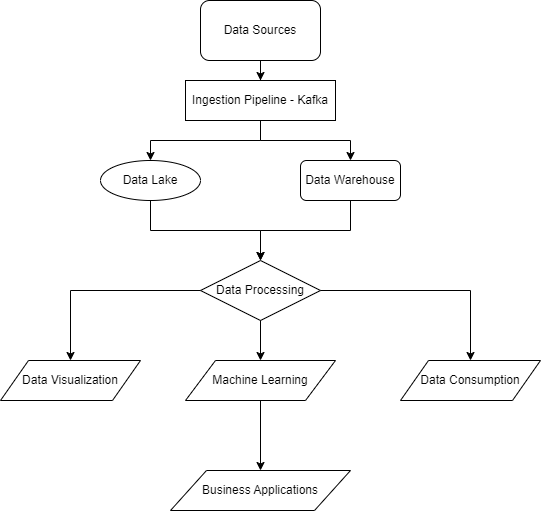
Atividade Final: RM351578 – Rodrigo de Feo Vieira

Arquitetura da Fundação de Dados em Nuvem:



1. Armazenamento de Dados:

Data Lake: Armazenamento centralizado para dados semi-estruturados e não estruturados.

Data Warehouse: Armazenamento para dados estruturados, permitindo consultas analíticas complexas.

Cloud Storage: Armazenamento escalável e acessível para backups, arquivos e dados de baixa latência.

2. Processamento de Dados:

Apache Spark ou AWS EMR: Para processamento distribuído de grandes volumes de dados.

Apache Kafka ou AWS Kinesis: Para ingestão e streaming de dados em tempo real.

3. Orquestração de Dados:

Apache Airflow ou AWS Step Functions: Orquestração de pipelines de dados e agendamento de tarefas.

4. Serviços de Banco de Dados:

Amazon RDS ou Azure SQL Database: Para dados transacionais e relacionais.

Amazon Redshift ou Azure Synapse Analytics: Para análise de dados em larga escala.

Serviços NoSQL (por exemplo, DynamoDB ou Cosmos DB): Para dados semi-estruturados e não estruturados.

5. Segurança e Governança:

AWS IAM ou Azure Active Directory: Controle de acesso e autenticação.

Encriptação de dados em repouso e em trânsito: Para garantir a segurança dos dados.

Auditoria e Monitoramento: Ferramentas para monitorar atividades e garantir conformidade.

6. Data Visualization e Machine Learning:

Power BI, Tableau ou AWS QuickSight: Para visualização de dados e painéis interativos.

AWS SageMaker ou Azure Machine Learning: Para construir, treinar e implantar modelos de machine learning.

Principais Desafios e Ações Recomendadas:

Desafios:

Segurança dos Dados: Garantir que os dados estejam protegidos contra ameaças cibernéticas.

Integração de Dados: Lidar com a variedade de fontes de dados e integrá-las de forma eficiente.

Gerenciamento de Custos: Controlar os custos associados ao armazenamento e processamento de dados na nuvem.

Ações Recomendadas:

Implementar Políticas de Segurança Robustas: Criptografar dados, implementar controle de acesso e monitorar atividades suspeitas.

Padronizar Formatos de Dados: Utilizar formatos como JSON ou Parquet para facilitar a integração e processamento de dados.

Monitorar e Otimizar Custos: Utilizar serviços de nuvem com base no uso real e ajustar recursos conforme necessário para otimizar os custos.

Ações recomendações para Ingestão de Dados:

Utilize ferramentas de ingestão de dados em tempo real, como Kafka ou Kinesis, para capturar dados de fontes em tempo real.

Implemente pipelines de ingestão de dados robustos e escaláveis para lidar com grandes volumes de dados.

Valide e normalize os dados durante o processo de ingestão para garantir consistência e qualidade.

Pré-requisitos necessarios antes de disponibilizar dados para consumo:

Garanta que os dados estejam limpos, validados e devidamente documentados.

Implemente políticas de acesso baseadas em funções para controlar quem pode acessar e modificar os dados.

Forneça documentação detalhada sobre a estrutura dos dados, seu significado e qualquer transformação aplicada a eles.