Index

I&D (MS AZURE) LoT Course Structure

I&D (MS AZURE) LoT provides exposure to a band of data warehousing technologies. It focuses on application development for data warehouses. The following table lists the course structure for BI LoT.

Sr. No.	Course	Duration (In Days)
1	Discover(Induction)	-
2	Power skills (Behavioural) -Foundation – session 1	1
3	Data base fundamentals & SQL Server with strong emphasis on TSQL	4
4	Power skills (Behavioural) -Foundation – session 2	0.25
5	Data base fundamentals & SQL Server with strong emphasis on TSQL	2.75
6	Data Warehousing + ETL Concepts	1
7	Python Programming with Data Engineering concepts	1
8	Power skills (Behavioural) -Foundation – session 3	0.25
9	Python Programming with Data Engineering concepts	1.75
10	Basics of Power Shell scripting	1
11	Fundamental Module Test (Platform based Coding Assessment)	0.5
12	Introduction to Big Data	0.5
13	Azure Fundamentals	1
14	Power skills (Behavioural) -Foundation – session 4	1
15	Azure Data Factory (UI based and Power shell based)	2
16	Azure SQL (Advanced)	2
17	Power skills (Behavioural) -Foundation – session 5	1
18	Azure SQL (Advanced)	2
19	Power skills (Behavioural) -Foundation – session 6	0.25
20	Azure BLOB Services & Relational Database	0.75
21	Azure Analysis Services	3
22	Power skills (Behavioural) -Foundation – session 7	0.25
23	PySpark + Databricks(DataBricks Associate Engineer Certification)	2.75
24	Power skills (Behavioural) -Foundation – session 8	1
25	PySpark + Databricks(DataBricks Associate Engineer Certification)	3
26	Power skills (Behavioural) -Foundation – session 9	0.25
27	PySpark + Databricks(DataBricks Associate Engineer Certification)	2.75
28	Sprint Evaluation (Coding Assessment)	1
29	L1 Preparation	1
30	L1 Assessment (MCQ - Concept & Code-based Qs)	1
Total Training Duration		**40**

I&D Curriculum

Data base fundamentals & SQL Server with strong emphasis on TSQL

Program Duration: 7 days
Contents:

What is database?
Why we need database?

History of Database

1. Data Modeling
2. File Systems
3. Hierarchical Databases
4. Network Databases
5. Relational Databases
6. Object Databases
7. Object Relational Databases

Database concepts

1. File Based Approach
2. Disadvantages of file based approach
3. Database Approach

What is DBMS?

1. Advantage in using DBMS
2. Functions of DBMS
3. ACID Properties

What is RDBMS?

1. Difference between DBMS & RDBMS?
2. What is table?
3. What is a field?
4. What is row and column?
5. What is NULL value?

Entity Relationship Modeling

1. What is a relation?
2. What is mean by relationship?
3. Relationship types
4. What is primary key & foreign key?
5. Constraints Data Integrity

Normalization
What is normalization?
List of Normalization forms
1. UNF
2. 1NF
3. 2NF
4. 3NF
5. BCNF

SQL

DDL - Create, Alter, Rename, Truncate, Drop
DML- Insert, Delete, Update, Select
DCL – Grant & revoke
TCL – Commit & Rollback

SQL server 2016 for BI

Introduction to SQL Server

- Connecting to SQL Server using SSMS
- Creating and Working with Tables
- Adding a Default Constraint
- Cascading Referential Integrity Constraint
- Adding a Check Constraint
- Identity Column in SQL Server
- How to check for last generated Identity column value
- Unique Key Constraint
- Select Statement in t-SQL
- Like Operator
- Group by Clause
- Having Clause
- Order by Clause
- T-SQL Functions
- T-SQL String Functions
- T-SQL Date Functions
- T-SQL Numeric Functions
- Cast and Convert functions

- Different ways to replace NULLs in T-SQL
- Coalesce function in t-SQL
- T-SQL Joins
- Self-Join
- Sub query in T-SQL
- Correlated Sub query in T-SQL

- UNION, UNION ALL, EXCEPT, INTERSECT Operator in T-SQL

- Special Functions in t-SQL
- Row Number Function
- Rank and Dense Rank Function
- Calculate Running Total in t-SQL
- NTILE Function
- Lead and Lag Functions
- FIRST VALUE Function
- Window Functions
- LAST VALUE Function
- PIVOT and UNPIVOT
- CHOOSE Function
- IIF Function
- EOMONTH Function
- DATEFROMPATS Function
- Views in SQL Server
- Indexes in SQL Server
- Working with Sequence
- Stored Procedures

Data Warehouse + ETL Concepts

Program Duration: 1 day

Contents:

- Business Intelligence
- Need for Business Intelligence
- Terms used in BI
- Components of BI
- General concept of Data Warehouse
- Data Warehouse
- History of Data Warehousing
- Need for Data Warehouse
- Data Warehouse Architecture
- Data Mining Works with DWH
- Features of Data warehouse
- Data Mart
- Application Areas
- Dimensional modeling
- Dimension modeling

- Fact and Dimension tables
- Database schema
- Schema Design for Modeling
- Star
- Snowflake
- Fact Constellation schema
- ETL and Metadata
- ETL process
- Metadata used in ETL
- Metadata in Data Warehousing
- Simple Data warehouse model
- Online Analytical Processing (OLAP)
- Online Analytical Processing (OLAP)
- Nature of OLAP analysis
- Types of OLAP
- OLAP Tools
- OLTP and OLAP
- OLAP Functional requirements
- OLAP Fast and Selective
- Operational versus Informational System
- Data Mining
- Data mining
- The Knowledge Discovery process
- Need of Data Mining
- Use of Data mining
- Data mining and Business Intelligence
- Types of data used in Data mining
- Data Mining applications
- Data Mining products
- Data Mining market
- Best Practices for Building Data Warehouse
- Recipe for a Successful data warehouse
- Data warehouse pitfalls
- Popular BI DW tools and suits
- Trends in BIDW

AZURE Fundamentals

Program Duration: 1 Day
Content:

1. Cloud computing introduction
2. Cloud deployment models
3. Getting Started with Microsoft Azure
4. Microsoft Azure Management Tools
5. Web Apps and Cloud Services
6. Creating and configuring Virtual networks
7. Cloud Storage
8. Microsoft Azure Databases
9. Introduction to Data Bricks and Snowflake

Azure Data Factory

Program Duration: 2 Days

Content:

Introduction

- What Is Azure Data Factory?
- Why We Need It?
- How does Data Factory work?
- Key Components of Azure Data Factory

Copy Data Flow activity in ADF

- Azure Data Factory Instance

- How to Connect to Azure SQL Data Base from On-Premises
- Linked Services
- Input & Output Data Set in Azure Data Factory

Control Flow Activity

- Get Metadata Activity
- Filter Activity
- If Activity
- Append Activity
- Wait Activity
- ForEach Loop Activity
- Lookup Activity

Dataflow Transformation

- Source Transformation
- Sink Transformation
- Conditional Split Transformation
- Derived Column Transformation
- Lookup Transformation
- Select Transformation
- Filter Transformation
- Join Transformation
- Exists Transformation

Parameterize

- Parameterize Linked Services and Data Sets
- Pipeline Parameters
- Data Flow Parameters

Monitor

- Monitor Visually

Azure Monitor

Azure SQL

Program Duration: 4 days

Contents:

- Introduction/Overview.
- Comparing SQL Azure Database to Azure / On-Premises SQL Server.
- Creating and Using SQL Server and SQL Database.
- Azure SQL Database Tools.
- Using Azure SQL Database with EF Code First.
- Migrating on premise database to SQL Azure.
- Planning the Deployment
- Elastic Storage.
- Monitoring Azure SQL Database
- Configure SQL Database Auditing
- Manage Business Continuity
- Azure SQL Database vs SQL Server in IaaS VM

Azure Blob Services & Relational Databases

Program Duration: 1 day

Contents:

Overview of Azure Blob storage

- Blob storage is designed for

Key Features of Blob Storage

- Consistency
- Mutability
- Blob types
- Geo redundancy

Types of resources

- Storage accounts
- Containers
- Blobs
- Move data to Blob storage

Azure Analysis Services

Program Duration: 3 days

Contents:

- What is Azure Analysis Services?
- How to connect to different data resources

Concepts

- Authentication & User Permissions
- Service Principles
- Client Libraries
- Compatibility level

Types of Models

- Tabular models
- Multidimensional models
- Comparing tabular and multidimensional solutions
- Tabular model solution deployment

Azure Data Bricks PySpark + Databricks(DataBricks Associate Engineer Certification)

Program Duration: 8.5 days

Contents:

SPARK Basics

- What is Spark?
- History of Spark
- Spark Architecture
- Spark Shell

Working with RDDs in Spark

- RDD Basics and Operations
- Transformations and Actions in Spark
- Spark RDD Persistence

Working with Key/Value Pairs

- Pair RDDs
- Data Partitioning (Advanced)
- Loading and Saving the Data.

Spark Advanced

- Accumulators
- Broadcast Variables
- Piping to External Programs
- Numeric RDD Operations
- Spark Runtime Architecture
- Deploying Applications

SPARK with SQL

- Spark SQL Overview
- Spark SQL Architecture
- Catalyst
- Plan Optimization & Execution
- ROW API

Spark streaming

- What is Spark streaming?
- Spark streaming: How it works?
- Spark DStreams

Introduction

- Overview of Big Data Architectures
- Top-down vs bottom-up
- What is Azure Databricks?

Databricks concepts

- Workspace
- Interface
- Data Management
- Computation Management
- Model Management
- Authentication and Authorization

Apache Spark

- What is Apache Spark?
- Spark Architecture
- What is Ecosystem of Apache Spark?
- Data Frames and Datasets

Databricks development and Deployment

- Collaborative Workspace
- Perform ETL Operations
- Deploy production jobs and workflows
- Optimized data bricks runtime engine

Databricks Jobs & Cluster

- Introduction to Jobs and Cluster
- General Spark Cluster Architecture
- How to Submit Jobs using Job Cluster?
- Pool in Databricks
- Azure Databricks Integration with AAD
- Clusters: Auto Scaling and auto termination

Databricks Data Lake

- Data lake defined
- Hadoop as the data lake

Modern data warehouse

- Federated querying
- Solution in the cloud
- SMP Vs MPP