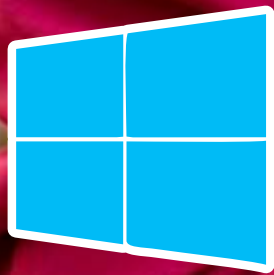


CELEBRATING
13 YEARS

QualityThought®

Hyderabad | Vijayawada | Chennai



Azure

Cloud Data Engineer

Azure Basics, Azure SQL DB

- ⇒ Cloud Introduction and Azure Basics
- ⇒ Azure Implementation: IaaS, PaaS, SaaS
- ⇒ Benefits of Azure Cloud Environment
- ⇒ Azure Data Engineer: Job Roles
- ⇒ Azure Storage Components
- ⇒ Azure ETL & Streaming Components
- ⇒ Need for Azure Data Factory (ADF)
- ⇒ Need for Azure Synapse Analytics
- ⇒ Azure Resources and Resource Types
- ⇒ Resource Groups in Azure Portal
- ⇒ Azure SQL Server [Logical Server]
- ⇒ Firewall Rules and Azure Services
- ⇒ Connections with SSMS & ADS Tools
- ⇒ DB Migration: OnPremise to Azure
- ⇒ Schema Migration & Data Migration

Synapse SQL Pools (DWH)

- ⇒ Dedicated SQL Pools in Azure
- ⇒ Enterprise Data Warehouse with Synapse
- ⇒ DWU: Data Warehouse Units, Resources
- ⇒ Massively Parallel Processing (MPP)
- ⇒ Control Nodes and Compute Nodes
- ⇒ SQL Pool Access from SSMS Tool
- ⇒ T-SQL Queries @ SQL Pools
- ⇒ Start/Resume/Pause, Scaling Options
- ⇒ Creating Tables in Azure SQL Pool
- ⇒ Compression, MAX DOP & Indexes
- ⇒ Distributions: Round Robin, Hash
- ⇒ Distributions: Replicate and Usage
- ⇒ Data Imports with COPY Table
- ⇒ Dynamic Views (DMV) with PDW
- ⇒ Data Loads Monitoring, Resource Class

Azure Data Factory Concepts

- ⇒ Azure Data Factory (ADF) Concepts
- ⇒ Hybrid Data Integration at Scale
- ⇒ ADF Pipeline Components & Usage
- ⇒ Configure ADF Resource in Azure
- ⇒ Understanding ADF Portal and IR
- ⇒ Linked Services and Connections
- ⇒ Datasets and Tables / Files for ETL
- ⇒ ADF Pipelines: Design, Publish & Trigger
- ⇒ ADF Pipeline with Copy Data Tool
- ⇒ Creating Azure Storage Account
- ⇒ Storage Container, BLOB File Uploads

- ⇒ Data Loads with Azure BLOB Files
- ⇒ DIU Allocations and Concurrency
- ⇒ Creating Linked Services, Datasets
- ⇒ Pipeline Trigger, Author and Monitor

ADF Pipelines, Polybase

- ⇒ Copy Data Tool For ETL Operations
- ⇒ Azure SQL DB to Synapse Data Loads
- ⇒ Working with Multi Tables Data Loads
- ⇒ Query Options for Source Datasets
- ⇒ Transformations with Copy Data Tool
- ⇒ Rename, Rearrange & Remove Options
- ⇒ Pipeline Execution: DTU & DOCP
- ⇒ ADF Pipeline Monitoring Options
- ⇒ ADF Pipelines: Execution Settings
- ⇒ ADF Logging Options & Storage Account
- ⇒ Compression Option, DOP and DOCP
- ⇒ ETL Staging Advantages & Performance
- ⇒ Staging with Storage Account, Container
- ⇒ ADF Pipeline Triggers and Monitoring
- ⇒ Polybase For Azure Synapse, Advantages

OnPremise Data with ADF

- ⇒ On-Premise Data Sources with Azure
- ⇒ Self Hosted Integration Runtime (IR)
- ⇒ Access Keys, Remote Linked Services
- ⇒ Synapse SQL Pool (DW) with OnPremise
- ⇒ Staged Data Copy and Performance
- ⇒ Pipeline Executions and Monitoring
- ⇒ Pipeline RunIDs and Audits / Tracing
- ⇒ Incompatible Rows Skips, Fault Tolerance
- ⇒ Incremental Loads with Files (BLOB)
- ⇒ Pipeline Executions and Schedules
- ⇒ Regular Schedules and Tumbling Window
- ⇒ Execution Retry and Delay Options
- ⇒ Binary Copy, Last Modified Date in Blob
- ⇒ Automated Loops and Trigger Schedules
- ⇒ Incremental Loads Verification Tests

ADF Data Flow - 1

- ⇒ Limitations with Copy Data Tool
- ⇒ Data Flow Task, Data Flow Activity
- ⇒ Transformations with Data Flow
- ⇒ Spark Cluster For Debugging
- ⇒ Cluster Node Configurations
- ⇒ Data Preview Options with DFT
- ⇒ SELECT Transformation & Options

- ⇒ JOIN Transformation and Usage
- ⇒ Conditional Split Transformation
- ⇒ Aggregate & Group By Transformations
- ⇒ Synapse Sink Options with DFT
- ⇒ DFT Optimization Techniques
- ⇒ Pipeline Debug Runs and ETL Testing
- ⇒ Spark Cluster For Pipeline Executions
- ⇒ Pipeline Monitoring & Run Ids

ADF Data Flow - 2

- ⇒ ADF Pipelines For ETL Operations
- ⇒ Data Flow Tasks and Activities in Synapse
- ⇒ Pivot Transformation For Normalization
- ⇒ Generating Pivot Column, Aggregations
- ⇒ Pivot Transformation and Pivot Settings
- ⇒ Pivot Key Selection, Value and Nulls
- ⇒ Pivoted Columns and Column Pattern
- ⇒ Column Prefix, Help Graphic & Metadata
- ⇒ Window Functions & Usage in Data Flow
- ⇒ Rank / DenseRank / Row Number
- ⇒ Over Clause and Input Options
- ⇒ Derived Column Transformations
- ⇒ Exists & Lookup Transformations
- ⇒ Reusing Data Flow Tasks in Synapse
- ⇒ Pipeline Validations & Executions

Azure Synapse Analytics

- ⇒ Azure Synapse Analytics Resource
- ⇒ Azure Synapse Analytics Workspace
- ⇒ Managed Resource Group, SQL Account
- ⇒ SQL Admin Account and its Purpose
- ⇒ Operations with Synapse Workspace
- ⇒ ADLS Gen 2 Storage Account, Container
- ⇒ Synapse Studio (Synapse Portal)
- ⇒ Dedicated SQL Pools & Spark Pools
- ⇒ Creating Dedicated SQL Pools
- ⇒ Synapse Tables, Data Loads with T-SQL
- ⇒ COPY INTO Statements with T-SQL
- ⇒ Clustered Column Store Indexes
- ⇒ Row Terminator and Compressions
- ⇒ T-SQL Queries and Aggregations
- ⇒ Aggregation Data Loads in Synapse

Synapse Analytics with Spark

- ⇒ Apache Spark Pool in Azure Synapse
- ⇒ Spark Cluster Nodes: Vcores, Memory
- ⇒ Creating Spark Clusters @ Synapse Studio
- ⇒ Python Notebooks For Remote Access
- ⇒ Creating Databases in Apache Spark Pool
- ⇒ Data Loads from Dedicated SQL Pools
- ⇒ Table Creations, Aggregation Operations
- ⇒ PySpark Code for Data Operations, Writes
- ⇒ Serverless Pool in Azure Synapse
- ⇒ Connections, Usage with Serverless Pool
- ⇒ Using Azure OpenDatasets in Synapse
- ⇒ OPENROWSET and BULK Data Loads
- ⇒ Azure Storage Account : Data Analysis
- ⇒ Working with Parquet Files in Synapse
- ⇒ Python Notebooks (Pyspark) in Synapse

Incremental Loads @ Synapse

- ⇒ Incremental Loads with Synapse Studio
- ⇒ Multi Table Merge Operations
- ⇒ On-Premise Data Sources & Timestamps
- ⇒ Azure SQL DB Destinations, Watermarks
- ⇒ Watermark Table Usage & Audits
- ⇒ Stored Procedures for Timestamp Updates
- ⇒ Table Data Type and Dynamic MERGE
- ⇒ SQL Queries for Datasets and Fetch
- ⇒ Lookup Activity and its Usage un Synapse
- ⇒ Expressions in ADF Portal for Lookup
- ⇒ Expressions in ADF Portal for Source
- ⇒ Output Pipeline Expression, Data Window
- ⇒ Concat Function, Run IDs Expressions
- ⇒ JSON Parameters, Pipeline Scheduling
- ⇒ Pipeline Validation, Trigger and Monitoring

Optimizations, Power Query

- ⇒ ADF ETL with GUI : Power Query
- ⇒ Power Query Resource Creation, Use
- ⇒ Source Data Configurations & Settings
- ⇒ Rename, Remove, Pivot, Group By, Order
- ⇒ Index, Filter, Remove Error Rows
- ⇒ Using Power Query Activity, ADF Pipelines
- ⇒ Spark Cluster Configurations for Pipelines
- ⇒ Concurrency, Big Data Recommendations
- ⇒ Storage Optimization Techniques
- ⇒ ETL Optimization Techniques
- ⇒ SQL Pool (Synapse) Optimizations

- ⇒ Indexes, Partitions, Distributions, DOP
- ⇒ Pipeline Optimization Techniques
- ⇒ Partitions, DOP, Compressions, DIU
- ⇒ Staging, Polybase and Core Counts

Pipeline Monitoring, Security

- ⇒ Azure Monitor Resource and Usage
- ⇒ Pipeline Monitoring Techniques
- ⇒ ADF: Pipeline Monitoring and Alerts
- ⇒ Synapse: Pipeline Monitoring and Alerts
- ⇒ Synapse: Storage Monitoring and Alerts
- ⇒ Conditions, Signal Rules and Metrics
- ⇒ Email Notifications with Azure
- ⇒ Concurrency, Big Data Recommendations
- ⇒ Azure Active Directory (AAD) Users, Groups
- ⇒ IAM: Identity & Access Management
- ⇒ Synapse Workspace Security with RBAC
- ⇒ ADF Security with RBAC: Owner, Contributor
- ⇒ Azure Synapse SQL Pool Security: Logins
- ⇒ Users, Roles and Resource Classes (RC)
- ⇒ Assigning RCs to Users. Object Level Security

Azure Data Bricks

- ⇒ capabilities of Azure Databricks and the Apache Spark notebook for processing huge files. architecture of an Azure Databricks Spark Cluster and Spark Jobs.
- ⇒ how to use Azure Databricks supports day-to-day data-handling functions, such as reads, writes, and queries
- ⇒ Process data in Azure Databricks by defining DataFrames to read and process the Data
- ⇒ Use the DataFrame Column Class Azure Databricks to apply column-level transformations, such as sorts, filters and aggregations
- ⇒ platform architecture and how it is secured Use Azure Key Vault to store secrets used by Azure Databricks and other services
- ⇒ how to use Delta Lake to create, append, and upsert data to Apache Spark tables, taking advantage of built-in reliability and optimizations
- ⇒ Process streaming data with Azure Databricks structured streaming. Create production workloads on Azure Databricks with Azure Data Factory
- ⇒ how to put Azure Databricks notebooks under version control in an Azure DevOps repo and build deployment pipelines to manage your release process

Power BI

- ⇒ Basic report Design
- ⇒ Visual Syn and grouping
- ⇒ Filters and Hierarchies
- ⇒ Bookmarks , Azure Modeling
- ⇒ Visualization properties
- ⇒ DAX Funcrions
- ⇒ Power Query
- ⇒ Power BI Cloud and Power BI Service
- ⇒ Stacked bar chart, Stacked column chart
- ⇒ Clustered bar chart, Clustered column chart
- ⇒ Adding Report Titles. Report Format Options
- ⇒ Focus Mode, Explore and Export Settings
- ⇒ Paginated reports
- ⇒ Power App
- ⇒ Report Fields, Formats and Analytics
- ⇒ Page-Level Filters and Column Formatting, Filters
- ⇒ Background Properties, Borders and Lock Aspect

CELEBRATING
13 YEARS

QualityThought®



QualityThought

90302 25436

Quality Thought Infosystems India (P) Ltd.

#1, 7th Street, Tansi Nagar, Taramani 100ft Road, Velachery, Chennai-600042 | www.qualitythought.in | chennai@qualitythought.in
Hyderabad | Vijayawada | Chennai