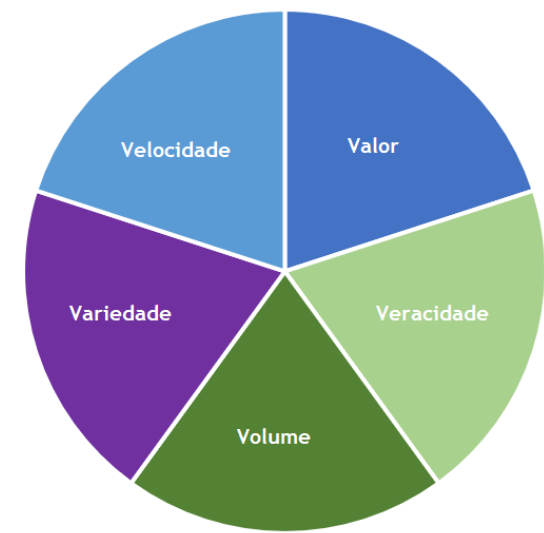
- Alguns pesquisadores adotaram os 5Vs: Valor e Veracidade

- VALOR

- Atributo que faz referência ao quão valioso e significativo um dado pode ser em uma solução
- O uso do Big Data torna-se imprescindível quando há foco na geração de valor, convertendo essa base de dados em informações relevantes, que serão aplicadas nas estratégias de negócio.

- VERACIDADE

- Está relacionado à confiabilidade dos dados.
- Realização de filtros e extrair das informações (que são muitas!)o que realmente agrega valor.



# Introdução a Big Data

- Existem ainda outros V's, são eles Visibilidade, Variabilidade e Visualização que complementam o conceito.
- VISIBILIDADE. É a relevância dos dados. A organização está ciente de todos os dados que ele gera? Estes poderiam ser (aparentemente) registros de dados inconsequentes.
- VARIABILIDADE (e complexidade). A SAS apresentou variabilidade (e complexidade) como duas dimensões adicionais para Big Data. Variabilidade refere-se à variação nas taxas de fluxo de dados. Muitas vezes, a velocidade de Big Data não é consistente e tem picos e depressões periódicas. Complexidade refere-se ao fato de Big Data gerar ou receber informações através de uma multiplicidade de fontes. Isso impõe um desafio crucial: a necessidade de se conectar, integrar, limpar e transformar os dados recebidos de diferentes fontes.
- VISUALIZAÇÃO. É o modo como aqueles dados complexos serão representados ou apresentados

# Introdução a Big Data

- Os dados devem ser processados com ferramentas avançadas (análise e algoritmos) para revelar informações significativas.



Hadoop



Spark



Mongo DB



HBase



Cloudera



SAP



Cassandra



Pentaho

# Introdução a Big Data

- Mitos
- Big Data engloba somente dados não estruturados
- Big Data refere-se somente a soluções com petabytes
- Big Data é aplicado somente às empresas do Vale do Silício
- Big Data é aplicado somente em grandes empresas
- Big Data requer o uso de dados externos
- As tecnologias de Big Data já estão bem estabelecidas

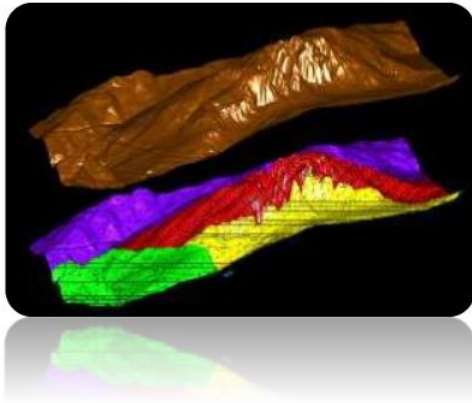
# Introdução a Big Data

- Exemplos

Área	Onde Big Data está sendo aplicado
Cuidados da saúde e medicina	Monitoramento de pacientes em tempo real; Análise de dados de redes sociais para descobertas de pandemias; Análise de padrões de doenças; Extração de informação em imagens médicas; Descoberta e desenvolvimento de medicamentos; Análise de dados genéticos.
Serviços financeiros	Análise de risco; Detecção de fraude; Programas de lealdade; Venda cruzada.
Setor público	Digitalização dos dados; Detecção de fraude e ameaças; Vigilância por vídeo; Manutenção preventiva de veículos públicos; Otimização de rotas no transporte público.
Telecomunicação	Análise de registro de chamadas; Alocação de banda em tempo real; Desenvolvimento de novos produtos; Planejamento da rede; Análise de <i>churn</i> ; Gerenciamento de fraude; Monitoramento de equipamentos.
Varejo	Análise de sentimento; Segmentação de mercado e cliente; Marketing personalizado; Previsão de demanda; Precificação dinâmica.

# Introdução a Big Data

- Exemplos



Geoestatística



Planejamento de Mina



Despacho de Mina



# Introdução a Big Data



**Data  
Science**

# Referencias



cloudera



Gartner®