

DP-200

Implementing an Azure Data Solution

Exam number: DP-200

Exam title: Implementing an Azure Data Solution

Language(s) this exam will be available in: English

Audience (IT professionals, Developers, Information workers, etc.): IT professionals

Technology: Microsoft Azure

Exam provider (VUE, Certiport, or both): VUE

Exam Design

Note: This document shows tracked changes that are effective as of June 21, 2019.

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.

~~Candidates for this exam are Azure data engineers who are responsible for data-data-related 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. ingesting, egressing, and transforming data from multiple sources using various services and tools. The Azure data engineer collaborates with business stakeholders to identify and meet data requirements while designing and implementing the management, monitoring, security, and privacy of data using the full stack of Azure services to satisfy business needs.~~

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

Skills measured

Note: This document shows tracked changes that are effective as of June 21, 2019.

Implement Data Storage Solutions (2540-3045%)

~~Implement Azure cloud data warehouses~~

may include but is not limited to:

- ~~• Design Data Lake architecture~~
- ~~• Design the data schema~~
- ~~• Provision the data warehouse~~

~~Implement No SQL Databases~~non-relational data stores

may include but is not limited to:

- ~~• Implement a solution that uses Cosmos DB, Data Lake Storage Gen2, or Blob storage~~
- Manage~~Implement~~ data distribution and partitions
- ~~• Select the database platform~~
- ~~• Model data storage based on use cases~~
- ~~• Select storage types~~
- ~~• Provision storage accounts~~
- ~~• Provision Data Lake storage~~
- ~~• Integrate WebHDFS applications with Data Lake Storage~~
- Provision in~~Implement a consistency model in~~ CosmosDB
- Provision a non-relational data store
- Provide access to data to meet security requirements
- Implement for high availability, disaster recovery, and global distribution

~~Implement Azure SQL Database~~relational data stores

may include but is not limited to:

- ~~• Provision Azure SQL Database~~
- Configure elastic pools
- ~~• Configure data backup strategies~~
- ~~• Configure elastic jobs~~
- ~~• Provision Azure SQL database managed instance~~
- ~~• Configure connections~~
- ~~• Manage data synchronization~~
- 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 SQL Data Warehouse.
- Implement PolyBase.

~~Implement hybrid data scenarios~~

may include but is not limited to:

- ~~• Design hybrid solution~~
- ~~• Design data replication and synchronization~~

Manage Azure DevOps Pipelines

may include but is not limited to:

- ~~Use a build service~~
- Deploy using Azure Resource Manager templates

Manage data security

may include but is not limited to:

- Implement data masking
- Encrypt data at rest and in motion

Manage and Develop Data Processing (3025-3530%)

Implement big data environments

may include but is not limited to:

- ~~Implement Hadoop clusters~~
- ~~Implement Databricks environment~~

Develop batch processing solutions

may include but is not limited to:

- ~~Develop batch processing solutions using Spark~~
- 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

may include but is not limited to:

- Configure input and output
- Select the appropriate windowing functions
- Implement event processing using Azure Stream Analytics
- ~~Query data using Azure Stream Analytics~~
- ~~Configure Azure Stream Analytics to read from Event Hub~~
- Configure Azure Stream Analytics to read from BLOB storage

Develop integration solutions

may include but is not limited to:

- ~~Create data pipeline systems in Azure~~
- ~~Develop data integration pipelines with Azure Data Factory~~
- ~~Develop data integration pipelines with Azure Databricks~~

Implement data migration

may include but is not limited to:

- Transform data
- Bulk load data with PolyBase

Automate Data Factory Pipelines

may include but is not limited to:

- Deploy data factory pipelines
- Configure Data Factory

Manage Data Security (15-20%)

Manage source data access security

may include but is not limited to:

- Connect to sources
- Create connection objects
- Install Gateways

Configure authentication and authorization

may include but is not limited to:

- Set up firewall rules
- Integrate Azure AD
- Design Cosmos DB security
- Design Data Lake security
- Design Azure SQL DB security
- Manage Access Control
- Manage permissions on resources

Manage and enforce data policies and standards

may include but is not limited to:

- Mask data
- Encrypt data at rest
- Encrypt data in motion
- Encrypt data elements
- Configure audit

Set up notifications

may include but is not limited to:

- Set up alerts on security threats
- Set up alerts on unexpected resource usage

Monitor and Optimize Data Solutions (1030-1535%)

Monitor data storage

may include but is not limited to:

- Monitor relational and non-relational data sources
- Implement BLOB storage monitoring
- Implement Data Lake Store monitoring
- Implement SQL Database monitoring
- Implement SQL Data Warehouse monitoring
- Implement Cosmos DB monitoring

- Configure Azure Monitor alerts
- Implement auditing by using Azure Log Analytics
- ~~Implement HDInsight monitoring~~

~~Monitor databases for a specified scenario~~

may include but is not limited to:

- ~~Implement SQL Database monitoring~~
- ~~Implement SQL Data Warehouse monitoring~~
- ~~Implement Cosmos DB monitoring~~

Monitor data processing

may include but is not limited to:

- Design and implement Data Factory monitoring
- Monitor Azure Databricks
- Monitor HDInsight processing
- Monitor stream analytics

~~Manage and Troubleshoot Azure Data Solutions (10-15%)~~

~~Manage Optimization~~ Optimize Azure data solutions

may include but is not limited to:

- Troubleshoot data partitioning bottlenecks
- ~~Optimize HIVE processing~~
- Optimize Data Lake Storage
- ~~Optimize SPARK processing~~
- Optimize ~~Azure~~ Stream Analytics
- Optimize SQL Data Warehouse
- Optimize SQL ~~DB~~ Database
- Manage data life cycle

~~Manage business continuity~~

may include but is not limited to:

- ~~Implement a disaster recovery strategy~~
- ~~Design for High Availability~~
- ~~Import and export data~~
- ~~Design a data retention policy~~