

This is **36-week training curriculum** designed to build **intermediate-level Data Engineering skills** so developers can justify 2-3 years of experience in the role. The curriculum focuses on practical and theoretical aspects.

Week	Topic	Key Learning Objectives
1	Introduction to Data Engineering	Overview of the role, responsibilities, and importance of data engineering. Key concepts like ETL, ELT, data pipelines, and workflow management.
2	Python for Data Engineering	Mastering Python fundamentals, libraries (Pandas, NumPy), and file handling for data manipulation.
3	SQL Fundamentals	Writing complex SQL queries, including joins, subqueries, aggregations, and window functions.
4	Databases and Data Modeling	Introduction to RDBMS (PostgreSQL, MySQL) and NoSQL databases (MongoDB). Basics of data modeling: Star and Snowflake schemas.
5	Data Formats and File Systems	Understanding CSV, JSON, Parquet, and Avro. Using distributed file systems like HDFS and S3.
6	Data Warehousing Concepts	Concepts of OLAP vs. OLTP, and tools like Snowflake and Amazon Redshift.
7	Apache Spark Basics	Introduction to Apache Spark: RDDs, DataFrames, and SparkSQL. Getting hands-on with PySpark.
8	Workflow Orchestration with Airflow	Creating DAGs, scheduling, monitoring, and optimizing workflows using Apache Airflow.
9	Big Data Ecosystem	Introduction to Hadoop, HDFS, and the MapReduce paradigm. Basics of Hive and its use in data querying.
10	Streaming Data Basics	Introduction to real-time data streaming using Apache Kafka. Basics of Spark Streaming and its use cases.
11	Cloud Platforms Overview	Using cloud platforms like AWS, Azure, and GCP for storage (S3, Azure Blob) and ETL pipelines.
12	Data Security and Governance	Understanding data security (encryption, access control) and compliance (GDPR). Tools for governance like Collibra and Alation.
13	Advanced SQL	Writing advanced SQL queries, query optimization, execution plans, and index management.
14	APIs and Data Integration	Using REST APIs for data ingestion and GraphQL basics. Integrating APIs into ETL workflows.
15	Data Lake Architecture	Designing and implementing data lakes. Understanding data lake partitioning and integration with data warehouses.
16	Data Quality and Testing	Ensuring data accuracy, completeness, and consistency using tools like Great Expectations. Writing test cases for data pipelines.
17	Data Engineering in Production	Monitoring, logging, and alerting for data pipelines. Handling failures and maintaining SLAs for data downtime.

NavAdr Technologies Solutions



A-04, Street of Europe, Hinjewadi-Phase1

Pune-411057 ,Ph: +91-8857019010

Email : training@navadr.com

Week	Topic	Key Learning Objectives
18	Performance Tuning in Spark and Databases	Optimizing Spark applications, database queries, and overall ETL pipeline performance.
19	Machine Learning Integration	Integrating ML workflows into data pipelines. Using Spark for feature engineering and model deployment.
20	Serverless Data Engineering	Building serverless ETL pipelines using AWS Lambda, Azure Functions, or Google Cloud Functions.
21	Version Control and CI/CD	Using Git for version control. Building CI/CD pipelines for automated deployment of data engineering workflows.
22	Data Observability and Lineage	Implementing data observability tools (Monte Carlo, Datafold). Ensuring traceability and building lineage workflows.
23	Real-World Data Engineering Use Cases	Hands-on implementation of batch and streaming pipelines. Working on real-world data engineering challenges.
24	Emerging Trends	Exploring data mesh, generative AI in data engineering, and preparing for certifications like Databricks, Snowflake, or GCP.
25-36	Live Project	3 Months Live Project Experience