

# Exam DP-200: Implementing an Azure Data Solution – Skills Measured

**This exam was updated on November 24, 2020. Following the current exam guide, we have included a version of the exam guide with Track Changes set to “On,” showing the changes that were made to the exam on that date.**

## Audience Profile

Candidates for this exam are Microsoft Azure data engineers who collaborate with business stakeholders to identify and meet the data requirements to implement data solutions that use Azure data services.

Azure data engineers are responsible for data-related implementation tasks that include provisioning data storage services, ingesting streaming and batch data, transforming data, implementing security requirements, implementing data retention policies, identifying performance bottlenecks, and accessing external data sources.

Candidates for this exam must be able to implement data solutions that use the following Azure services: Azure Cosmos DB, Azure Synapse Analytics (formerly Azure SQL DW), Azure Data Factory, Azure Stream Analytics, Azure Databricks, and Azure Blob storage.

## Skills Measured

NOTE: The bullets that appear below each of the skills measured are intended to illustrate how we are assessing that skill. This list is not definitive or exhaustive.

NOTE: In most cases, exams do NOT cover preview features, and some features will only be added to an exam when they are GA (General Availability).

## Implement Data Storage Solutions (40-45%)

### Implement non-relational data stores

- implement a solution that uses Cosmos DB, Data Lake Storage Gen2, or Blob storage
- implement data distribution and partitions
- implement a consistency model in Cosmos DB
- provision a non-relational data store
- provide access to data to meet security requirements
- implement for high availability, disaster recovery, and global distribution

### Implement relational data stores

- provide access to data to meet security requirements
- implement for high availability and disaster recovery
- implement data distribution and partitions for Azure Synapse Analytics
- implement PolyBase

### **Manage data security**

- implement data masking
- encrypt data at rest and in motion

## **Manage and Develop Data Processing (25-30%)**

### **Develop batch processing solutions**

- develop batch processing solutions by using Data Factory and Azure Databricks
- ingest data by using PolyBase
- implement the integration runtime for Data Factory
- create linked services and datasets
- create pipelines and activities
- create and schedule triggers
- implement Azure Databricks clusters, notebooks, jobs, and autoscaling
- ingest data into Azure Databricks

### **Develop streaming solutions**

- configure input and output
- select the appropriate built-in functions
- implement event processing by using Stream Analytics

## **Monitor and Optimize Data Solutions (30-35%)**

### **Monitor data storage**

- monitor relational and non-relational data stores
- implement Blob storage monitoring
- implement Data Lake Storage Gen2 monitoring
- implement Azure Synapse Analytics monitoring
- implement Cosmos DB monitoring
- configure Azure Monitor alerts
- implement auditing by using Azure Log Analytics

### **Monitor data processing**

- monitor Data Factory pipelines

- monitor Azure Databricks
- monitor Stream Analytics
- configure Azure Monitor alerts
- implement auditing by using Azure Log Analytics

### **Optimize of Azure data solutions**

- troubleshoot data partitioning bottlenecks
- optimize Data Lake Storage Gen2
- optimize Stream Analytics
- optimize Azure Synapse Analytics
- manage the data lifecycle

**The exam guide below shows the changes that were implemented on November 24, 2020.**

## **Audience Profile**

Candidates for this exam are Microsoft Azure data engineers who collaborate with business stakeholders to identify and meet the data requirements to implement data solutions that use Azure data services.

Azure data engineers are responsible for data-related implementation tasks that include provisioning data storage services, ingesting streaming and batch data, transforming data, implementing security requirements, implementing data retention policies, identifying performance bottlenecks, and accessing external data sources.

Candidates for this exam must be able to implement data solutions that use the following Azure services: Azure Cosmos DB, ~~Azure SQL Database~~, Azure Synapse Analytics (formerly Azure SQL DW), ~~Azure Data Lake Storage~~, Azure Data Factory, Azure Stream Analytics, Azure Databricks, and Azure Blob storage.

## **Skills Measured**

NOTE: The bullets that appear below each of the skills measured are intended to illustrate how we are assessing that skill. This list is not definitive or exhaustive.

NOTE: In most cases, exams do NOT cover preview features, and some features will only be added to an exam when they are GA (General Availability).

## **Implement Data Storage Solutions (40-45%)**

### **Implement non-relational data stores**

- implement a solution that uses Cosmos DB, Data Lake Storage Gen2, or Blob storage
- implement data distribution and partitions

- implement a consistency model in Cosmos DB
- provision a non-relational data store
- provide access to data to meet security requirements
- implement for high availability, disaster recovery, and global distribution

### **Implement relational data stores**

- provide access to data to meet security requirements
- implement for high availability and disaster recovery
- implement data distribution and partitions for Azure Synapse Analytics
- implement PolyBase

### **Manage data security**

- implement data masking
- encrypt data at rest and in motion

## **Manage and Develop Data Processing (25-30%)**

### **Develop batch processing solutions**

- develop batch processing solutions by using Data Factory and Azure Databricks
- ingest data by using PolyBase
- implement the integration runtime for Data Factory
- create linked services and datasets
- create pipelines and activities
- create and schedule triggers
- implement Azure Databricks clusters, notebooks, jobs, and autoscaling
- ingest data into Azure Databricks

### **Develop streaming solutions**

- configure input and output
- select the appropriate built-in functions
- implement event processing by using Stream Analytics

## **Monitor and Optimize Data Solutions (30-35%)**

### **Monitor data storage**

- monitor relational and non-relational data [storagesources](#)
- implement Blob storage monitoring
- implement Data Lake Storage [Gen2](#) monitoring
- implement Azure Synapse Analytics monitoring

- implement Cosmos DB monitoring
- configure Azure Monitor alerts
- implement auditing by using Azure Log Analytics

### **Monitor data processing**

- monitor Data Factory pipelines
- monitor Azure Databricks
- monitor Stream Analytics
- configure Azure Monitor alerts
- implement auditing by using Azure Log Analytics

### **Optimize of Azure data solutions**

- troubleshoot data partitioning bottlenecks
- optimize Data Lake Storage [Gen2](#)
- optimize Stream Analytics
- optimize Azure Synapse Analytics
- manage the data lifecycle