Azure Data Factory Training Path for DP-200 and DP-201

This path will prepare you for exams DP-200 and DP-201 for implementing and designing Azure data solutions, enabling you to design and perform data management, monitoring, security and privacy using the complete Azure data services stack. Further, you will learn the below list of topics:

1. Implement Azure Data Solution (DP 200)

Implementing a data storage solution

Managing and developing data processing

Monitoring and optimizing data solutions

2. Design Azure Data Solutions (DP 201)

Design an Azure Data Storage Solution

Design a data processing solution

Data security and compliance design

Why should I learn Azure Stream Analytics in this DP-200 certification?

Using Azure Stream Analytics, you can manage real-time event processing applications by scaling it. It promotes greater performance through partitioning, so complicated queries can be parallelized and run on many stream nodes.

Will I learn Azure Databricks in this Azure Data Engineering training?

With Azure Databricks, you can get the most advanced version of Apache Spark, which will enable you to seamlessly integrate it with the open-sourced libraries. Also, with Azure's inherent scalability that surpasses every other cloud services, you can easily update the clusters in the Apache Spark.

What is Azure Cosmos DB? Why is it important for getting DP-200 certification?

Azure Cosmos DB leverages the Jupyter notebooks and Apache Spark to reduce the time to insights by collecting and serving data and performing analysis on local database copies in Azure regions. In this DP-200 certification, you will cover this topic in detail.

Who should take up this best Azure Data Factory training course?

This Microsoft Azure Data Factory certification is ideal for candidates who are looking to start their career or already working in the following roles:

Data Engineers

Data Architects

Data Scientists

Data Analysts

Further, candidates who design analytics solutions and build operationalized solutions on Azure and who are familiar with the features and capabilities of batch data processing, real-time processing, and operationalization technologies, etc., can also opt for this Azure Data Engineering training course.

What are the prerequisites for taking up this best Azure Data Factory training online?

Relevant work experience in Data Engineering issues with Azure SQL Data Warehouse, Azure Data Lake, Azure Data Factory, and Azure Stream Analytics

Module 01 - Non-Relational Data Stores and Azure Data Lake Storage

https://docs.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-overview

https://azure.microsoft.com/en-us/overvxiew/nosql-database/

1.1 Document data stores

1.2 Columnar data stores

1.3 Key/value data stores

1.4 Graph data stores

1.5 Time series data stores

1.6 Object data stores

1.7 External index

1.8 Why NoSQL or Non-Relational DB?

1.9 When to Choose NoSQL or Non-Relational DB?

https://docs.microsoft.com/en-us/dotnet/architecture/cloud-native/relational-vs-nosql-data

Best Uses

Scenarios

1.10 Azure Data Lake Storage

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction

Definition

Azure Data Lake-Key Components

How it stores data?

Azure Data Lake Storage Gen2

Why Data Lake?

Data Lake Architecture

Module 02 - Data Lake and Azure Cosmos DB

https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-cosmos-db

2.1 Data Lake Key Concepts

2.2 Azure Cosmos DB

2.3 Why Azure Cosmos DB?

2.4 Azure Blob Storage

2.5 Why Azure Blob Storage?

2.6 Data Partitioning

https://docs.microsoft.com/en-us/azure/architecture/best-practices/data-partitioning

Horizontal partitioning

Vertical partitioning

Functional partitioning

2.7 Why Partitioning Data?

2.8 Consistency Levels in AzureCosmos DB

https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

Semantics of the five-consistency level

Hands-on:

1. Load Data fromAmazonS3 to ADLS Gen2 with Data Factory

https://docs.microsoft.com/en-us/azure/data-factory/connector-amazon-simple-storage-service

https://docs.microsoft.com/en-us/azure/data-factory/solution-template-migration-s3-azure

2. Working with Azure Cosmos DB

Module 03 - Relational Data Stores

https://docs.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-overview

https://docs.microsoft.com/en-us/learn/paths/work-with-relational-data-in-azure/

3.1 Introduction to Relational Data Stores

3.2 Azure SQL Database

Deployment Models

Service Tiers

Hands-on:

1. Create a Single Database Using Azure Portal

2. Create a managed instance

3. Create an elastic pool

3.3 Why SQL Database Elastic Pool?

https://docs.microsoft.com/en-us/azure/azure-sql/database/elastic-pool-overview

Hands-on:

1. Create a SQL virtual machine

2. Configure active geo-replication for Azure SQL Database in the Azure portal and initiate failover.

Module 04 - Why Azure SQL?

https://docs.microsoft.com/bs-cyrl-ba/azure/azure-sql/database/sql-database-paas-overview

4.1 Azure SQL Security Capabilities

4.2 High-Availability and Azure SQL Database

Standard Availability Model

Premium Availability Model

4.3 Azure Database for MySQL

Hands-on:

1. Design an Azure Database for MySQL database using the Azure portal

2. Connect using MySQL Workbench

https://docs.microsoft.com/en-us/azure/mysql/connect-workbench

4.4 Azure Database for PostgreSQL

https://docs.microsoft.com/en-us/azure/postgresql/overview

Hands-on:

1. Design an Azure Database for PostgreSQL – Single Server

4.5 Azure Database For MariaDB

https://docs.microsoft.com/en-us/azure/mariadb/overview

Hands-on:

1. Create an Azure Database for MariaDB server by using the Azure portal

4.6 What is PolyBase?

https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-guide?view=sql-server-ver15

Why PolyBase?

4.7 What is Azure Synapse Analytics (formerly SQL DW)?

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-overview-what-is#:~:text=Synapse%20SQL%20pool%20refers%20to,Data%20Warehousing%20Units%20(DWU).

SQL Analytics and SQL pool in Azure Synapse

Key component of a big data solution

SQL Analytics MPP architecture components

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/massively-parallel-processing-mpp-architecture

Hands-on:

1. Import Data From Blob Storage to Azure Synapse Analytics by Using PolyBase

https://docs.microsoft.com/en-us/learn/modules/import-data-into-asdw-with-polybase/

Module 05 - Azure Batch

https://docs.microsoft.com/en-us/azure/data-factory/transform-data-using-dotnet-custom-activity

https://docs.microsoft.com/en-us/azure/data-factory/v1/data-factory-data-processing-using-batch

https://docs.microsoft.com/en-us/azure/batch/tutorial-run-python-batch-azure-data-factory

5.1 What is Azure Batch?

5.2 Intrinsically Parallel Workloads

https://docs.microsoft.com/en-us/azure/batch/batch-technical-overview#:~:text=Run%20parallel%20workloads,-Batch%20works%20well&text=Intrinsically%20parallel%20workloads%20are%20those,other%20instances%20of%20the%20application.

5.3 Tightly Coupled Workloads

5.4 Additional Batch Capabilities

5.5 Working of Azure Batch

Hands-on:

1. Run a batch job using Azure Portal

2. Parallel File Processing with Azure Bath using the .NET API

https://docs.microsoft.com/en-us/azure/batch/tutorial-parallel-dotnet

3. Render a Blender Scene using Batch Explorer

https://docs.microsoft.com/en-us/azure/batch/tutorial-rendering-batchexplorer-blender

4. Parallel R Simulation with Azure Batch

https://docs.microsoft.com/en-us/azure/batch/tutorial-r-doazureparallel

Module 06 - Azure Data Factory

6.1 Flow Process of Data Factory

6.2 Why Azure Data Factory

6.3 Integration Runtime in Azure Data Factory

https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime

6.4 Mapping Data Flows

Hands-on:

1. Transform data using Mapping data flows

Module 07 - Azure Data Bricks

7.1 What is Azure Databricks?

7.2 Azure Spark-based Analytics Platform

7.3 Apache Spark in Azure Databricks

https://azure.microsoft.com/en-us/services/databricks/#:~:text=Azure%20Databricks%20provides%20the%20latest,scale%20and%20availability%20of%20Azure.

Hands-on:

1. Run a Spark Job on Azure Databricks using the Azure portal

https://docs.microsoft.com/en-us/azure/databricks/scenarios/quickstart-create-databricks-workspace-portal?tabs=azure-portal

2. ETL Operation by using Azure Databricks

https://docs.microsoft.com/en-us/azure/databricks/scenarios/databricks-extract-load-sql-data-warehouse

3. Stream data into Azure Databricks using Event Hubs

https://docs.microsoft.com/en-us/azure/databricks/scenarios/databricks-stream-from-eventhubs

Module 08 - Azure Stream Analytics

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-manage-job

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-power-bi-dashboard

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-solution-patterns

8.1 Working of Stream Analytics

8.2 Key capabilities and benefits

Hands-on:

1. Analyse phone call data with stream analytics and visualize results in Power BI dashboard

8.3 Stream Analytics Windowing Functions

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions

Tumbling window

Hopping Window

Sliding Window

Session Window

Module 09 - Monitoring & Security

9.1 What is Azure Monitor?

Metrics

Logs

Metrics Vs Logs

9.2 What data does Azure Monitor collect?

https://azure.microsoft.com/en-us/services/monitor/#:~:text=Azure%20Monitor%20collects%20monitoring%20telemetry,optimized%20for%20cost%20and%20performance.

9.3 What can you Monitor?

Insights and Core Solutions

9.4 Alerts in Azure

Flow of Alerts

Key Attributes of an Alert Rule

https://docs.microsoft.com/en-us/azure/azure-monitor/platform/alerts-overview

What can you set alert on?

Manage alerts

Alert States

How to create an alert?

Hands-on:

1. Create, View, and Manage Metric alerts using Azure Monitor

https://docs.microsoft.com/en-us/azure/azure-monitor/platform/alerts-metric

2. Monitor your Azure Data Factory Pipelines proactively with Alerts

https://azure.microsoft.com/fr-fr/blog/create-alerts-to-proactively-monitor-your-data-factory-pipelines/

9.5 Azure Security Logging & Auditing

Types of Logs in Azure

Azure SQL Database Auditing

Server-level vs. Database-level Auditing Policy

Hands-on:

1. Azure SQL Database Auditing

https://docs.microsoft.com/en-us/azure/azure-sql/database/auditing-overview

https://docs.microsoft.com/en-us/azure/azure-sql/database/audit-log-format

What projects I will be working on this Azure Data Factory training?

In this Azure Data Factory Project, you are supposed to automate the transformation of the real-time video list from the YouTube channel. You will be storing multiple files at the dynamic location of Azure Data Lake Store and the same needs to transformed and copied to any data store. The list of the channels should be displayed on PowerBI dynamically.

Project 01: Fetch the list of videos from the attached dataset of YouTube channel with the highest views and likes to promote advertisements on the channel which has maximum traffic.

Topics: Azure Data Factory, Azure Data Lake, Triggers, SQL, Power BI

Highlights:

1.1 Creating Azure Data Factory

https://docs.microsoft.com/en-us/azure/data-factory/quickstart-create-data-factory-portal

https://docs.microsoft.com/en-us/azure/data-factory/tutorial-copy-data-portal

1.2 Creating Pipelines

https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipelines-activities

1.3 Creating a trigger that runs a pipeline on a schedule

https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers

1.4 Transforming data using SQL

https://docs.microsoft.com/en-us/azure/data-factory/connector-sql-server

1.5 Connecting Azure Data Lake to Power BI

Project 02: Working with Azure Data Factory, Data Lake and Azure SQL

Industry: General

Problem Statement: You are working as an Azure Architect for Zendrix Corp. This company is a service-based company and has its major revenue from the sales it makes for its subscription-based service.

The company needs to continuously monitor its lead flow from different countries. This helps them in strategizing how much they need to invest in Ad-Marketing for a particular country, this, in turn, helps them to achieve their desired sales targets.

Currently, the company has to manually synchronize data from their live SQL database to their BI tool, for checking the lead flow from different countries.

The company wants an automated solution, using which they will be able to see a live dashboard of the lead count. You as an Architect have suggested the following things:

Highlights:

2.1 Use of Power BI Heat maps

2.2 Use of Azure SQL instead of On-Premise SQL

2.3 Use of Data Factory to automate the data lifecycle from SQL to the BI tool.

https://www.microsoft.com/en-us/itshowcase/managing-the-data-life-cycle-using-azure-data-factory

https://azure.microsoft.com/fr-fr/blog/accelerate-your-cloud-data-warehouse-with-automation-tools/

Help them achieve the above goals.

Project 03: Identify the videos that get maximum traffic in selected YouTube channels

Industry: Marketing

Problem Statement: Getting the real-time list of maximum traffic fetching videos from YouTube channels to promote advertisements in the same channels (traffic should be considered on a weekly basis)

Description: There is a company 'XYZ Pvt. Ltd' that promotes advertisements in the maximum traffic generating YouTube channels (on a weekly basis) to drive profits. To maximize profitability, the marketing team that manages the posting of advertisements requires an interface using which they can get a real-time list of YouTube channels for promoting advertisements and monitoring the analytics of traffic on those channels.

Objective: As an Azure Data Factory specialist, you are supposed to automate the transformation of the real-time video list from YouTube channels on a weekly basis. This will help the marketing team promote advertisements on the right YouTube videos on targeted channels.

Note: The traffic can be analyzed on various parameters like the number of views, and likes or comments on a particular day. You can get these publicly available data from the YouTube API.

Case Study 01: Non-Relational Data Stores

Problem Statement: Knowledge check of non-relational databases: Categories and where to use them

Topics: NoSQL or Non-Relational Database, Azure Data Lake Storage and its key components.

Highlights:

1.1 Scenarios where you can use NoSQL or Non-Relational Database.

https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/non-relational-data#:~:text=Instead%2C%20non%2Drelational%20databases%20use,type%20of%20data%20being%20stored.&text=The%20term%20NoSQL%20refers%20to,constructs%20to%20query%20the%20data.

1.2 categories of Non-Relational or No SQL databases with relevant Azure services.

1.3 Azure Data Lake Storage and its key components.

https://www.red-gate.com/simple-talk/cloud/cloud-data/azure-data-lakes/#:~:text=The%20data%20lake%20promise&text=The%20three%20primary%20components%2DData,even%20then%2C%20only%20on%20Windows.

Case Study 02: Non-Relational Data Stores

Problem Statement: Copy data from Azure Blob Storage to Azure Data Lake Storage Gen2; Create an Azure Cosmos DB account and Demonstrate adding and removing regions from your Database account; Strategies for Partitioning data; Semantics of consistency levels in Cosmos DB

Topics: Azure Cosmos DB, Azure Data Factory, Blob Storage, Strategies for Partitioning Data, Semantics of consistency levels in Cosmos DB

Highlights:

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction

2.1 Azure Blob Storage

2.2 Azure Data Lake Storage Gen2

2.3 Azure Cosmos DB

2.4 Partitioning data

2.5 Consistency levels

Case Study 03: Relational Data Stores

Problem Statement: Knowledge check of Relational databases: Deployment models in Azure SQL; Create an elastic pool, Azure SQL Security Capabilities; Import Data From Blob Storage to Azure Synapse Analytics by Using PolyBase

Topics: Azure SQL, PolyBase, Azure Synapse Analytics

Highlights:

3.1 Deployment models in Azure SQL

https://cloud.netapp.com/blog/azure-cvo-blg-azure-sql-database-18-options-for-sql-server-on-the-cloud

3.2 Elastic Pool

https://docs.microsoft.com/en-us/azure/azure-sql/database/elastic-pool-manage

3.3 Azure Synapse Analytics

3.4 PolyBase

Case Study 04: Azure Batch, Azure Data Factory

Problem Statement: Working of Azure Batch; Flow Process of Data Factory; Types of Integration Runtime in Azure Data Factory; Transform data using Mapping data flows

Topics: Azure Batch, Data Factory, Integration Runtime, Mapping Data Flows

Highlights:

4.1 Working of Azure Batch

4.2 Integration Runtime in Azure Data Factory

4.3 Transform data using Mapping data flows

https://docs.microsoft.com/en-us/azure/data-factory/tutorial-data-flow

Case Study 05: Azure Data Bricks, Azure Stream Analytics

Problem Statement: ETL Operation by using Azure Databricks; Working of Stream Analytics; Stream Analytics Windowing Functions

Topics: Azure Data Bricks, Azure Stream Analytics, Windowing Functions

Highlights:

5.1 ETL operation by using Azure Databricks

https://docs.microsoft.com/en-us/azure/databricks/scenarios/databricks-extract-load-sql-data-warehouse

5.2 Working of Stream Analytics

5.3 Windowing Functions

Case Study 06: Monitoring & Security

Problem Statement: Create, View, and Manage Metric alerts using Azure Monitor; Azure SQL Database Auditing

Topics: Azure Monitor, Alerts in Azure, Azure Security Logging & Auditing

https://docs.microsoft.com/en-us/azure/security/benchmarks/security-control-logging-monitoring

https://azure.microsoft.com/en-us/resources/videos/security-logging-and-audit-log-collection/

Highlights:

6.1 Azure Monitor

6.2 Alerts

6.3 Azure SQL Database Auditing

https://docs.microsoft.com/en-us/azure/azure-sql/database/auditing-overview