

Documento Técnico de Desenvolvimento

1. Introdução

- Objetivo do Documento:

Este documento visa descrever as especificações técnicas e o processo de desenvolvimento do sistema de visualização de dados para análise de Big Data. Ele abrange a arquitetura do sistema, as tecnologias e ferramentas utilizadas, e as práticas de desenvolvimento adotadas.

- Escopo do Projeto:

O sistema de visualização de dados tem como objetivo fornecer representações gráficas de dados para facilitar a análise e a tomada de decisão. Ele será desenvolvido utilizando tecnologias de Big Data e Python.

2. Arquitetura do Sistema

- Visão Geral da Arquitetura:

O sistema será composto por três camadas principais:

1. Camada de Coleta de Dados: Responsável pela aquisição e armazenamento dos dados.

2. Camada de Processamento: Onde os dados serão limpos, transformados e analisados.

3. Camada de Visualização: Responsável pela criação de gráficos e relatórios a partir dos dados processados.

- Diagrama de Arquitetura:

- Componentes e Tecnologias Utilizadas:

- Coleta de Dados: APIs, Web Scraping, Banco de Dados

- Processamento de Dados: Python (pandas, NumPy, PySpark)

- **Visualização de Dados: Python (matplotlib, seaborn, Plotly)**

3. Tecnologias e Ferramentas

- **Python: Linguagem principal para desenvolvimento, utilizada devido à sua robustez em análise de dados.**

- **Bibliotecas Python:**

- **pandas: Manipulação e análise de dados**
- **NumPy: Operações matemáticas e numéricas**
- **PySpark: Processamento distribuído de grandes volumes de dados**
- **matplotlib e seaborn: Criação de gráficos e visualizações estáticas**
- **Plotly: Criação de visualizações interativas**

- **Banco de Dados:**

- **PostgreSQL: Armazenamento e consulta de dados estruturados**
- **NoSQL (opcional): MongoDB, para dados não estruturados**

- **Ferramentas de Desenvolvimento:**

- **Git: Controle de versão**
- **Jupyter Notebook: Ambiente de desenvolvimento interativo**
- **Docker: Contêineres para ambientes consistentes**

4. Metodologia de Desenvolvimento

- **Desenvolvimento Ágil:**

Utilização de metodologias ágeis para garantir a flexibilidade e a adaptação às mudanças de requisitos. Sprints semanais para revisão e ajustes.

- **Controle de Versão:**

Utilização do Git para versionamento do código e colaboração entre desenvolvedores.

Testes e Validação:

- **Testes Unitários:** Testar individualmente cada módulo do sistema.
- **Testes de Integração:** Garantir que os módulos funcionem corretamente em conjunto.
- **Testes de Aceitação:** Validar com a comunidade se o sistema atende aos requisitos definidos.

5. Processo de Implementação

- Coleta de Dados:

Configuração de scripts para coleta automática e periódica dos dados necessários.

- Processamento e Análise:

Desenvolvimento de pipelines para limpeza, transformação e análise dos dados.

- Desenvolvimento de Visualizações:

Criação de gráficos e dashboards interativos baseados nas análises realizadas.

- Documentação e Treinamento:

Disponibilização de documentação técnica e materiais de treinamento para os usuários finais.

6. Conclusão

- Resumo:

Este documento técnico fornece uma visão abrangente do desenvolvimento do sistema de visualização de dados, abordando a arquitetura, as tecnologias utilizadas, e a metodologia de desenvolvimento. O objetivo é assegurar que todas as etapas do projeto sejam bem planejadas e executadas de forma a atender às necessidades da comunidade.

- Referências:

- Documentos de especificação de requisitos**
- Guias e tutoriais das tecnologias utilizadas**
- Repositórios de código e documentação de ferramentas**