

Exam DP-201: Designing an Azure Data Solution – Skills Measured

This exam was updated on July 31, 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 design data solutions that use Azure data services.

Azure data engineers are responsible for data-related design tasks that include designing Azure data storage solutions that use relational and non-relational data stores, batch and real-time data processing solutions, and data security and compliance solutions.

Candidates for this exam must design data solutions that use the following Azure services: Azure Cosmos DB, [Azure SQL Database](#), Azure Synapse Analytics, 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).

Design Azure Data Storage Solutions (40-45%)

Recommend an Azure data storage solution based on requirements

- choose the correct data storage solution to meet the technical and business requirements
- choose the partition distribution type

Design non-relational cloud data stores

- design data distribution and partitions
- design for scale (including multi-region, latency, and throughput)
- design a solution that uses Cosmos DB, Data Lake Storage Gen2, or Blob storage
- select the appropriate Cosmos DB API
- design a disaster recovery strategy

- design for high availability

Design relational cloud data stores

- design data distribution and partitions
- design for scale (including latency and throughput)
- design a solution that uses Azure Synapse Analytics
- design a disaster recovery strategy
- design for high availability

Design Data Processing Solutions (25-30%)

Design batch processing solutions

- design batch processing solutions that use Data Factory and Azure Databricks
- identify the optimal data ingestion method for a batch processing solution
- identify where processing should take place, such as at the source, at the destination, or in transit

Design real-time processing solutions

- design for real-time processing by using Stream Analytics and Azure Databricks
- design and provision compute resources

Design for Data Security and Compliance (25-30%)

Design security for source data access

- plan for secure endpoints (private/public)
- choose the appropriate authentication mechanism, such as access keys, shared access signatures (SAS), and Azure Active Directory (Azure AD)

Design security for data policies and standards

- design data encryption for data at rest and in transit
- design for data auditing and data masking
- design for data privacy and data classification
- design a data retention policy
- plan an archiving strategy
- plan to purge data based on business requirements

The exam guide below shows the changes that were implemented on July 31, 2020.

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).

Design Azure Data Storage Solutions (40-45%)

Recommend an Azure data storage solution based on requirements

- choose the correct data storage solution to meet the technical and business requirements
- choose the partition distribution type

Design non-relational cloud data stores

- design data distribution and partitions
- design for scale (including multi-region, latency, and throughput)
- design a solution that uses Cosmos DB, Data Lake Storage Gen2, or Blob storage
- select the appropriate Cosmos DB API
- design a disaster recovery strategy
- design for high availability

Design relational cloud data stores

- design data distribution and partitions
- design for scale (including ~~multi-region~~, latency, and throughput)
- design a solution that uses ~~SQL Database and~~ Azure Synapse Analytics
- design a disaster recovery strategy
- design for high availability

Design Data Processing Solutions (25-30%)

Design batch processing solutions

- design batch processing solutions that use Data Factory and Azure Databricks
- identify the optimal data ingestion method for a batch processing solution
- identify where processing should take place, such as at the source, at the destination, or in transit

Design real-time processing solutions

- design for real-time processing by using Stream Analytics and Azure Databricks ~~and~~ ~~Azure Data Explorer~~

- design and provision compute resources

Design for Data Security and Compliance (25-30%)

Design security for source data access

- plan for secure endpoints (private/public)
- choose the appropriate authentication mechanism, such as access keys, shared access signatures (SAS), and Azure Active Directory (Azure AD)

Design security for data policies and standards

- design data encryption for data at rest and in transit
- design for data auditing and data masking
- design for data privacy and data classification
- design a data retention policy
- plan an archiving strategy
- plan to purge data based on business requirements