

## Fase 1: Fundamentos de Programação e Banco de Dados

Linguagens de Programação (Python/SQL):

Python: Foque em manipulação de dados (Pandas, Numpy) e automação.

SQL: Pratique consultas complexas, joins, subconsultas e criação de views.

Exercícios Práticos:

Plataforma: LeetCode, HackerRank, Codecademy.

Fundamentos de Bancos de Dados:

Modelagem de Dados: Entenda normalização, desnormalização, chaves primárias e estrangeiras.

SGBDs Relacionais: Trabalhe com PostgreSQL, MySQL, SQL Server.

SGBDs NoSQL: Conheça MongoDB, Cassandra.

Versionamento com Git:

Versione seus projetos de dados no GitHub ou GitLab.

## Fase 2: Armazenamento e Processamento de Dados

Sistemas de Armazenamento de Dados:

Data Lakes: Explore ferramentas como Apache Hadoop, AWS S3 e Azure Data Lake.

Data Warehousing: Estude Redshift, BigQuery, Snowflake, e conceitos de OLAP/OLTP.

ETL/ELT:

Entenda o fluxo de dados desde a extração até a transformação e carga.

Ferramentas: Apache Airflow, Talend, dbt.

Processamento de Dados em Lote:

Apache Spark: Estude como processar grandes volumes de dados em paralelo.

Hadoop: Conheça o ecossistema Hadoop (MapReduce, HDFS).

## Fase 3: Arquiteturas de Dados e Integrações

Arquitetura de Dados:

Lambda e Kappa Architecture: Compare arquiteturas de processamento de dados.

Event-Driven Architecture: Explore ferramentas como Kafka para ingestão de dados em tempo real.

DataOps e Automação:

Pratique automação de pipelines de dados com CI/CD (Jenkins, GitLab CI).

Data Lakes vs Data Warehouses:

Aprofunde-se no Apache Iceberg, Delta Lake, e outras ferramentas de gerenciamento de lakes open-source.

## Fase 4: Processamento de Dados em Tempo Real

Streaming de Dados:

Ferramentas como Kafka, Apache Flink e Spark Streaming.

Pratique ingestão de dados em tempo real.

APIs e Integração de Sistemas:

Construa pipelines que integram APIs REST para ingestão e monitoramento de dados em tempo real.

## Fase 5: Escalabilidade e Segurança de Dados

Escalabilidade de Pipelines de Dados:

Sharding e Partitioning: Aprenda a distribuir dados entre servidores.

Balanceamento de Carga e Cache: Utilize Redis, Memcached.

Segurança de Dados:

GDPR e LGPD: Entenda as regulamentações de proteção de dados.

Ferramentas de criptografia, autenticação e auditoria de acessos.

## Fase 6: Projetos Práticos

### Projetos Reais:

Desenvolva pipelines de dados completos (de ingestão à análise) usando ferramentas aprendidas.

Use datasets públicos ou crie APIs que geram dados em tempo real para prática.

### Portfólio no GitHub:

Documente e publique seus projetos com repositórios bem organizados.

### Recursos Complementares:

#### Livros:

"Designing Data-Intensive Applications" (Martin Kleppmann).

"The Data Warehouse Toolkit" (Ralph Kimball).