Index

I&D (MS AZURE) LoT Course Structure

Sl No	Topics	Duration	Immersive Approach
1	Discover (Induction)	1	
2	Soft Skills Day 1	1	Soft Skills Foundation – Part 1
3	Data base fundamentals & SQL Server with strong emphasis on TSQL	5	
4	Data Warehousing Concepts	1	
5	Cloud Fundamentals	1	
6	Azure Fundamentals (To cover IAAS-PAAS-SAAS concepts)	1	
7	Basics of Power Shell scripting	1	
8	Introduction to Big Data	1	
9	Module 1 Test	1	Module Test(MCQ + Coding)
10	Soft Skills Day 2	1	Soft Skills Foundation – Part 2
11	Apache Hadoop (Deep dive)	2	Sprint 1- Apache Hadoop (Deep dive),Azure Data Factory (UI based and Power shell based),Azure Data Lake Analytics Gen 2, Azure BLOB Services, Azure Analysis Services,Azure Synapse Analytics
12	Azure Data Factory (UI based and Power shell based)	4	
13	Azure SQL	2	
14	Azure Data Lake Analytics Gen 2	2	
15	Soft Skills Day 3	1	Soft Skills Foundation – Part 3
16	Azure BLOB Services	0.5	Sprint 1- Apache Hadoop (Deep dive),Azure Data Factory (UI based and Power shell based),Azure Data Lake Analytics Gen 2, Azure BLOB Services, Azure Analysis Services,Azure Synapse Analytics
17	Azure Analysis Services	2.5	
18	Azure Synapse Analytics	3	
19	Sprint 1 Evaluation	2	Sprint 1- Evaluation
20	Soft Skills Day 4 – Include interview skills, customer expectations	1	Soft Skills Foundation – Part 4
21	Apache SPARK	4	Sprint 2- Apache SPARK, Python Programming, Azure Data Bricks
22	Python Programming	3	
23	Azure Data Bricks	3	
24	Overview AWS and GCP	1	
25	Sprint 2 Evaluation	2	Sprint 2- Evaluation
26	L1 Test	1	Code and Concept based MCQ
Total Training Duration		48	

Data base fundamentals & SQL server 2016 for BI

Program Duration: 5 day

Day1-

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

Day 2-

a. Connecting to SQL Server using SSMS
b. Creating and Working with Tables
c. Adding a Default Constraint
d. Cascading Referential Integrity Constraint
e. Adding a Check Constraint
f. Identity Column in SQL Server
g. How to check for last generated Identity column value
h. Unique Key Constraint
i. Select Statement in t-SQL
j. Like Operator
k. Group by Clause
l. Having Clause
m. Order by Clause
n. T-SQL Functions
o. T-SQL String Functions
p. T-SQL Date Functions
q. T-SQL Numeric Functions
r. Cast and Convert functions

Day3-

a. Different ways to replace NULLs in T-SQL
b. Coalesce function in t-SQL
c. T-SQL Joins
d. Self-Join
e. Sub query in T-SQL
f. Correlated Sub query in T-SQL
g. UNION, UNION ALL, EXCEPT, INTERSECT Operator in T-SQL

Day 4-

a. Special Functions in t-SQL
b. Row Number Function
c. Rank and Dense Rank Function
d. Calculate Running Total in t-SQL
e. NTILE Function
f. Lead and Lag Functions
g. FIRST VALUE Function
h. Window Functions
i. LAST VALUE Function
j. PIVOT and UNPIVOT
k. CHOOSE Function
l. IIF Function
m. EOMONTH Function
n. DATEFROMPATS Function

Day5-

a. Views in SQL Server
b. Indexes in SQL Server
c. Working with Sequence
d. Stored Procedures

Data Warehouse Concepts

Program Duration: 1 days.
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
- Snow Flake
- 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

Cloud Fundamentals & Azure Fundamentals

Program Duration-1 Days
Content

What is cloud computing?

What makes cloud computing different?

Types of cloud computing

Types of cloud services: IaaS, PaaS, serverless and SaaS

Advantages and disadvantages of cloud computing

AZURE Fundamentals

Contents:

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

Basics of Power Shell scripting

Program Duration: 1 day

Introduction

- What is PowerShell?
- Why Use PowerShell?
- PowerShell History
- PowerShell Concepts
- Applications of PowerShell
- PowerShell Vs Command Prompt
- What is PowerShell ISE?

Features of PowerShell

- PowerShell Remoting
- Background Jobs
- Transactions
- Evening
- Network File Transfer
- How to launch PowerShell

PowerShell - Cmdlet

- Cmdlet vs Command
- Advanced Cmdlets

PowerShell – Scripts

- Creating A PowerShell Script
- Executing A Script
- Data types (Integer, Char, Date, String and so on)
- Variables ($Error, $Host, $Profile and so on)
- Constants
- Comparison Operators
- Looping concepts
- Regular Expressions

PowerShell Providers

- Understanding Providers
- The File System Provider

- The Alias Provider
- The Variable and Function Providers
- The Environment Provider
- The Registry Provider
- The Certificate Provider

Working with Files and Folders

- Reading and Writing
- Output to HTML
- Output to XML
- Working with CSV

Big Data Overview

Program Duration: 1 day

Introduction:

- Big Data overview
- Hadoop insight
- Big data Analytics

Components

- Storage Layer – HDFS
- Processing Layer - MapReduce
- In-memory Processing

Hadoop Ecosystem

- Pig Latin
- Hive query language

Introduction to Real-time Processing and brief on Kafka
One Big Data use case briefing

s

Apache Hadoop (Deep Drive)

Program Duration: 2 days

Contents:

Hadoop Introduction

- What is Hadoop?
- Hadoop Features

- Hadoop Run Modes and Job Types

Hadoop Distributed File System (HDFS)

- HDFS Architecture
- HDFS Components
- HDFS Client creating new file
- Rack Description
- HDFS Write Operation
- Selection of Data Nodes & Node Distance
- Serialization
- HDFS Caching & Failover
- HDFS Federation
- HDFS High Availability

Map Reduce

- MapReduce Introduction
- MapReduce Architecture
- Map Reduce Data Types
- Advanced MapReduce - Partioners, Combiners, Comparators and More
- Partitioner Code Walkthrough
- Resource Manager Failure

YARN

- YARN overview
- YARN processing

HIVE

- HIVE Basics
- HIVE Architecture
- Hive Database
- HIVE Query
- Hive Managed Tables
- Hive External Tables
- Hive Patterns and Anti-Patterns
- Working with Parquet
- Hive Partitioning
- Hive Bucketing

PIG

- Apache PIG Introduction
- PIG Modes
- Pig Patterns and Anti-Patterns

- Load and Store Operators
- Diagnostic Operators
- Grouping and Joining
- Combining and Splitting
- Pig Latin Built-In Functions
- Comparison of PIG, HIVE, MapReduce

SQOOP

- Sqoop Overview
- Sqoop Import Export
- Managing Target Directories
- Working with Different File Formats
- Conditional Imports
- Split-by and Boundary Queries
- Field delimiters
- Incremental Appends
- Sqoop Hive Import
- Sqoop List Tables/Database
- Export from HDFS to Mysql
- Export from Hive to Mysql

Apache Flume

- Flume Introduction & Architecture
- Exec Source and Logger Sink
- Moving data from Twitter to HDFS
- Flume Interceptors
- Flume Multi-Agent Flow
- Flume Consolidation

Oozie

- Oozie Introduction
- Workflow
- Property File
- Coordinator
- Bundle
- CLI and Extensions

Azure Data Factory

Program Duration: 4 Day

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-Premise
- 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 Data Lake Analytics Gen 2

Program Duration: 2 day

Content

Introduction

- What is Azure Data Lake Storage Gen2
- How to create Data Lake Storage Gen2?
- Difference between Gen1 and Gen2

Manage a Data Lake Gen2

- Explore Storage Explore (Object & hierarchical file storage)
- Access control & data transfer options
- Shared Access Signature (SAS)
- Encryption
- Monitoring
- Security & Firewall
- Performance

Key features of Data Lake Storage Gen2

- Massive Scalability
- Cost effectiveness
- Supported Blob storage features
- Supported Azure service integrations
- Supported open source platforms

Azure SQL Database

Program Duration: 2 days

- Introduction/Overview.
- Comparing SQL Azure Database to Azure / On-Premise 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 IaaA VM

Azure Blob Storage

Program Duration: 1 day

Introduction

- 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: 2.5 day

- 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 Synapse Analytics

Program Duration: 3 day

- What is Azure Synapse Analytics (formerly SQL DW)?
- Azure Synapse Benefits

Azure Synapse Analytics (formerly SQL DW) architecture

- Azure Synapse MPP Architecture
- Storage and Sharding patterns
- Data Distribution and Distributing Keys
- Partitioning

Azure Synapse Analytics features

- Limitless scale
- Powerful insights
- Unified experience

Security layers

- Advance data security
- Network Security
- Transparent data encryption
- Dynamic Data Masking
- Access Management

Spark in memory

Program Duration: 4 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

Python Programming

Program Duration: 3 days

Contents:

Introduction to Python

- Why do we need Python?
- Program structure in Python

Python Basics

- Data types
- Variables and Assignments
- Strings
- Execution steps
- Pattern Matching with Regular Expressions
- introduction to lists

Flow Control

- comparison operators
- Statements and Syntax in Python
- Loop Concepts

Functions in Python

- Function definition and call
- Function Scope and Types
- Return Values and return Statements
- Arguments
- Exception Handling

Modules and Packages-Basic

- Module Creations and Usage
- Package Creation and Importing

Classes in Python

- Classes and instances
- Classes method calls

Libraries

- Importing a library using PIP, CONDO etc
- Math
- Numpy

Working with RDBMS

- Connection to Database
- Cursor Creation
- Fire Query & Collect results from Tables/Queries
- How to Insert Data into Tables and Types?

Debugging

- Raising Exceptions
- Assertions
- Logging Module and File
- IDLE's Debugger
- Breakpoints

Working with CSV Files and JSON Data

- CSV Module and delimiters
- Reader and Writer Objects
- JSON Module and I/O Functions

Multithreading
Running Other Python Scripts

Azure Data bricks

Program Duration: 3 days

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?
- DataFrames and Datasets

Databricks development and Deployment
- Collaborative Workspace
- Perform ETL Operations
- Deploy production jobs and workflows
- Optimized databricks 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

Overview of AWS and GCP

Program Duration: 1 day

- Introduction to AWS
- What is “Cloud Computing”?
- Amazon and Cloud Computing
- The Differences that Distinguish AWS
- Flexible

Cost-Effective

- Scalable and Elastic
- Secure
- Experienced

Amazon Web Services Cloud Platform

- Compute & Networking
- Storage & Content Delivery Network
- Database
- Analytics
- Application Services
- Deployment and Management
- What is Google Cloud Platform (GCP)?
- GCP services & Benefits
- Why GCP?
- What is Google Compute Engine (GCE)
- How to create a VM using GCE
- Overview of Google Cloud Shell & gcloud CLI tool