

PROJETO APLICADO DE PIPELINE DE DADOS - FASE 04

Disciplina: Pipeline de Dados - IA-CD

VISÃO GERAL DA FASE 04

Esta é a fase final do projeto! Vocês irão refatorar o código para produção, processar o projeto completo, concluir toda a documentação, registrar as atividades de cada membro do grupo e suas funções, e gravar um vídeo de apresentação onde o pipeline será executado e cada membro falará sobre o que fez.

O que vocês farão nesta fase:

- **Refatorar** o código em funções e classes reutilizáveis
- **Processar** o projeto completo de ponta a ponta
- **Concluir** toda a documentação do projeto
- **Registrar** as atividades de cada membro do grupo e suas funções
- **Gravar** um vídeo de apresentação com execução do pipeline
- **Incluir dashboards** (opcional, a critério do grupo)

ENTREGÁVEIS DA FASE 04

ENTREGÁVEL 1: Refatoração do Código

Organizem todo o código desenvolvido nas fases anteriores em módulos reutilizáveis e bem estruturados. O objetivo é ter um código limpo, organizado e pronto para produção.

ENTREGÁVEL 2: Processamento do Projeto Completo

Execute o pipeline completo de ponta a ponta, desde a extração até a carga final. Garanta que todas as etapas funcionem corretamente e que o fluxo esteja documentado.

Checklist de Processamento Completo

- ☐ **Extração (Bronze Layer)**
 - Dados extraídos da fonte original
 - Dados salvos na camada Bronze (Parquet ou CSV)
- ☐ **Transformação (Silver Layer)**
 - Duplicatas removidas
 - Valores nulos tratados
 - Tipos de dados corrigidos
 - Validações de qualidade aplicadas
 - Dados salvos na camada Silver
- ☐ **Agregação (Gold Layer)**
 - Métricas calculadas
 - Agregações realizadas
 - Dados prontos para análise
 - Dados salvos na camada Gold
- ☐ **Carga no Banco de Dados**
 - Dados carregados no PostgreSQL/MySQL ou outro da escolha do grupo
 - Tabelas criadas corretamente
 - Relacionamentos estabelecidos
 - Índices criados (se necessário)
- ☐ **Orquestração (Airflow)**
 - DAG criada e funcionando
 - Tarefas executando na ordem correta
 - Monitoramento ativo
 - Logs sendo gerados

Documentação do Processamento

Incrementem os processos de execução na documentação existente, descrevendo:

- Ordem de execução das etapas
- Tempo de processamento de cada etapa
- Volume de dados processados

- Problemas encontrados e soluções aplicadas
- Métricas de qualidade final

ENTREGÁVEL 3: Dashboards de Visualização (Opcional)

Nota: A inclusão de dashboards é opcional e fica a critério do grupo. Se optarem por incluir, criem dashboards interativos para visualização dos dados processados em streamlit, seaborn, matplotlib, power bi ou outra ferramenta a critério do grupo.

ENTREGÁVEL 4: Documentação Completa

Concluem toda a documentação do projeto. Essa documentação deve ser completa e detalhada.

ENTREGÁVEL 5: Registro das Atividades dos Membros

Na documentação, incluam uma seção detalhando as atividades de cada membro do grupo e suas funções em cada fase do projeto.

Sugestão de como estruturar o registro das atividades dos membros

Contribuições dos Membros do Grupo

Membros do Grupo

1. [Nome Completo 1]
2. [Nome Completo 2]
3. [Nome Completo 3]
4. [Nome Completo 4]

FASE 01: Planejamento e Exploração

[Nome 1]

- Função Principal: [Ex: Líder de Projeto / Analista de Dados]
- Atividades Realizadas:
 - Escolha do tema e justificativa
 - Exploração inicial dos dados
 - Criação do notebook de análise exploratória
 - Documentação da proposta inicial
- Arquivos Criados/Modificados:
 - `notebooks/01_exploracao_inicial.ipynb`
 - `Projeto Aplicado - Fase 01.md`

[Nome 2]

- Função Principal: [Ex: Desenvolvedor de Pipeline]
- Atividades Realizadas:
 - [Listar atividades]
- Arquivos Criados/Modificados:
 - [Listar arquivos]

[Nome 3]

- Função Principal: [Ex: Especialista em Qualidade de Dados]
- Atividades Realizadas:
 - [Listar atividades]
- Arquivos Criados/Modificados:
 - [Listar arquivos]

[Nome 4]

- Função Principal: [Ex: Documentador]
- Atividades Realizadas:
 - [Listar atividades]
- Arquivos Criados/Modificados:
 - [Listar arquivos]

FASE 02: Pipeline ETL Básico

```
# [Nome 1]
- Função Principal: [Ex: Desenvolvedor Bronze/Silver]
- Atividades Realizadas:
  - Implementação da camada Bronze
  - Implementação da camada Silver
  - Tratamento de valores nulos
  - Remoção de duplicatas
- Arquivos Criados/Modificados:
  - `notebooks/01_bronze_layer.ipynb`
  - `notebooks/02_silver_layer.ipynb`

# [Nome 2]
- Função Principal: [Ex: Desenvolvedor Gold/Database]
- Atividades Realizadas:
  - [Listar atividades]
- Arquivos Criados/Modificados:
  - [Listar arquivos]

# [Nome 3]
- Função Principal: [Ex: Especialista em Qualidade]
- Atividades Realizadas:
  - [Listar atividades]
- Arquivos Criados/Modificados:
  - [Listar arquivos]

# [Nome 4]
- Função Principal: [Ex: Documentador]
- Atividades Realizadas:
  - [Listar atividades]
- Arquivos Criados/Modificados:
  - [Listar arquivos]
```

FASE 03: Big Data e Cloud

```
# [Nome 1]
- Função Principal: [Ex: Especialista em Spark]
- Atividades Realizadas:
  - Migração do código Pandas para Spark
  - Otimização das transformações
  - Configuração do ambiente Spark
- Arquivos Criados/Modificados:
  - `notebooks/01_spark_bronze_layer.ipynb`
  - `notebooks/02_spark_silver_layer.ipynb`

# [Nome 2]
- Função Principal: [Ex: Especialista em Banco de Dados]
- Atividades Realizadas:
  - Migração de SQLite para PostgreSQL
  - Criação do schema do banco
  - Otimização de queries
- Arquivos Criados/Modificados:
  - `notebooks/04_create_database_schema.ipynb`
  - `notebooks/05_load_to_postgres.ipynb`

# [Nome 3]
- Função Principal: [Ex: Especialista em Airflow]
- Atividades Realizadas:
  - Configuração do Airflow
  - Criação da DAG
  - Monitoramento do pipeline
```

- Arquivos Criados/Modificados:
 - `airflow/dags/pipeline_etl_dag.py`
 - `docker-compose.yaml`

[Nome 4]

- Função Principal: [Ex: Documentador]
- Atividades Realizadas:
 - [Listar atividades]
- Arquivos Criados/Modificados:
 - [Listar arquivos]

FASE 04: Refatoração e Finalização

[Nome 1]

- Função Principal: [Ex: Arquiteto de Software]
- Atividades Realizadas:
 - Refatoração do código em módulos
 - Criação da estrutura de pastas
 - Implementação do script principal (main.py)
- Arquivos Criados/Modificados:
 - `src/extract.py`
 - `src/transform.py`
 - `main.py`

[Nome 2]

- Função Principal: [Ex: Desenvolvedor de Dashboard]
- Atividades Realizadas:
 - Criação do dashboard Streamlit
 - Implementação de visualizações
 - Integração com banco de dados
- Arquivos Criados/Modificados:
 - `dashboard/app.py`

[Nome 3]

- Função Principal: [Ex: Documentador Técnico]
- Atividades Realizadas:
 - Criação do README completo
 - Documentação do código
 - Criação de guias de instalação
- Arquivos Criados/Modificados:
 - `README.md`
 - `INSTALACAO.md`
 - `PROCESSAMENTO.md`

[Nome 4]

- Função Principal: [Ex: Coordenador de Apresentação]
- Atividades Realizadas:
 - Organização do vídeo de apresentação
 - Edição do vídeo
 - Preparação da apresentação final
- Arquivos Criados/Modificados:
 - `video_apresentacao.mp4`
 - `slides_apresentacao.pdf`

Distribuição de Trabalho

- Desenvolvimento: XX%
- Documentação: XX%
- Testes: XX%
- Apresentação: XX%

Observações

[Qualquer observação adicional sobre a divisão de trabalho, desafios enfrentados, etc.]

ENTREGÁVEL 6: Vídeo de Apresentação

Grave um vídeo de apresentação do projeto onde:

1. O pipeline será executado de ponta a ponta
2. Cada membro do grupo terá um espaço para falar sobre o que fez
3. Todas as funcionalidades serão demonstradas

Roteiro Sugerido para o Vídeo

Duração Total: 8-12 minutos

1. Introdução

- Apresentação do grupo
- Problema de negócio resolvido
 - Objetivos do projeto
- **Membro responsável:** [Nome]

2. Arquitetura e Tecnologias

- Visão geral da arquitetura
 - Tecnologias utilizadas
 - Fluxo de dados
- **Membro responsável:** [Nome]

3. Demonstração do Pipeline - Bronze Layer

- Execução da extração
- Visualização dos dados brutos
- Explicação dos metadados
- **Membro responsável:** [Nome que trabalhou na Bronze]

4. Demonstração do Pipeline - Silver Layer

- Execução das transformações
- Visualização dos dados limpos
 - Métricas de qualidade
- **Membro responsável:** [Nome que trabalhou na Silver]

5. Demonstração do Pipeline - Gold Layer

- Execução das agregações
- Visualização das métricas
- Dados prontos para análise
- **Membro responsável:** [Nome que trabalhou na Gold]

6. Demonstração do Banco de Dados

- Visualização das tabelas
- Queries de exemplo
- Relacionamentos
- **Membro responsável:** [Nome que trabalhou no banco]

7. Demonstração do Airflow

- Interface do Airflow
- Execução da DAG
- Monitoramento
- **Membro responsável:** [Nome que trabalhou no Airflow]

8. Dashboard (Opcional)

- Visualização do dashboard
- Interatividade
- Insights
- **Membro responsável:** [Nome que trabalhou no dashboard]

9. Conclusão e Aprendizados

- Resumo das conquistas
- Principais desafios

- Lições aprendidas
 - Próximos passos
- **Membros:** Todos participam

Dicas para Gravação

1. Preparação:

- Testem o pipeline antes de gravar
- Preparem um roteiro detalhado
- Ensaiem a apresentação

2. Qualidade do Vídeo:

- Use uma ferramenta de gravação de tela (OBS, Zoom, etc.)
- Garanta boa qualidade de áudio
- Use um microfone se possível

3. Demonstração:

- Mostrem o código sendo executado
- Expliquem o que está acontecendo em cada etapa
- Mostrem os resultados visuais

4. Participação:

- Cada membro deve falar sobre sua parte
- Seja claro e objetivo
- Mostre conhecimento técnico

5. Edição:

- Remova pausas longas
- Adicione legendas se necessário
- Inclua transições suaves

Formato de Entrega

- **Formato:** MP4 ou similar
- **Duração:** 8-12 minutos
- **Qualidade:** Mínimo 720p
- **Áudio:** Claro e sem ruídos
- **Upload:** Google Drive (se arquivo muito grande)

CRITÉRIOS DE AVALIAÇÃO

Refatoração de Código (25%)

- Código organizado em módulos
- Funções reutilizáveis
- Boas práticas de programação
- Código limpo e documentado

Processamento Completo (25%)

- Pipeline executando de ponta a ponta
- Todas as etapas funcionando
- Dados processados corretamente
- Airflow operacional

Documentação (20%)

- README completo e detalhado
- Código comentado
- Instruções claras de instalação
- Documentação do processamento

Registro de Atividades (15%)

- CONTRIBUICOES.md completo
- Atividades de cada membro registradas
- Funções claramente definidas
- Distribuição de trabalho documentada

Vídeo de Apresentação (15%)

- Pipeline executado no vídeo
- Cada membro participa

- Demonstração clara
 - Qualidade técnica do vídeo
-

DATA DE ENTREGA

Prazo: 27/11/2025

Formato de Entrega:

- Repositório GitHub/Google drive completo com todo o código
- Documentação completa
- Vídeo de apresentação (link ou arquivo)
- Reprodutibilidade (README.md com as instruções para rodar o projeto)

Itens Obrigatórios:

- Código refatorado e organizado
- Pipeline executando completamente
- Documentação completa
- Vídeo de apresentação com execução do pipeline e participação de todos

Itens Opcionais:

- Dashboard interativo
-

Parabéns por chegarem até aqui! Este projeto representa uma jornada completa de dados, desde a exploração até um pipeline de produção.

Boa sorte na fase final!