



DURATION
100 Days



100+ Hours
Live Training



Weekend
Assignments



Daily Practice
Questions



Real-Time
Industry Projects

FEES STRUCTURE

Online : ₹40,000
Offline : ₹50,000

ELIGIBILITY

- Freshers
- Software Engineers
- ETL Developers
- Database Developers
- BI Developers
- Test Engineers
- Data Analysts
- Professionals with 0 – 5 Years Experience

TOOLS COVERED

-  SQL Server, PostgreSQL
-  Python
-  Pandas, NumPy
-  PySpark, Apache Spark
-  Azure Data Factory (ADF)
-  Azure Databricks
-  Azure Data Lake Gen2
-  Azure Synapse Analytics
-  Microsoft Fabric
-  Delta Lake
-  Apache Kafka
-  Apache Airflow
-  Git & GitHub
-  Docker
-  Azure DevOps
-  Power BI
-  dbt (Data Build Tool)
-  Great Expectations
-  REST APIs
-  AI Tools (Copilot, ChatGPT)

SYLLABUS – ARRANGED BY EXPERIENCE LEVEL

PHASE 1
FOUNDATION LEVEL
(0 – 1 YEAR EXPERIENCE / FRESHERS)

1. Data Engineering Foundation

- Introduction to Data Engineering
- Data Lifecycle, ETL vs ELT, OLTP vs OLAP
- Data Warehouse, Data Lake, Lakehouse
- Medallion Architecture, Career Path

2. SQL for Data Engineering

- SQL Basics, Joins, Functions
- CTE, Window Functions, CASE, COALESCE
- MERGE, UPSERT, Views, Indexes
- Stored Procedures, Transactions
- Query Optimization & Performance Tuning
- Hands-on SQL Scenarios

3. Python Programming

- Python Basics to Advanced
- OOP, Exception Handling, Logging
- File Handling, JSON, CSV, XML
- Modules, Virtual Environment, REST APIs
- Hands-on Python Programs

4. Pandas & NumPy

- DataFrames, Series, Read/Write Files
- Data Cleaning, Merge, Join, Pivot
- Aggregation, Group By, EDA
- Hands-on Data Analysis Projects

5. Linux & Git

- Linux Commands, Shell Scripting
- File System, Permissions, Cron Jobs
- GitHub, Branching, Merge, Pull Request
- Version Control Best Practices

Outcome: Build Strong Foundation in SQL, Python & Core Concepts

PHASE 2
INTERMEDIATE LEVEL
(1 – 2 YEARS EXPERIENCE)

6. Data Warehousing & Modeling

- Star & Snowflake Schema
- Fact & Dimension Tables
- SCD Types (1,2,3), Surrogate Keys
- Partitioning, Data Marts, Kimball Concepts

7. Azure Cloud Fundamentals

- Azure Portal, Resource Groups
- Storage Account, ADLS Gen2, Blob Storage
- Key Vault, Managed Identity, IAM
- Networking Basics, Cost Optimization

8. Azure Data Factory (ADF)

- Pipelines, Activities, Datasets, Linked Services
- Parameters, Variables, Lookup, ForEach
- Copy Activity, Mapping Data Flow
- Triggers, Scheduling, Incremental Load
- Error Handling, Monitoring
- Project: Enterprise ETL Pipeline

9. Azure Databricks & PySpark

- PySpark (RDD, DataFrame, SQL)
- Transformations, Actions, Joins
- Spark SQL, Aggregations, UDFs
- Databricks Workspace, Clusters, Jobs
- Secrets, Delta Live Tables, Unity Catalog
- Performance Optimization
- Mini Project

10. Delta Lake & Lakehouse

- Delta Architecture, ACID Transactions
- Bronze, Silver, Gold Layers
- Time Travel, Merge, Vacuum
- Optimize, Z-Ordering

Outcome: Build & Manage End-to-End Cloud Data Pipelines

PHASE 3
ADVANCED LEVEL
(3 – 5 YEARS EXPERIENCE)

11. Streaming Data Engineering (Kafka)

- Kafka Architecture & Concepts
- Producer, Consumer, Topics, Partitions
- Offset Management, Event Streaming
- Real-Time Data Processing

12. Workflow Orchestration (Airflow)

- DAGs, Operators, Scheduling
- Dependencies, Sensors
- Monitoring, Logging
- Production Grade Pipelines

13. Data Quality & DevOps

- Great Expectations, Data Validation
- Data Profiling, Reconciliation
- Unit Testing, Integration Testing
- CI/CD, Azure DevOps, Release Pipelines
- Docker, Environment Management

14. Microsoft Fabric & dbt

- Microsoft Fabric (OneLake, Lakehouse, Warehouse, Pipelines, Notebooks)
- dbt Fundamentals, Models, Tests
- Data Transformation, Documentation

15. AI for Data Engineers

- GitHub Copilot, ChatGPT for SQL
- Prompt Engineering, Code Generation
- Pipeline Documentation, AI Debugging
- AI-assisted Development

16. Power BI for Data Engineers

- Connecting Azure Data, Databricks, SQL
- Data Modeling, Incremental Refresh
- Power BI Service, Deployment Pipeline

Outcome: Build Production-Ready Data Engineering Solutions

PHASE 4 – CAPSTONE PROJECT

End-to-End Enterprise Data Engineering Project

- REST APIs
- CSV Files
- SQL Database
- JSON Files



Azure Data Factory (Orchestration)



Azure Data Lake Gen2 (Storage)



Azure Databricks (PySpark)



Delta Lake (Bronze → Silver → Gold)



Microsoft Fabric (Lakehouse / Warehouse)



Power BI Dashboard

REAL-TIME INDUSTRY PROJECTS



Project 1
Retail Sales Data Pipeline



Project 2
Banking Transaction Processing



Project 3
Healthcare Claims Analytics



Project 4
E-Commerce Customer Analytics



Project 5
IoT Streaming Data using Kafka



Project 6
Microsoft Fabric Lakehouse Project

INTERVIEW & CAREER PREPARATION



300+ SQL
Interview Questions



200+ Python
Programs



100+ PySpark
Questions

ADF, Databricks, Kafka, Airflow Interview Qs



Resume & LinkedIn Optimization



GitHub Portfolio Creation



Mock Interviews & HR Prep



INSTITUTE ADDRESS

5th Floor, Office No. 516,
West Avenue, Bremen Chowk,
Opp. PMRDA, Aundh,
Pune – 411067



CONTACT*

9405048041 | 7385755621



EMAIL

mailto:cdsshai@gmail.com

**JOIN OUR
WHATSAPP GROUP!**

Scan the QR Code
to join our group

