

CENTRO UNIVERSITÁRIO ESTÁCIO

Polo Taguatinga Sul – DF

Princípios de Big Data

Nome do(s) discente(s):

Bruno Ribeiro Ataides Braga – Matrícula: 202104361343

Erick Assunção Montalvão – Matrícula: 202309419203

Luís Fernando Santos Souza – Matrícula: 202302462936

Nome do(a) professor(a) orientador(a):

José Miranda

2024

Taguatinga - DF

Resumo detalhado dos 5 V's com exemplos práticos:

1. Volume

O volume refere-se à grande quantidade de dados gerados todos os dias. Com o aumento do uso de mídias sociais, dispositivos IoT, entre outros sistemas. O volume de dados gerados é impressionante:

– Exemplo: Redes sociais como Facebook e Instagram geram petabytes de dados todos os dias com fotos, vídeos, mensagens, curtidas, entre outros. O armazenamento e o processamento eficiente deste volume são essenciais para gerar insights.

2. Velocidade

A velocidade refere-se à rapidez com que os dados são gerados, processados e analisados. Com o uso cada vez maior de dados e informações em tempo real, como transmissão de vídeo ao vivo ou transações financeiras, é um fator crítico.

– Exemplo: Plataformas e-commerce monitoram dados em tempo real para ajustar os preços, reduzir a fraude e otimizar a experiência do usuário, dependendo de ações a cada segundo.

3. Variedade

Como citado, os dados vêm de várias fontes e formatos, por exemplo, texto, áudio, vídeo, e dados próprios estruturados e não-estruturados. A variedade torna o processamento um desafio.

– Exemplo: em uma fábrica de carros conectados, os dados podem vir dos sensores (todos são numéricos), câmeras (vídeos), gerenciamento de sistemas híbridos (dados estruturados), então o processamento é diferente.

4. Veracity. Esse princípio trata da confiabilidade dos dados e se refere a dados precisos, pois dados imprecisos ou incorretos levam a decisões precárias.

Exemplo: empresas de saúde coletam muitos dados de pacientes, mas é importante que a informação seja precisa e validada para que não ocorra um diagnóstico incorreto ou tratamento inadequado.

5. Valor. Esse princípio está vinculado ao conceito de informações que podem ser utilizados, e não tem relevância até que os dados sejam processados e se transformem em conhecimento. Exemplo: A análise de dados no campo do marketing ajuda a identificar os padrões de compra dos consumidores, o que ajuda a tornar a campanha de publicidade mais local e fazer com que a quantidade de vendas aumente. Com o crescimento exponencial de nosso diário tudo, a escalabilidade e a arquitetura distribuída são necessárias para processar grande e democratizar. Esse é um sistema distribuído usado em plataformas como o Hadoop e o Spark, que processam enormes quantidades de dados por várias partes do servidor.