



AZURE DATA ENGINEER



Programming & development

SERVING INFORMATION WORLDWIDE

Course Curriculum



AZURE DATA ENGINEER

Module 1 : Introduction to Azure Data Engineer

- Understanding of data world fundamental
- Basic familiarity with Azure Services

Module 2 : Design & implement Storage

Understand your data

- Structured data
- Semi-structured data
- Unstructured data

Azure data store services



Module 3 : Azure Data Store Services

Relational database management systems

- Azure SQL Database
- Azure Database for MySQL
- Azure Database for PostgreSQL
- Azure Database for MariaDB

Key / value stores

- Azure cosmos DB Tables API
- Azure Cache for Redis
- Azure Table Storage

Document Databases

- Azure Cosmos DB SQL API



- Graph Databases
- Azure cosmos DB Gremlin API
- SQL Server

Module 4 : (Design and implement storage) Azure Data Lake Gen 2

- Design an Azure Data Lake solution
- Recommend file types for storage
- Recommend file types for analytical queries
- Design for efficient querying
- Design a folder structure that represents the levels of data transformation

Module 5 : Azure Cosmos DB

- Structure of Azure Cosmos DB
- Database Partitions
- Shared and Dedicated throughput
- Create Azure Cosmos DB Account – core (SQL)

Module 6 : Azure Synapse Analytics

- Introduction to Azure Synapse Analytics
- Before Synapse Analytics
- Create a workspace
- Data integration
- Spark SQL Pool
- Dedicated SQL Pool
- SQL Demo
- Debugging
- Spark Code, Spark Demo

Module 7 : Azure SQL

- Introduction to Azure SQL



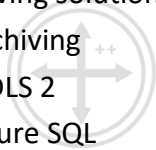
- Azure SQL Purchasing Models

Module 8 : Partitioning

- Introduction to Partitioning
- Agenda
- Logical Partitioning
- Hot Partitions
- Partitioning in Snaps
- Replication
- Round Robin
- Hash implementation
- Azure Data Lake Gen 2 Partitioning

Module 9 : Archiving

- Design a data archiving solution
- Implement data archiving
- Azure Archive in ADLS 2
- Implementation Azure SQL



Module 10 : Data Processing

Clean and transform data by using Spark

- Transform data by using Apache Spark
- Cleanse Data
- Perform data exploratory analysis
- Clean and transform data by using Transact-SQL

Batch data pipeline in Azure

- Develop batch processing solutions by using
- Data factory
- Data lake
- Spark lake
- Azure Synapse Pipelines
- PolyBase and Azure Databricks



- Create data pipelines

Incremental Data loads

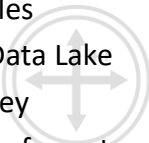
- Design and implement incremental data loads
- Incremental data loading using Azure Data Factory
- Pipeline load

Surrogate key

- Surrogate key and IDENTITY property
- Using IDENTITY to create surrogate keys using dedicated SQL pool in Azure Synapse Analytics

Module 11 : External Tables in Azure Synapse Analytics

- What are External Tables?
- Types of External Tables
 - Hadoop
 - Native
- Create External Tables
- Uploading Data to Data Lake
- Creating a Master Key
- Creating External file format
- Drop External Tables
- Change Master Key



Module 12 : Stream Analytics Windowing Functions

- Non-overlapping Time Windows
 - Tumbling
 - Session
 - Snapshot
- Overlapping Time Windows
 - Hopping
 - Sliding

Module 13: Azure Data Bricks

- What is Azure Databricks?



- Components of Azure Databricks?
- Types of clusters
- Cluster setup

Module 14: Data Encryption & Dynamic data masking

- Introduction to Data Encryption
- Types of Data Encryption
- Data encryption for services
- Data encryption for Data Lake
- Data encryption for snap analytics
- Data encryption for workspace
- Data masking
- Dynamic Data Masking

Module 15: Auditing for Azure SQL Database and Azure Synapse Analytics

- Introduction to Auditing
- Auditing Policy
 - Server-level
 - Database-level
- Auditing policy writes an audit log in
 - Azure storage account
 - Log Analytics workspace
 - Event Hubs



Module 16: Row and Column level security in Azure SQL & Azure Synapse Analytics

- What is Row Level Security
- How to Implement Row Level Security
- What is Column Level Security
- How to Implement Column Level Security