

DATE ENGINEERING COURSE

Roadmap to Data Scientist

AI - Powered

Unleash the potential of your career as a Data Scientist

Session 1-2: Introduction to Data Engineering

- Data Warehouse Architecture
- Data Driven Platform
- OLTP | OLAP
- Data Modelling - Star and Snowflake Schema
- ETL flow
- DataMart

Session 3-4: Introduction to Microsoft Azure | Big Data | Data Lake

- How Azure influenced the Data warehousing
- Introduction to Azure Data Engineering
- Azure Free account creation
- DataLake | Blob | DataBase | Structured | Unstructured data
- Big Data to AzureDataLake Storage Gen2
- DataLake Gen2 storage Architecture - data Storage Explorer,
- Data storing pattern and Data archival strategy

Session 5-8: Azure Key Vault Engineering

- Azure Key Vault Understanding
- Security principals
- Secrets and Keys
- Key value Policy
- Scoped Credential
- SAS token
- Understanding of Different Azure Data tools
- ADLS2 connection approach to different data tools

DATE ENGINEERING COURSE

Roadmap to Data Scientist

AI - Powered

Unleash the potential of your career in Data Analytics

Session 9-10: Azure Data Factory - ADF

- Understanding of ADF and it's architecture
- ADF as ELT mechanism
- Understanding Pipeline | Data flow | Control flow
- Understanding of Linked Service | authorization mechanism
- Build our first basic pipeline with thorough understanding (incl. variable and parameter)
- Understanding of varoaus activities of ADF - Copy, Lookup, Store Procedure,set, IF- Else, Loop, Try-Catch, Web etc
- Understanding of Integration RunTime - Atoresolve, Self Hosted (On Premise)

Session 11-12: Hands-On End to End Data Pipeline building in ADF

- Build Data pipeline in ADF using important activities
- Understanding of Data Flow in detail in pipeline
- Triggers in ADF

Session 13-14: Azure Monitoring & Error logging

- Monitoring in ADF in detail
- Error Logging
- ADF Ingestion dynamic Framework in pipeline building

DATE ENGINEERING COURSE

Roadmap to Data Scientist

AI - Powered

Unleash the potential of your career in Data Analytics

Session 15-16: Introduction to SPARK & Databricks

- Understanding of Spark
- Spark in Databricks
- Stage | Jobs | Notebook
- Understanding PySpark and Databricks
- Our first notebook in Databricks.

Session 17-18: Understanding Azure Databricks (ADB) & PySpark

- Databricks Architecture (ADB)
- Play with Notebooks for data ingestion using PySpark
- PySpark - python + SQL
- DataBricks - Data ingestion from external file
- Datbricks - connect to ADLS2
- Understanding ADB utility

Session 19-20: Introduction to Synapse and its architecture

- Why using Synapse?
- Synapse vs Databricks
- Different Pools in Synapse
- Understanding of Serverless pool
- Data analysis in serverless pool in detail

DATE ENGINEERING COURSE

Roadmap to Data Scientist

AI - Powered

Unleash the potential of your career in Data Analytics

Session 21-22: Synapse and SPARK Pool (Microsoft Spark utility)

- Synapse Dedicated pool and pipeline
- End to End Data load pipeline to a data warehouse in Fact and Dimension modelling.
- Different kind of distribution
- Introduction of Spark Pool (understanding Microsoft spark utility)
- How spark pool helps in data analysis

Session 23-24: Conclusion

- Synapse spark pool conclusion using a notebook
- Logic App understanding and alert mechanism
- Azure Stream analytics and Event hub high-level understanding for data engineering

NOTE: This course outline provides a structured learning path from foundational skills to more advanced concepts, including hands-on projects and career development sessions. Each session can be further broken down into lectures, hands-on exercises, and discussions to enhance learning and engagement. This is an initiative taken by **CLIFAV TECHNOLOGIES PRIVATE LIMITED - We Make IT Happen**