

Copilot

Perfeito! Vamos detalhar o plano do projeto com marcos a cada 3 dias, completando em 30 dias. Aqui está a nova estrutura:

Plano do Projeto de Análise de Dados e Engenharia de Dados

1. Objetivo do Projeto

- **Análise de Dados:** Realizar análises exploratórias e avançadas.
- **Pipelines de Dados:** Criar e gerenciar pipelines de dados eficientes.
- **MLOps:** Implementar práticas de MLOps para automação e monitoramento de modelos de machine learning.
- **Machine Learning e IA:** Desenvolver e treinar modelos de machine learning e inteligência artificial.

2. Frameworks e Ferramentas

- Apache Spark, Delta Table, AWS, Kubernetes, Airflow, dbt, Databricks, Grafana, Jupyter Notebook, VS Code, WSL.

3. Softwares Necessários

- WSL, VS Code, Jupyter Notebook, Docker, PostgreSQL, Git, Airflow, dbt, Grafana.

4. Volumetria de Dados

- **Pequenos dados para testes:** Dados gerados artificialmente ou extraídos de APIs públicas.

5. Fontes de Dados

- Banco de Dados PostgreSQL, Arquivos CSV e JSON, APIs.

6. Tarefas Semestrais

- Atualização de Dados, Manutenção de Pipelines, Desempenho e Observabilidade, Histórico e Projeção.

7. Duração do Projeto (Marcos a cada 3 dias)

Dias 1 a 3: Instalação e Configuração

- Instalar WSL, VS Code, Docker, PostgreSQL, e outras ferramentas necessárias.
- Configurar ambientes de desenvolvimento e contêineres Docker.

Dias 4 a 6: Criação dos Pipelines usando Airflow e dbt

- **Airflow:**
 - Criar repositório no VS Code.
 - Estrutura de pastas:
 - dags/: Contém os DAGs (Directed Acyclic Graphs).
 - plugins/: Plugins personalizados.
 - config/: Arquivos de configuração.
 - scripts/: Scripts auxiliares.
 - Criar um DAG simples para ingestão de dados.
- **dbt:**
 - Criar repositório no VS Code.

- Estrutura de pastas:
 - models/: Modelos SQL.
 - seeds/: Dados de sementes.
 - snapshots/: Snapshots de dados.
 - tests/: Testes de qualidade de dados.
- Configurar dbt para se conectar ao PostgreSQL.
- Criar um modelo simples de transformação de dados.

Dias 7 a 9: Criar Modelos de Machine Learning

- **Preparação dos Dados:**
 - Carregar dados no Jupyter Notebook.
 - Realizar limpeza e pré-processamento dos dados.
- **Desenvolvimento do Modelo:**
 - Escolher um algoritmo de machine learning (e.g., regressão linear, árvore de decisão).
 - Treinar o modelo com os dados preparados.
 - Avaliar o desempenho do modelo.
- **MLOps:**
 - Integrar o modelo treinado com Airflow para automação.
 - Configurar monitoramento do modelo com Grafana.

Dias 10 a 12: Implementação de Pipelines de Dados

- Criar pipelines de dados mais complexos com Airflow.
- Implementar transformações de dados avançadas com dbt.
- Configurar alertas e monitoramento de pipelines.

Dias 13 a 15: Integração com AWS e Kubernetes

- Configurar serviços AWS para armazenamento e processamento de dados.
- Implementar orquestração de contêineres com Kubernetes.
- Desplegar pipelines e modelos em um cluster Kubernetes.

Dias 16 a 18: Monitoramento e Observabilidade

- Configurar Grafana para monitoramento de pipelines e modelos.
- Criar dashboards para visualização de métricas de desempenho.
- Implementar alertas para falhas e anomalias.

Dias 19 a 21: Manutenção e Atualização de Dados

- Atualizar dados provenientes de APIs.
- Realizar manutenção preventiva dos pipelines.
- Ajustar modelos de machine learning com novos dados.

Dias 22 a 24: Análise de Desempenho e Otimização

- Analisar o desempenho dos pipelines e modelos.
- Identificar gargalos e otimizar processos.
- Documentar melhorias e ajustes realizados.

Dias 25 a 27: Documentação e Relatórios

- Documentar todo o processo de desenvolvimento.
- Criar relatórios detalhados sobre o desempenho dos pipelines e modelos.
- Preparar documentação para futuros desenvolvedores.

Dias 28 a 30: Revisão Final e Entrega

- Revisar todo o projeto para garantir que todos os requisitos foram atendidos.
- Realizar testes finais e ajustes necessários.
- Subir o projeto para o GitHub e compartilhar com a comunidade.

Próximos Passos

1. **Dias 1 a 3:** Vamos começar com a instalação e configuração do ambiente. Pronto para iniciar?