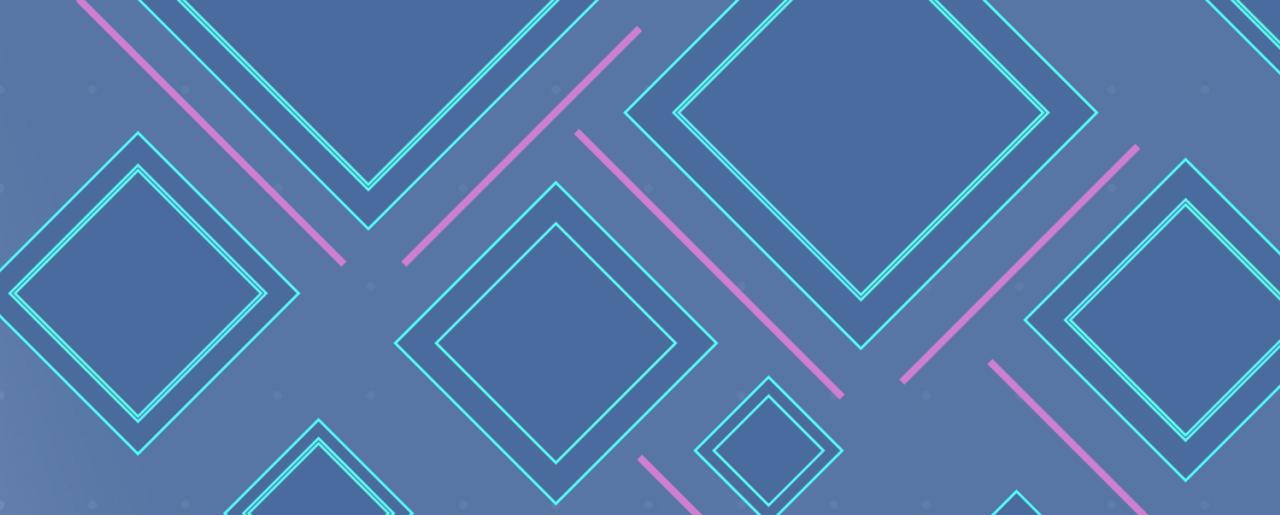


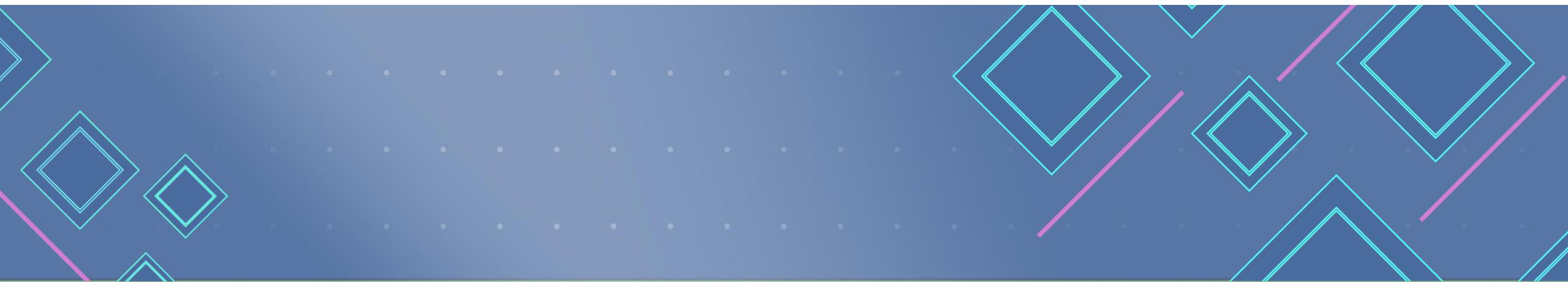


PUC Minas



BIG Data

Diego Roberto Gonçalves de Pontes



Introdução

- O processamento de volumes massivos de dados, também conhecido como Big Data, envolve a coleta, armazenamento, análise e visualização de grandes quantidades de dados gerados continuamente por diversas fontes. Essas fontes incluem transações comerciais, interações em redes sociais, sensores IoT (Internet das Coisas), dispositivos móveis, entre outros.

Características do Big Data

- Volume;
- Velocidade;
- Variedade;
- Veracidade; e
- Valor.

Tecnologias e Ferramentas

- Armazenamento e Gestão de Dados
 - Hadoop
- Processamento e Análise de Dados
 - MapReduce
 - Apache Spark
- Banco de Dados NoSQL
- Visualização de Dados
 - Tableau
 - Power BI

Categoria de Dados

- Dados Estruturados
- Dados Semiestruturados
- Dados Não Estruturados

Dados Estruturados

- Dados estruturados são informações organizadas em um formato predefinido e fixo, como tabelas de bancos de dados, facilitando sua consulta, análise e processamento.
- Um exemplo de dados estruturados é uma tabela de clientes em um banco de dados relacional, onde cada linha representa um cliente e cada coluna armazena atributos como nome, endereço e telefone.

Dados Semi Estruturados

- Dados semiestruturados são informações que não seguem um esquema rígido como os dados estruturados, mas possuem alguma organização que facilita a interpretação e análise.
- Um exemplo de dados semiestruturados é um documento XML ou JSON, onde os dados são organizados em hierarquias e etiquetas, permitindo flexibilidade na forma como as informações são armazenadas e acessadas.

Dados Não Estruturados

- Dados não estruturados são informações que não possuem uma organização ou esquema predefinido, tornando sua análise e processamento mais complexos. Eles podem ser de diversos tipos e formatos, como texto livre, imagens, vídeos, e-mails, posts em redes sociais e documentos em PDF.
- Um exemplo de dados não estruturados é um conjunto de e-mails, onde cada mensagem pode variar amplamente em termos de conteúdo, formato e estrutura, dificultando a extração de informações consistentes e a realização de análises automatizadas.



PUC Minas