Data Engineering Course Glossary

Basic Data Engineering

- Basic Concepts

   - Introduction & Practical Application of Data

   - What is Big Data

   - Structured vs Unstructured Data

   - Batch vs Stream Processing

   - File Formats, Schema & Compression

- Data Modeling

   - Dimensional Data Modeling

   - Star Schema

   - Snowflake Schema

   - Slowly Changing Dimensions (SCD)

   - Additive vs Non-Additive Facts

- Data Warehousing

   - Medallion Architecture

   - Data Lake vs Data Warehouse vs Data Marts

   - Data Lakehouse

- SQL

   - Basic Operations

   - Joins & Aggregations

   - Views & Materialized Views

   - DDL Operations

   - Normalization

- ETL with Spark

   - Introduction to Spark

   - Basic Programming

   - Batch vs Streaming

   - Insert vs Upsert

   - Change Data Capture - CDC

Data Engineering Advanced

- Spark Advanced Features

Data Engineering Course Glossary

  - Transformations

  - Joins and Optimizations

  - Partitioning and File Formats

  - Orchestration with Apache Airflow

- Lakehouse Architecture

  - Introduction

  - Hive Table & Hive Metastore

  - Presto & Trino & Databricks SQL

  - Open Table Formats: Delta Lake, Iceberg

Azure Cloud

- Azure SQL Database

  - Introduction - Database as a Service (DaaS)

  - Creating & Using SQL Server

  - Creating SQL Database Service

  - Managed Instance vs Single vs Elastic Pool

  - Scalability & Availability

  - Migrations

- Azure Storage Service

  - Features of Azure Storage Service

  - Introduction to Blob Storage Service

  - Blob Storage Architecture

  - Azure Storage Access Tiers

  - Understanding Data Replication

  - Versioning & Lifecycle

- Azure Data Factory

  - Introduction to ADF

  - ADF Key Components: Pipeline, Activity, Triggers, Data Flows, Integration Runtime

  - Activity 1: Copy data from Blobstore to Blobstore, Copy data to Azure SQL Database, Working with Copy, Lookup, Filter act

  - Activity 2: Triggering Pipelines Manually, On Schedule, Monitoring Pipelines

- Azure Databricks

  - Introduction

  - Databricks Runtime & Cluster Management

Data Engineering Course Glossary

- Databricks Main Concepts

- Databricks Unity Catalog

- Databricks SQL & SQL Serverless

- Activity 1: Reading and Writing data from Azure Blob

- Azure Synapse Analytics

- Introduction

- Creating Synapse Workspace

- Understanding Synapse SQL Pool and SQL On-Demand

- Using Synapse Pipelines and Copy Data Tool

- Orchestrating with Synapse Pipelines