

Azure Data Engineering Content

Basics Of Cloud Computing

- What is On Premise?
- What is Cloud?
- Cloud Deployment Models
- Types Of Services
- Cloud Vendors

Microsoft Azure - Introduction

- Introduction
- Creating Account
- Subscription
- Resource Group
- Resources
- Best Practices for Billing

Azure Storage Platform - ADLS

- Introduction
- Creating ADLS Gen2 Account
- Uploading Files into Container
- Storage Types
- Pricing Tiers
- Redundancy
- Security & Access
- Azure Storage Explorer

Azure Storage Platform - SQL Server

- Introduction
- SQL Server Deployment Options
- Purchasing Models
- Pricing Tiers
- Creating Azure SQL Server
- Query Editor
- SSMS Installation
- SQL Server Management Studio
- Azure Data Studio

Microsoft SQL Server - OnPrem

- Introduction
- Data Definition Language
- Data Manipulation Language
- Data Querying Language
- SELECT Statement
- WHERE Clause
- Operators & Conditions
- Numeric Functions
- Text Functions
- Date & Time Functions
- Aggregate Functions
- GROUP BY Clause
- HAVING Clause
- Data Merging - Joins
- Data Appending - UNION & UNION ALL
- Sub Queries
- CTE's
- Window Functions
- Views and Stored Procedures

Azure Data Factory - Introduction

- What is ETL/ELT?
- What is Data Factory?
- Data Factory Architecture
- Components Of Data Factory
- Integration Runtimes
- Linked Services
- Datasets
- Dataflows
- Pipeline
- Triggers

ADF - Copy Data Activity

- Introduction
- Source and Target/Sink
- File System as A Source
- Creating Linked Services
- Creating Datasets
- Creating Pipeline
- Debug/Run Pipeline
- Azure SQL Database As A Source
- Installing Self-Hosted IR
- On Prem SQL Server As A Source

ADF - Variable Activity

- Introduction
- Creating Variable
- Set A Variable with Expression
- Set Current Date MMddYYYY
- Dataset-Parameter
- Add Dynamic Content - Expression
- Passing Value to Dataset
- Pipeline to load current day file
- Pipeline to load previous day file

ADF - Get Metadata Activity

- Introduction
- Field List - Exists
- Field List - Child Items
- Field List - Size
- Getting File System Details
- Getting Database Details

ADF - Filter Activity

- Introduction
- Getting Output From Get Metadata
- Filter Only .csv Files and Only Specific Files
- Functions Introduction

ADF - Lookup Activity

- Introduction
- Lookup On File System
- Lookup On Database
- Getting Single Row
- Getting Multiple Rows
- Limitations

ADF - Foreach Activity

- Introduction
- Wait Activity
- Items()
- Taking List Of Items
- Changing File Names and Adding Date To File Names
- Creating File Names Dynamically
- Loading Data From Files and Loading Only .csv Files
- Loading Data From Multiple Tables
- Pipeline To Load Specific Tables

ADF - If Condition Activity

- Introduction
- True Block
- False Block
- Pipeline To Load Data If Data Exists

ADF - Delete Activity

- Introduction
- Deleting File
- Deleting Folder
- Delete File Based on Size

ADF - Switch Activity

- Introduction
- Adding Block by Expressions
- Copy & Delete By Expression

ADF - Execute Pipeline Activity

- Introduction
- Execute Pipeline Activity
- Using Foreach Inside Foreach
- Parent Pipeline
- Child Pipeline
- Passing Params To Child From Parent
- Load Data From Multiple Folders

ADF - Stored Procedure Activity

- Introduction
- Creating Procedure For Logging
- Passing Values To Procedure
- Executing Procedure

ADF - Until & Validation Activity

- Until Activity Introduction
- Pipeline To Move If File Existed
- Pipeline To Wait Till File Existed
- Validation Activity Introduction
- Move a File If Dataset Reference Existed

ADF - Fail Activity

- Introduction
- Fail A Pipeline If File Is Not Existed
- Fail A Pipeline If Job in Progress

ADF - Dataflow Activity

- Introduction
- Dataflow Debug
- Source and Sink/Target
- Filter Transformation
- Conditional Split Transformation
- Derived Column Transformation
- Select Transformation
- Sort Transformation
- Join Transformation
- Lookup Transformation
- Exists Transformation
- Union Transformation
- Aggregate Transformation

ADF - Code Repository

- Introduction
- Creating DevOps Organization
- Creating DevOps Project
- Integrating Repository To ADF
- Create Working Branch
- Creating SCD Type-1
- Create Pull Request
- Merging Cod With Collaboration Branch
- Create Working Branch Again
- SCD Type - 2
- Compare The Code Changes
- Create Pull Request Again
- Merge Code With Collaboration Branch

Azure Databricks - Introduction

- What is Bigdata ?
- Distributed File Systems
- HDFS, DBFS and ADLS
- Create Azure Databricks
- Create Cluster
- Types Of Clusters
- Cluster Modes
- Managing Clusters

Azure Databricks - Python Basics

- Python Introduction
- Create Databricks Community Edition Account
- Notebook
- Variables
- Datatypes
- Type Conversions
- Collection Datatypes
- List
- Tuple
- Set
- Dictionaries
- Operators
- Conditional Statements
- For Loop
- String Operations
- List Operations
- Tuple Operations
- Set Operations
- Custom Functions
- Lambda Functions
 - Map Function
 - Filter Function
 - Reduce Function

Azure Databricks - DBFS

- Introduction
- Dbutils
- mkdirs Command
- Create Directories
- cp command
- Copy a File
- Copy a Folder
- mv command
- Move a File
- Move a Folder
- ls Command
- List Out All The Files in Folder
- put Command
- head Command
- Copy Current Day File
- rm Command
- Remove a File
- Remove a Folder

Azure Databricks - RDD

- Introduction
- Types Of Transformations
 - Narrow Transformations
 - Wide Transformations
- Lazy Evaluation
- Actions
- Data Lineages
- Create RDD - Python Object
- Create RDD With Existed RDD
- Create RDD - By Reading File
- RDD - map Transformation
- RDD - reduceByKey transformation
- RDD - flatMap Transformation
- RDD - filter Transformation
- RDD - collect Action
- RDD - getNumPartitions
- RDD - glom Transformation
- RDD - coalesce Transformation
- RDD - repartition Transformation
- RDD - Word Count Program
- RDD - Select Required Columns
- RDD - Datatype Conversion
- RDD - Single Aggregation, Grouping
- RDD - Single Aggr & Multi Grouping
- RDD - Multi Aggr & Single Grouping
- RDD - Multi Aggregations & Groupings

Azure Databricks - Dataframe

- Introduction
- Dataframe - Python Object
- Dataframe - RDD
- Dataframe - Without Columns
- Dataframe - With Columns
- Dataframe - printSchema()
- Dataframe - Spark Reader
- Dataframe - csv()
- Dataframe - format()
- Dataframe - load()
- Dataframe - parquet()
- Dataframe - select()
- Dataframe - selectExpr()
- Dataframe - filter()
- Dataframe - where()
- Dataframe - orderBy()
- Dataframe - sortBy()
- Dataframe - describe()
- Dataframe - distinct()
- Dataframe - dropDuplicates()
- Dataframe - drop()
- Dataframe - first()
- Dataframe - take()
- Dataframe - tail()
- Dataframe - replace()
- Dataframe - withColumnRenamed()
- Dataframe - count()
- Dataframe - Group By
- Dataframe - Aggregations
- Dataframe - Joins
- Dataframe - broadcasting
- Dataframe - union()
- Dataframe - intersect()
- Dataframe - write()
- Dataframe - save()
- Dataframe - Creating View

Spark – SQL

Real Time Project