

Azure Data Engineer Placement oriented training

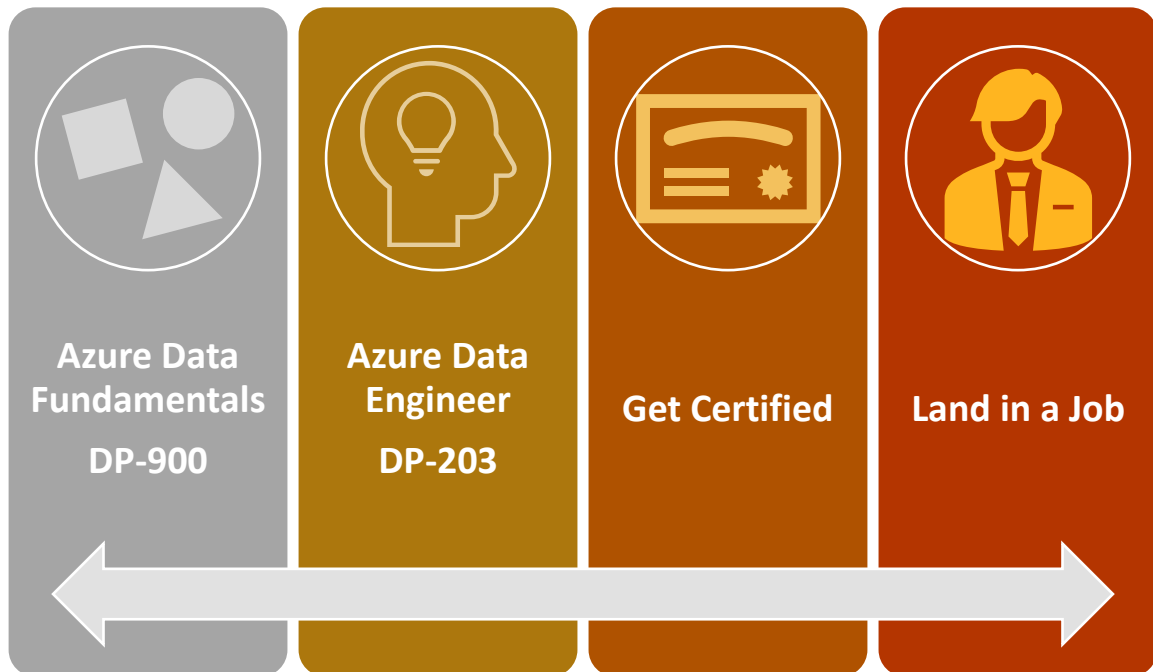
From **KloudCourse** _Venkat Kurela

Are you looking to advance your skills as a data professional? The Azure Data Engineer course is the perfect opportunity. With this course, you'll acquire the skills to optimize data storage, manage data processing, and implement data security using Azure services like Azure Data Factory, Azure Databricks, and Azure Synapse Analytics.

With the Azure Data Engineer course, you'll learn how to integrate data from various sources, transform it into meaningful insights, and visualize it in a clear and concise manner. You'll also become proficient in data warehousing, data modeling, and data pipelining, all of which are essential skills for data engineers in today's fast-paced digital landscape.

At KloudCourse we believe in **"your success is our success"** to make it happen we prepare course outline with all success ingredients needed, this course is combination of DP-900, DP-203, Certification, CV preparation, mock-Interview, and close group referral program.

Don't miss out on this opportunity to advance your career in the field of data engineering. Join the Azure Data Engineer course today and gain the skills and knowledge you need to succeed in this exciting and rapidly growing field.



Describe core data concepts (25—30%)

Describe ways to represent data

- Describe features of structured data
- Describe features of semi-structured
- Describe features of unstructured data

Identify options for data storage

- Describe common formats for data files
- Describe types of databases

Describe common data workloads

- Describe features of transactional workloads
- Describe features of analytical workloads

Identify roles and responsibilities for data workloads

- Describe responsibilities for database administrators
- Describe responsibilities for data engineers
- Describe responsibilities for data analysts

Identify considerations for relational data on Azure (20—25%)

Describe relational concepts

- Identify features of relational data
- Describe normalization and why it is used
- Identify common structured query language (SQL) statements
- Identify common database objects

Describe relational Azure data services

- Describe the Azure SQL family of products including Azure SQL Database, Azure SQL Managed Instance, and SQL Server on Azure Virtual Machines
- Identify Azure database services for open-source database systems

Describe considerations for working with non-relational data on Azure (15—20%)

Describe capabilities of Azure storage

- Describe Azure Blob storage
- Describe Azure File storage
- Describe Azure Table storage

Describe capabilities and features of Azure Cosmos DB

- Identify use cases for Azure Cosmos DB
- Describe Azure Cosmos DB APIs

Describe an analytics workload on Azure (25—30%)

Describe common elements of large-scale analytics

- Describe considerations for data ingestion and processing
- Describe options for analytical data stores
- Describe Azure services for data warehousing, including Azure Synapse Analytics, Azure Databricks, Azure HDInsight, and Azure Data Factory

Describe consideration for real-time data analytics

- Describe the difference between batch and streaming data
- Describe technologies for real-time analytics including Azure Stream Analytics, Azure Synapse Data Explorer, and Spark structured streaming

Describe data visualization in Microsoft Power BI

- Identify capabilities of Power BI
- Describe features of data models in Power BI
- Identify appropriate visualizations for data

Design and implement data storage (15–20%)

Implement a partition strategy

- Implement a partition strategy for files
- Implement a partition strategy for analytical workloads
- Implement a partition strategy for streaming workloads
- Implement a partition strategy for Azure Synapse Analytics
- Identify when partitioning is needed in Azure Data Lake Storage Gen2

Design and implement the data exploration layer

- Create and execute queries by using a compute solution that leverages SQL serverless and Spark cluster
- Recommend and implement Azure Synapse Analytics database templates
- Push new or updated data lineage to Microsoft Purview
- Browse and search metadata in Microsoft Purview Data Catalog

Develop data processing (40–45%)

Ingest and transform data

- Design and implement incremental loads
- Transform data by using Apache Spark
- Transform data by using Transact-SQL (T-SQL)
- Ingest and transform data by using Azure Synapse Pipelines or Azure Data Factory
- Transform data by using Azure Stream Analytics
- Cleanse data
- Handle duplicate data
- Handle missing data
- Handle late-arriving data



- Split data
- Shred JSON
- Encode and decode data
- Configure error handling for a transformation
- Normalize and denormalize data
- Perform data exploratory analysis

Develop a batch processing solution

- Develop batch processing solutions by using Azure Data Lake Storage, Azure Databricks, Azure Synapse Analytics, and Azure Data Factory
- Use PolyBase to load data to a SQL pool
- Implement Azure Synapse Link and query the replicated data
- Create data pipelines
- Scale resources
- Configure the batch size
- Create tests for data pipelines
- Integrate Jupyter or Python notebooks into a data pipeline
- Upsert data
- Revert data to a previous state
- Configure exception handling
- Configure batch retention
- Read from and write to a delta lake

Develop a stream processing solution

- Create a stream processing solution by using Stream Analytics and Azure Event Hubs
- Process data by using Spark structured streaming
- Create windowed aggregates
- Handle schema drift
- Process time series data
- Process data across partitions
- Process within one partition
- Configure checkpoints and watermarking during processing
- Scale resources
- Create tests for data pipelines
- Optimize pipelines for analytical or transactional purposes
- Handle interruptions
- Configure exception handling
- Upsert data
- Replay archived stream data



Manage batches and pipelines

- Trigger batches, Handle failed batch loads
- Validate batch loads
- Manage data pipelines in Azure Data Factory or Azure Synapse Pipelines
- Schedule data pipelines in Data Factory or Azure Synapse Pipelines
- Implement version control for pipeline artifacts
- Manage Spark jobs in a pipeline

Secure, monitor, and optimize data storage and data processing (30–35%)

Implement data security

- Implement data masking
- Encrypt data at rest and in motion
- Implement row-level and column-level security
- Implement Azure role-based access control (RBAC)
- Implement POSIX-like access control lists (ACLs) for Data Lake Storage Gen2
- Implement a data retention policy
- Implement secure endpoints (private and public)
- Implement resource tokens in Azure Databricks
- Load a DataFrame with sensitive information
- Write encrypted data to tables or Parquet files
- Manage sensitive information

Monitor data storage and data processing

- Implement logging used by Azure Monitor
- Configure monitoring services
- Monitor stream processing
- Measure performance of data movement
- Monitor and update statistics about data across a system
- Monitor data pipeline performance
- Measure query performance
- Schedule and monitor pipeline tests
- Interpret Azure Monitor metrics and logs
- Implement a pipeline alert strategy

Optimize and troubleshoot data storage and data processing

- Compact small files
- Handle skew in data
- Handle data spill
- Optimize resource management
- Tune queries by using indexers
- Tune queries by using cache
- Troubleshoot a failed Spark job
- Troubleshoot a failed pipeline run, including activities executed in external services

<https://kloudcourse.in/>

Learn by doing is the new way of learning.



hands-on labs play a critical role in modern education and training programs. They provide a dynamic, interactive learning environment that allows learners to actively engage with the subject matter, develop critical thinking and problem-solving skills, retain information better, and prepare for real-world scenarios. We at KloudCourse have given high priority in incorporating hands-on labs into curriculum, our trainers help learners develop the practical skills and knowledge they need to succeed.

Below are the list of Hands-on labs that will be performed, this is sample outline of activities but more Realtime scenarios will be performed during the course.

Explore Azure relational database services, SQL, PostgreSQL, MySQL

1. Provision an Azure SQL Database resource
2. Provision an Azure Database for PostgreSQL resource
3. Explore Azure Database for MySQL

Explore Azure Storage

1. Provision an Azure Storage account
2. Explore blob storage
3. Explore Azure Data Lake Storage Gen2
4. Explore Azure Files
5. Explore Azure Tables

Explore Azure Cosmos DB

1. Create a Cosmos DB account
2. Create a sample database
3. View and create items
4. Query the database

Explore data analytics in Azure with Azure Synapse Analytics

1. Provision an Azure Synapse Analytics workspace
2. Ingest data
3. Use a SQL pool to analyze data
4. Use a Spark pool to analyze data
5. Delete Azure resources

Explore Azure Stream Analytics

1. Create Azure resources
2. Explore the Azure resources
3. Use the resources to analyze streaming data

Explore Spark Streaming in Azure Synapse Analytics

1. Provision a Synapse Analytics workspace
2. Create a Spark pool
3. Explore stream processing
4. Delete Azure resources

<https://kloudcourse.in/>

Learn by doing is the new way of learning.



Explore Azure Synapse Data Explorer

1. Provision a Synapse Analytics workspace
2. Create a Data Explorer pool
3. Create a database and ingest data
4. Use Kusto query language to query the table in Synapse Studio
5. Delete Azure resources

Explore fundamentals of data visualization with Power BI

1. Import data
2. Explore a data model
3. Create a report

DP-203

Explore Azure Synapse Analytics

1. Provision an Azure Synapse Analytics workspace
2. Explore Synapse Studio
3. Ingest data with a pipeline
4. Use a serverless SQL pool to analyze data
5. Use a Spark pool to analyze data
6. Use a dedicated SQL pool to query a data warehouse
7. Explore data with a Data Explorer pool
8. Delete Azure resources

Query files using a serverless SQL pool

1. Provision an Azure Synapse Analytics workspace
2. Query data in files
3. Access external data in a database
4. Visualize query results
5. Delete Azure resources

Transform files using a serverless SQL pool

1. Provision an Azure Synapse Analytics workspace
2. Query data in files
3. Transform data using CREATE EXTERNAL TABLE AS SELECT (CETAS) statements
4. Encapsulate data transformation in a stored procedure
5. Delete Azure resources

Analyze data in a lake database

1. Provision an Azure Synapse Analytics workspace
2. Create a lake database
3. Create a table
4. Create a table from a database template
5. Create a table from existing data
6. Work with lake database tables
7. Delete Azure resources

<https://kloudcourse.in/>

Learn by doing is the new way of learning.



Analyze data with Spark

1. Provision an Azure Synapse Analytics workspace
2. Query data in files
3. Analyze data in a dataframe
4. Query data using Spark SQL
5. Visualize data with Spark
6. Delete Azure resources

Transform data using Spark in Synapse Analytics

1. Provision an Azure Synapse Analytics workspace
2. Use a Spark notebook to transform data
3. Delete Azure resources

Use Delta Lake with Spark in Azure Synapse Analytics

1. Provision an Azure Synapse Analytics workspace
2. Create delta tables
3. Create catalog tables
4. Use delta tables for streaming data
5. Query a delta table from a serverless SQL pool
6. Delete Azure resources

Explore a data warehouse

1. Provision an Azure Synapse Analytics workspace
2. Explore the data warehouse schema
3. Query the data warehouse tables
4. Challenge - Analyze reseller sales
5. Delete Azure resources

Load Data into a Relational Data Warehouse

1. Provision an Azure Synapse Analytics workspace
2. Prepare to load data
3. Load data warehouse tables
4. Perform post-load optimization
5. Delete Azure resources

Check for skewed data and space usage

1. Analyze the space used by tables
2. Use a more advanced approach to understand table space usage



View column store storage details

Compare storage requirements between optimal and sub-optimal column data types

Improve the execution plan of a query with a materialized view

Optimize a delete operation

Manage authorization through column and row level security

Build a data pipeline in Azure Synapse Analytics

1. Provision an Azure Synapse Analytics workspace
2. View source and destination data stores
3. Implement a pipeline
4. Debug the Data Flow
5. Publish and run the pipeline
6. Delete Azure resources

Use an Apache Spark notebook in a pipeline

1. Provision an Azure Synapse Analytics workspace
2. Run a Spark notebook interactively
3. Run the notebook in a pipeline
4. Delete Azure resources

Use Azure Synapse Link for Azure Cosmos DB

1. Provision Azure resources
2. Configure Synapse Link in Azure Cosmos DB
3. Configure Synapse Link in Azure Synapse Analytics
4. Query Azure Cosmos DB from Azure Synapse Analytics
5. Delete Azure resources

Use Azure Synapse Link for SQL

1. Provision Azure resources
2. Configure Azure SQL Database
3. Explore the transactional database
4. Configure Azure Synapse Link
5. Delete Azure resources

Get started with Azure Stream Analytics

1. Provision Azure resources
2. View the streaming data source
3. Create an Azure Stream Analytics job
4. Create an input for the event stream
5. Create an output for the blob store
6. Create a query
7. Run the streaming job
8. Delete Azure resources

<https://kloudcourse.in/>

Learn by doing is the new way of learning.



Ingest streaming data into Azure Synapse Analytics

1. Provision Azure resources
2. Ingest streaming data into a dedicated SQL pool
3. Summarize streaming data in a data lake
4. Delete Azure resources

Create a realtime report with Azure Stream Analytics and Microsoft Power BI

1. Provision Azure resources
2. Create a Power BI workspace
3. Use Azure Stream Analytics to process streaming data
4. Visualize the streaming data in Power BI
5. Delete resources

Use Microsoft Purview with Azure Synapse Analytics

1. Provision Azure resources
2. Catalog Azure Synapse Analytics data assets in Microsoft Purview
3. Integrate Microsoft Purview with Azure Synapse Analytics
4. Delete Azure resources

Explore Azure Databricks

1. Provision an Azure Databricks workspace
2. Create a cluster
3. Use Spark to analyze a data file
4. Create and query a database table
5. Delete Azure Databricks resources

Use Spark in Azure Databricks

1. Provision an Azure Databricks workspace
2. Create a cluster
3. Explore data using a notebook
4. Delete Azure Databricks resources

Use Delta Lake in Azure Databricks

1. Provision an Azure Databricks workspace
2. Create a cluster
3. Explore data using a notebook
4. Delete Azure Databricks resources

Use a SQL Warehouse in Azure Databricks

1. Provision an Azure Databricks workspace
2. View and start a SQL Warehouse
3. Create a database
4. Create a table
5. Create a query
6. Create a dashboard
7. Delete Azure Databricks resources

<https://kloudcourse.in/>

Learn by doing is the new way of learning.



Automate an Azure Databricks Notebook with Azure Data Factory

1. Provision Azure resources
2. Import a notebook
3. Enable Azure Databricks integration with Azure Data Factory
4. Use a pipeline to run the Azure Databricks notebook
5. Delete Azure Databricks resources