

Microsoft Certified: Azure Data Engineer Associate – Skills Measured

Implement data storage solutions

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

- configure elastic pools
- configure geo-replication
- provide access to data to meet security requirements
- implement for high availability, disaster recovery, and global distribution
- 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

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 windowing functions
- implement event processing by using Stream Analytics
- ingest and query streaming data with Azure Data Explorer

Monitor and optimize data solutions

Monitor data storage

- monitor relational and non-relational data sources
- implement Blob storage monitoring
- implement Data Lake Storage monitoring
- implement SQL Database monitoring
- implement Azure Synapse Analytics monitoring
- implement Cosmos DB monitoring
- implement Azure Data Explorer 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 Azure data solutions

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

Design Azure data storage solutions

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 SQL Data Warehouse
- design a disaster recovery strategy
- design for high availability

Design data processing solutions

Design batch processing solutions

- design batch processing solutions by using 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

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