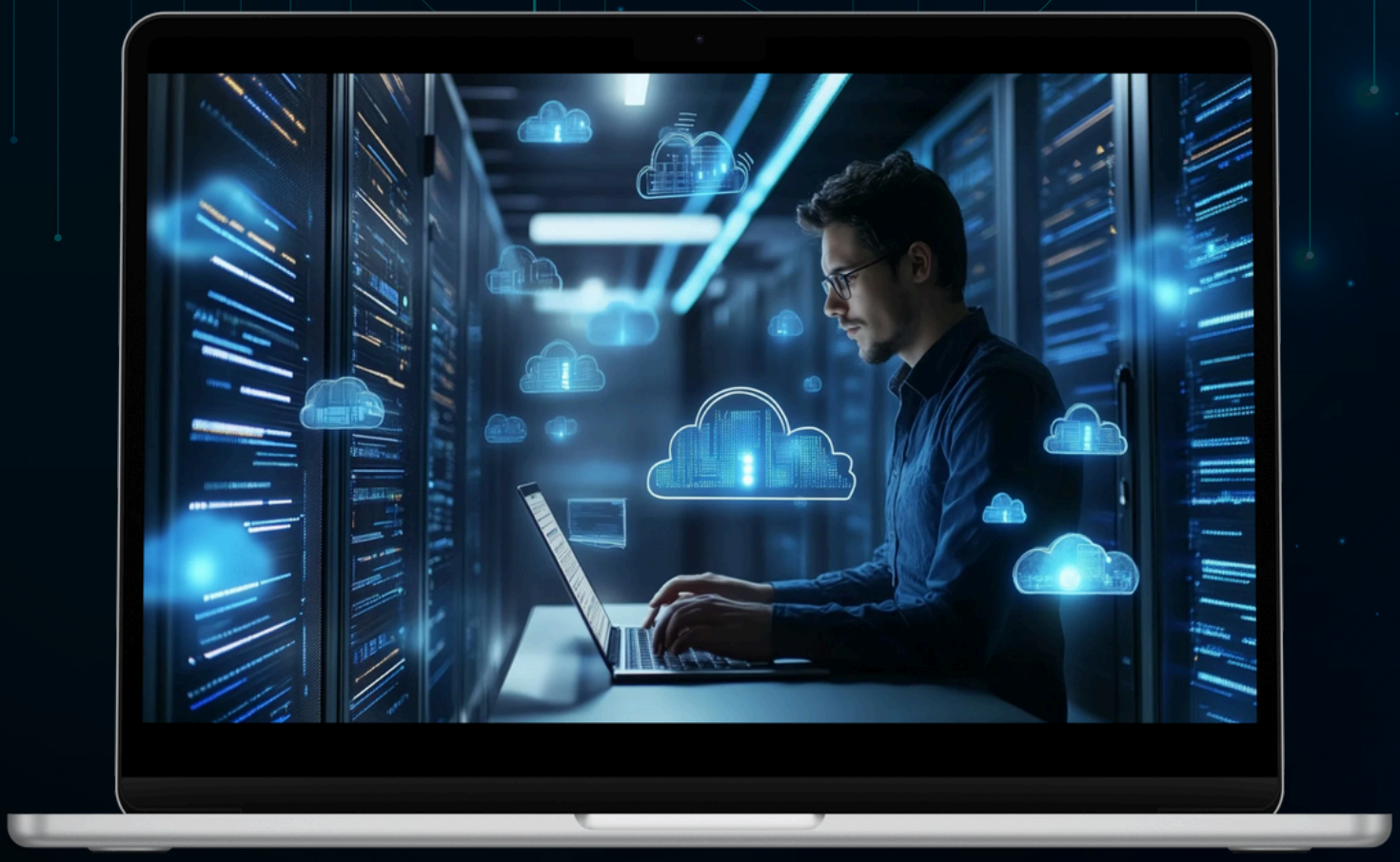
Cloud UpSkill

UP-SKILL YOURSELF WITH AZURE DATA ENGINEERING PROGRAM



Info@cloudupskill.com | https://cloudupskills.com

What is ADE (Azure Data Engineer)?

The Azure Data Engineer (ADE) course equips learners with the skills to design and implement data solutions using Microsoft Azure. It focuses on building pipelines, integrating data from various sources, transforming it for analytics, and ensuring secure, scalable, and high- performance data processing systems in the cloud.

Why Azure Data Engineering Matters?

In a world driven by data, organizations rely on cloud platforms like Azure to make smarter, faster decisions. Azure Data Engineers are pivotal in handling massive volumes of data, ensuring it is reliable, clean, and ready for insights. With businesses moving to the cloud at an accelerating pace, the demand for professionals who can manage and orchestrate data pipelines is higher than ever.

Career Opportunities After these Course

- **Cloud Data Engineer**
- **Big Data Engineer**
- **Data Engineer**
- **ETL Developer**
- **Azure Data Analyst**
- **Business Intelligence Engineer**

Cloud UpSkill

Info@cloudupskill.com | https://cloudupskills.com

Cloud UpSkill

Who Should attend Course?

- ✅ Fresh Graduates interested in Cloud & Data roles
- ✅ IT Professionals transitioning to Cloud Data Engineering
- ✅ Data Analysts & BI Developers upskilling for Azure roles
- ✅ Anyone preparing for Microsoft DP-203 Certification

✉️ Info@cloudupskill.com | https://cloudupskills.com

What You'll Learn in the ADE Program

You will:

- ✅ Design and build scalable data pipelines using Azure Data Factory
- ✅ Transform and clean data with Azure Databricks and SQL
- ✅ Implement data storage using Azure Data Lake and Synapse Analytics
- ✅ Monitor and optimize data workflows in production
- ✅ Apply best practices in data governance, security, and compliance

Tools You'll Master

- Azure Data Factory
  Data pipeline orchestration
- Azure Databricks
  Scalable big data processing
- Azure Synapse
  Analytics and data warehousing
- Azure Data Lake
  Big data storage
- Azure Monitor
  Pipeline monitoring and diagnostics
- Azure Key Vault
  Secure credential management

Career & Project Readiness

- 🎓 Real-world projects aligned with industry scenarios
- 🟦 Prepared for Microsoft DP-203 certification exam
- 👩‍💻 Contribute to cloud data teams from day one
- 📁 Build a portfolio with labs and case studies

✉️ Info@cloudupskill.com | https://cloudupskills.com

Azure Data Factory 01

Azure Data Factory:

- ✅ Introduction to Azure Data Factory
- ✅ Concepts of Azure Data Factory
- ✅ Creating our first Azure Data Factory
- ✅ Different ways to work with Azure Data Factory
- ✅ Pipelines & activities in Azure Data Factory
- ✅ Linked Services & Datasets in Azure Data Factory

Triggers:

- ✅ Triggers in Azure Data Factory
- ✅ Schedule Trigger in Azure Data Factory
- ✅ Tumbling Window in Azure Data Factory
- ✅ Tumbling Window Trigger Dependency in Azure Data Factory
- ✅ Event Based rigger in Azure Data Factory

Linked Services & Datasets in Azure Data Factory:

- ✅ Integration runtime in Azure Data Factory
- ✅ Azure Integration runtime in Azure Data Factory
- ✅ Self hosted Integration runtime in Azure Data Factory
- ✅ Setting up Self hosted Integration runtime in Azure Data Factory
- ✅ Shared Self hosted Integration runtime in Azure Data Factory
- ✅ Parameterize Linked Services in Azure Data Factory
- ✅ Parameterize datasets in Azure Data Factory
- ✅ Parameterize Pipelines in Azure Data Factory
- ✅ System Variables in Azure Data Factory
- ✅ Connectors Overview in Azure Data Factory
- ✅ Supported File formats in Azure Data Factory

Activities:

- ✅ Copy data activity in Azure Data Factory
- ✅ Monitor copy data activity in Azure Data Factory
- ✅ Delete activity in Azure Data Factory
- ✅ Variables in Azure Data Factory
- ✅ Set variable activity in Azure Data Factory
- ✅ Append variable activity in Azure Data Factory
- ✅ User properties in Azure Data Factory
- ✅ Execute pipeline activity in Azure Data Factory
- ✅ Filter activity in Azure Data Factory
- ✅ ForEach activity in Azure Data Factory
- ✅ Get metadata activity in Azure Data Factory
- ✅ If condition activity in Azure Data Factory
- ✅ Wait activity in Azure Data Factory
- ✅ Until activity in Azure Data Factory
- ✅ Web activity in Azure Data Factory
- ✅ Webhook activity in Azure Data Factory
- ✅ Switch activity in Azure Data Factory
- ✅ Validation activity in Azure Data Factory
- ✅ Lookup activity in Azure Data Factory
- ✅ Transformation data activities overview in Azure Data Factory
- ✅ Stored procedure activity in Azure Data Factory

Dataflow:

- ✅ Dataflow in Azure Data Factory
- ✅ Mapping dataflow in Azure Data Factory
- ✅ Dataflow activity in Azure Data Factory
- ✅ Mapping dataflow debug mode in Azure Data Factory

Transformations & Mapping Dataflows in Azure Data Factory:

- ✅ Filter transformation in mapping dataflow in ADF
- ✅ Aggregate transformation in mapping dataflow in ADF
- ✅ JOIN transformation in mapping dataflow in ADF
- ✅ Conditional split transformation in mapping dataflow in ADF
- ✅ Derived column transformation in mapping dataflow in ADF
- ✅ Exist transformation in mapping dataflow in ADF
- ✅ Union transformation in mapping dataflow in ADF
- ✅ Lookup Transformation in mapping dataflow in ADF
- ✅ Sort transformation in mapping dataflow in ADF
- ✅ New branch in mapping dataflow in ADF
- ✅ Select transformation in mapping dataflow in ADF
- ✅ Pivot transformation in mapping dataflow in ADF
- ✅ Unpivot transformation in mapping dataflow in ADF
- ✅ Surrogate key transformation in mapping dataflow in ADF
- ✅ Window transformation in mapping dataflow in ADF
- ✅ Alter Row transformation in mapping dataflow in ADF
- ✅ Flatten transformation in mapping dataflow in ADF
- ✅ Parameterize mapping dataflow in ADF
- ✅ Validate schema in mapping dataflow in ADF
- ✅ Schema drift in mapping dataflow in ADF
- ✅ Wrangling dataflow overview in ADF
- ✅ Merge Queries in wrangling dataflow in ADF
- ✅ GroupBy in wrangling dataflow in ADF
- ✅ Rank transformation in mapping dataflow in ADF
- ✅ Cache sink & cache lookup in mapping dataflow in ADF
- ✅ Parse Transformation in Mapping Data Flow in ADF
- ✅ Stringify transformation in Mapping Data Flow in ADF
- ✅ Assert Transformation in Mapping Data Flows in ADF
- ✅ Flowlets in Mapping data flow in ADF
- ✅ Extract Data from table of website page using ADF

Data Pipelines:

- ✅ Per Pipeline Billing View for Azure Data Factory
- ✅ TTL Setting in Azure IR to reduce cluster spin up time for dataflows
- ✅ Create Alert rules in ADF for Pipeline or activity Failures
- ✅ Pipeline return value in Set variable in Azure Data Factory
- ✅ Pagination rules - When API response have URL for next page
- ✅ Copy activity - Pagination rules - Variables in Query Parameters
- ✅ CDC(change data capture) for SQL Source in Mapping data flows
- ✅ CDC (change data capture) Resource in Azure Data Factory
- ✅ Deactivate an Activity in Azure Data Factory
- ✅ Managed Virtual Integration Runtime in Azure Data Factory
- ✅ TTL in Managed Vnet IR in Azure Data Factory

Azure Databricks 02

Introduction to Databricks & Lakehouse:

- ✅ What is Data Lakehouse & Databricks Data Intelligence Platform
- ✅ Benefits of Databricks Lakehouse Architecture
- ✅ High-Level Architecture of Databricks (Control Plane & Data Plane)
- ✅ Roles & Responsibilities in Databricks

Getting Started with Databricks:

- ✅ Setting up Databricks on Azure
- ✅ Creating Your First Databricks Workspace
- ✅ Accessing the Databricks Free Trial
- ✅ Databricks Account Console Overview
- ✅ User & Workspace Management in Databricks

Databricks Workspace Essentials:

- ✅ Introduction to Workspace & Notebooks
- ✅ Cell Magic Commands, Comments, and Variables
- ✅ Version History and Collaboration Features

Databricks & Azure Integration:

- ✅ Working with Azure Managed Storage
- ✅ Databricks Clusters using Azure VMs

Unity Catalog & Governance:

- ✅ Introduction to Unity Catalog and Governance
- ✅ Unity Catalog Object Model & Metastore
- ✅ Legacy Hive Metastore vs Unity Catalog
- ✅ Managed Tables vs External Tables
- ✅ Databricks File System (DBFS) Explained
- ✅ Setting Up Unity Catalog and Metastore
- ✅ Catalogs, External Locations & Storage
- ✅ Schemas & Managed Data Locations in Unity Catalog
- ✅ Comparison of Tables: Unity Catalog vs Legacy
- ✅ Using UNDROP Feature in Databricks

Delta Lake on Databricks:

- ✅ Deep & Shallow Clones in Delta Tables
- ✅ Temporary & Permanent Views
- ✅ MERGE, UPSERT, and SCD1 in Delta Tables
- ✅ Soft Deletes with Merge (Incremental Loads)
- ✅ Liquid Clustering & Deletion Vectors
- ✅ Optimizing & Clustering Delta Tables

Volumes, DBUtils & Utilities:

- ✅ Databricks Volumes (Managed & External)
- ✅ Working with Files in Volumes
- ✅ DBUTILS Overview & File System Utilities
- ✅ Creating Interactive Widgets in Notebooks

Job Orchestration & Schedulinge:

- ✅ Orchestrating Notebook Jobs & Parameters
- ✅ Chaining Notebooks & Parameter Passing
- ✅ Cluster Types: All-Purpose vs Job
- ✅ Cluster Access Modes, Policies & Permissions
- ✅ Creating Custom Cluster Policies & Instance Pools
- ✅ Warm Instance Pools for Faster Startups

Workflows & Automations:

- ✅ Creating and Managing Jobs, Tasks & Workflows
- ✅ Conditional Logic in Workflows (If, Else, For Each)
- ✅ Re-running Jobs and Handling Failures

Data Ingestion & COPY INTO:

- ✅ COPY INTO Command & Metadata
- ✅ Idempotent Data Pipelines
- ✅ Exactly Once Processing

Auto Loader in Databricks:

- ✅ Auto Loader Overview & Schema Evolution
- ✅ File Detection Modes in Auto Loader

Medallion Architecture & DLT:

- ✅ Medallion Architecture: Bronze, Silver, Gold Layers
- ✅ Delta Live Tables (DLT) Introduction – Part 1
- ✅ Streaming Tables & Materialized Views
- ✅ DLT Internals & Incremental Load – Part 2
- ✅ Column Modification, Rename, and Lineage
- ✅ DLT Append Flow with AutoLoader
- ✅ Passing Parameters & Dynamic Table Creation
- ✅ SCD1 & SCD2 with DLT | CDC & Backload Strategy
- ✅ Deletion & Truncation in SCD Tables
- ✅ Data Quality Rules & Expectations in DLT
- ✅ Monitoring & Observability for DLT
- ✅ Full Refresh & Truncate Load in DLT
- ✅ Workflow File Arrival Triggers & Scheduling

Security & Access Management:

- ✅ Secret Management & Secret Scopes
- ✅ Integration with Azure Key Vault
- ✅ User & Group Management in Unity Catalog
- ✅ Object-Level Security & Privileges

Data Governance in Unity Catalog:

- ✅ Functions in Unity Catalog (SCALAR & TABLE UDFs)
- ✅ Row-Level Security for Sensitive Data
- ✅ Column-Level Masking for Sensitive Fields
- ✅ Catalog Binding to Workspace & Access Restrictions

Data Sharing & Federation:

- ✅ Delta Sharing & Databricks Express Edition
- ✅ Sharing Data Outside the Organization
- ✅ Serverless Compute: Architecture & Use Cases

SQL, Warehousing & Performance:

- ✅ Data Warehousing using Databricks SQL (DBSQL)
- ✅ SQL Warehouses & Query Performance Tuning
- ✅ Streaming Tables & Materialized Views in DBSQL
- ✅ Query Scheduling & Background Refresh
- ✅ Query Federation & External Catalogs

Monitoring & Alerts:

- ✅ Creating and Managing SQL Alerts
- ✅ Configuring Alert Schedules & Destinations

Python 03

Python Basics:

- ✅ Introduction to Python
- ✅ Installing Python on Windows
- ✅ Getting Started with Visual Studio Code
- ✅ Python Indentation Rules
- ✅ Comments in Python

Python Basics:

- ✅ Variables in Python
- ✅ Global vs Local Variables
- ✅ Data Types Overview
- ✅ Numbers in Python
- ✅ Type Conversion and Type Casting

Working with Strings:

- ✅ Strings in Python
- ✅ Format Strings
- ✅ Escape Characters
- ✅ String Methods

Boolean and Operators:

- ✅ Booleans in Python
- ✅ Operators in Python
- ✅ Arithmetic Operators
- ✅ Assignment Operators
- ✅ Comparison Operators
- ✅ Logical Operators
- ✅ Identity Operators
- ✅ Membership Operators
- ✅ Bitwise Operators: AND (&), OR (|), XOR (^)
- ✅ Bitwise Operators: NOT (~), Left Shift, Right Shift

Collections:

- ✅ Lists in Python
- ✅ Tuples in Python
- ✅ Sets in Python
- ✅ Dictionaries in Python

Control Flow:

- ✅ If...Else Statements
- ✅ While Loops
- ✅ For Loops

Functions:

- ✅ Defining and Using Functions
- ✅ Recursion in Python
- ✅ Lambda (Anonymous) Functions

Object-Oriented Programming:

- ✅ Classes and Objects
- ✅ The __init__() Method
- ✅ Inheritance in Python
- ✅ Iterators

Scope and Modules:

- ✅ Variable Scope (LEGB Rule)
- ✅ Modules in Python

Dates, Math, and JSON:

- ✅ Working with Dates
- ✅ Math Operations
- ✅ JSON Handling

Python Tools and Packages:

- ✅ Using PIP in Python
- ✅ Packaging Python Code as a Wheel File
- ✅ Understanding the __name__ Variable

Error Handling and User Input:

- ✅ Exception Handling with Try...Except
- ✅ Taking User Input with input()

File Handling:

- ✅ File Handling with open()
- ✅ Reading Files
- ✅ Writing/Creating Files
- ✅ Deleting Files

Built-in Functions:

- ✅ Using the map() Function

SQL 04

Getting Started with SQL:

- ✅ SQL Introduction & Installing SQL Server
- ✅ What is a Database? | Types of Databases
- ✅ Practical Database Setup

Data Manipulation Basics:

- ✅ INSERT Statement in SQL
- ✅ UPDATE Statement in SQL
- ✅ DELETE vs DROP vs TRUNCATE

Constraints in SQL:

- ✅ Introduction to Constraints
- ✅ NOT NULL & UNIQUE Constraints
- ✅ CHECK & DEFAULT Constraints
- ✅ PRIMARY KEY Constraint
- ✅ FOREIGN KEY Constraint

Querying and Filtering Data:

- ✅ FILTER and SORT Data in SQL
- ✅ Conditional Statements: AND, OR, IN
- ✅ LIKE Operator in SQL
- ✅ DISTINCT Keyword
- ✅ TOP or LIMIT Clause
- ✅ COALESCE Function

Aggregation and Grouping:

- ✅ Aggregate Functions in SQL (SUM, AVG, COUNT, etc.)
- ✅ GROUP BY Clause
- ✅ HAVING Clause

Joins and Set Operations:

- ✅ SQL Joins Explained with Examples (INNER, LEFT, RIGHT, FULL)
- ✅ UNION vs UNION ALL

Schema and Object Management:

- ✅ ALTER Command in SQL
- ✅ VIEWS in SQL

Advanced SQL Concepts:

- ✅ Window Functions Overview
- ✅ RANK(), DENSE_RANK(), ROW_NUMBER()
- ✅ LEAD() and LAG() Functions
- ✅ CTE (Common Table Expression) using WITH Clause
- ✅ Subqueries in SQL

Stored Logic and Triggers:

- ✅ Stored Procedures in SQL
- ✅ Triggers with Real-Life Examples

PySpark 05

Introduction to PySpark:

- ✅ What is PySpark?
- ✅ PySpark vs. Spark: Understanding the difference
- ✅ Spark architecture and components
- ✅ Setting up PySpark environment
- ✅ Creating RDDs (Resilient Distributed Datasets)
- ✅ Transformations and actions in RDDs

PySpark DataFrames:

- ✅ Introduction to DataFrames
- ✅ Creating DataFrames from various data sources (CSV, JSON, Parquet)
- ✅ Basic Data Frame operations (filtering, selecting, aggregating)
- ✅ Handling missing data
- ✅ Data Frame joins and unions

PySpark SQL:

- ✅ Introduction to Spark SQL
- ✅ Creating temporary views and global temporary views
- ✅ Executing SQL queries on DataFrames
- ✅ Performance optimization techniques
- ✅ Working with user-defined functions (UDFs)

PySpark MLlib (Machine Learning Library):

- ✅ Introduction to MLlib
- ✅ Data preprocessing and feature engineering
- ✅ Building and evaluating regression models
- ✅ Classification algorithms and evaluation metrics
- ✅ Clustering and collaborative filtering
- ✅ Model selection and tuning

PySpark Streaming:

- ✅ Introduction to Spark Streaming
- ✅ DStream (Discretized Stream) and input sources
- ✅ Windowed operations and stateful transformations
- ✅ Integration with Kafka for real-time data processing

PySpark and Big Data Ecosystem:

- ✅ Overview of Hadoop, HDFS, and YARN
- ✅ Integrating PySpark with Hadoop and Hive
- ✅ PySpark and NoSQL databases (e.g., HBase)
- ✅ Spark on Kubernetes

PySpark Optimization and Best Practices:

- ✅ Understanding Spark's execution plan
- ✅ Performance tuning and optimization techniques
- ✅ Broadcast variables and accumulators
- ✅ PySpark configuration and memory management

Cloud UpSkill

Learning Journey & Certification Path

Understand Cloud & Data Fundamentals

- *Complete AZ-900: Microsoft Azure Fundamentals to grasp basic cloud concepts.*
- *Proceed to DP-900: Microsoft Azure Data Fundamentals to learn about core data services.*

Dive into Data Engineering

- *Enroll in the DP-203: Data Engineering on Microsoft Azure course.*
- *Gain hands -on experience with Azure services like Data Factory , Synapse Analytics, and Databricks.*

Practice and Prepare

- *Utilize Microsoft Learn's self-paced modules and labs.*
- *Engage in practice assessments to test your knowledge.*

Get Certified

- *Schedule and pass the DP-203 certification exam to earn the Azure Data Engineer Associate credential.*

Advance Your Career

- *Leverage your certification to pursue roles such as Data Engineer, Data Analyst, or Solutions Architect.*

Contact US

Prestige Sky Tech, Financial District, Nanakramguda, Telangana 500032

+91 -7892882483
+91 -7396053003
+91 -7780637278
+91 - 9398162283

Info@cloudupskill.com