# AZURE Data Foundations Certification

## Skills measured as of November 2, 2023

### Audience profile

This exam is an opportunity to demonstrate your knowledge of core data concepts and related Microsoft Azure data services. As a candidate for this exam, you should have familiarity with Exam DP-900’s self-paced or instructor-led learning material.

This exam is intended for you, if you’re a candidate beginning to work with data in the cloud.

You should be familiar with:

* The concepts of relational and non-relational data.
* Different types of data workloads such as transactional or analytical.

You can use Azure Data Fundamentals to prepare for other Azure role-based certifications like Azure Database Administrator Associate or Azure Data Engineer Associate, but it is not a prerequisite for any of them.

### Skills at a glance

* Describe core data concepts (25–30%)
* Identify considerations for relational data on Azure (20–25%)
* Describe considerations for working with non-relational data on Azure (15–20%)
* Describe an analytics workload on Azure (25–30%)

### Describe core data concepts (25–30%)

#### Describe ways to represent data

1. Describe features of structured data
   * Adheres to a fixed schema, so all of the data has the same fields or properties.
   * Most commonly, the schema for structured data is tabular with rows and columns.
   * Often stored in a database in which multiple tables can reference one another by using key values in a relational model.
2. Describe features of semi-structured
   * Has some structure, but allows for some variation between entity instances. i.e. some customers may have an email, multiple emails, or none at all.
   * One common format of semi-structured data is JSON
3. Describe features of unstructured data
   * Some examples of unstructured data are:
     + Documents
     + Images
     + Audio
     + Video
     + Binary Files

#### Identify options for data storage

* Describe common formats for data files
  + Delimited Text Files
    - Plain text format with specific field delimiters and row terminators
      * Comma-separated values (csv)
      * Tab-separated values (tsv)
      * Field width data
    - Delimited text is a good choice for structured data that needs to be accessed by a wide range of applications and services in a human-readable format.
  + JavaScript Object Notation (JSON)
    - JSON is a ubiquitous format in which a hierarchical document schema is used to define data entities (objects) that have multiple attributes.
    - JSON is a flexible format that’s good for both structured and semi-structured data.
  + Extensible Markup Language (XML)
    - Has largely been superseded by JSON, but there are still some systems that use XML to represent data.
    - Much like HTML, XML uses tags enclosed in angle brackets to represent data, defining elements and attributes.
  + Binary Large Object (BLOB)
    - Unstructured data in raw binary usually storing images, audio, video, and application specific documents.
  + Optimized file formats
    - Avro
    - ORC (Optimized Row Columnar format)
    - Parquet
* Describe types of databases

#### Describe common data workloads

* Describe features of transactional workloads
* Describe features of analytical workloads

#### Identify roles and responsibilities for data workloads

* Describe responsibilities for database administrators
* Describe responsibilities for data engineers
* Describe responsibilities for data analysts

### Identify considerations for relational data on Azure (20–25%)

#### Describe relational concepts

* Identify features of relational data
* Describe normalization and why it is used
* Identify common structured query language (SQL) statements
* Identify common database objects

#### Describe relational Azure data services

* Describe the Azure SQL family of products including Azure SQL Database, Azure SQL Managed Instance, and SQL Server on Azure Virtual Machines
* Identify Azure database services for open-source database systems

### Describe considerations for working with non-relational data on Azure (15–20%)

#### Describe capabilities of Azure storage

* Describe Azure Blob storage
* Describe Azure File storage
* Describe Azure Table storage

#### Describe capabilities and features of Azure Cosmos DB

* Identify use cases for Azure Cosmos DB
* Describe Azure Cosmos DB APIs

### Describe an analytics workload on Azure (25–30%)

#### Describe common elements of large-scale analytics

* Describe considerations for data ingestion and processing
* Describe options for analytical data stores
* Describe Azure services for data warehousing, including Azure Synapse Analytics, Azure Databricks, Azure HDInsight, and Azure Data Factory

#### Describe consideration for real-time data analytics

* Describe the difference between batch and streaming data
* Describe technologies for real-time analytics including Azure Stream Analytics, Azure Synapse Data Explorer, and Spark Structured Streaming

#### Describe data visualization in Microsoft Power BI

* Identify capabilities of Power BI
* Describe features of data models in Power BI
* Identify appropriate visualizations for data

## Study resources

We recommend that you train and get hands-on experience before you take the exam. We offer self-study options and classroom training as well as links to documentation, community sites, and videos.

| **Study resources** | **Links to learning and documentation** |
| --- | --- |
| Get trained | [Choose from self-paced learning paths and modules or take an instructor-led course](https://learn.microsoft.com/en-us/certifications/exams/dp-900#two-ways-to-prepare) |
| Find documentation | [Azure SQL documentation - Azure SQL](https://learn.microsoft.com/en-us/azure/azure-sql/?view=azuresql) [SQL Server technical documentation - SQL Server](https://learn.microsoft.com/en-us/sql/sql-server/?view=sql-server-ver16) [Azure Blob Storage documentation](https://learn.microsoft.com/en-us/azure/storage/blobs/) [Azure Table storage documentation](https://learn.microsoft.com/en-us/azure/storage/tables/) [Azure Storage documentation](https://learn.microsoft.com/en-us/azure/storage/) [Azure Cosmos DB](https://learn.microsoft.com/en-us/azure/cosmos-db/) [Azure Synapse Analytics](https://learn.microsoft.com/en-us/azure/synapse-analytics/) [Azure Databricks](https://learn.microsoft.com/en-us/azure/databricks/) [Data Factory](https://learn.microsoft.com/en-us/azure/data-factory/) [Power BI documentation - Power BI](https://learn.microsoft.com/en-us/power-bi/) |
| Ask a question | [Microsoft Q&A | Microsoft Docs](https://learn.microsoft.com/en-us/answers/products/) |
| Get community support | [Analytics on Azure | TechCommunity](https://techcommunity.microsoft.com/t5/analytics-on-azure/bd-p/AnalyticsonAzureDiscussion) [Azure Synapse Analytics | TechCommunity](https://techcommunity.microsoft.com/t5/azure-synapse-analytics/bd-p/AzureSynapseAnalytics) [Welcome to the SQL Server Community (microsoft.com)](https://techcommunity.microsoft.com/t5/sql-server/ct-p/SQL-Server) [Azure PaaS - Microsoft Tech Community](https://techcommunity.microsoft.com/t5/azure-paas/bd-p/AzurePaaS) |
| Follow Microsoft Learn | [Microsoft Learn - Microsoft Tech Community](https://techcommunity.microsoft.com/t5/microsoft-learn/ct-p/MicrosoftLearn) |
| Find a video | [Exam Readiness Zone](https://learn.microsoft.com/en-us/shows/exam-readiness-zone/) [Data Exposed](https://learn.microsoft.com/en-us/shows/data-exposed/) [Browse other Microsoft Learn shows](https://learn.microsoft.com/en-us/shows/browse) |

## Change log

Key to understanding the table: The topic groups (also known as functional groups) are in bold typeface followed by the objectives within each group. The table is a comparison between the two versions of the exam skills measured and the third column describes the extent of the changes.

| **Skill area prior to November 2, 2023** | **Skill area as of November 2, 2023** | **Change** |
| --- | --- | --- |
| Audience profile |  | No change |
| **Describe core data concepts** | **Describe core data concepts** | No change |
| Describe ways to represent data | Describe ways to represent data | No change |
| Identify options for data storage | Identify options for data storage | No change |
| Describe common data workloads | Describe common data workloads | No change |
| Identify roles and responsibilities for data workloads | Identify roles and responsibilities for data workloads | No change |
| **Identify considerations for relational data on Azure** | **Identify considerations for relational data on Azure** | No change |
| Describe relational concepts | Describe relational concepts | No change |
| Describe relational Azure data services | Describe relational Azure data services | Minor |
| **Describe considerations for working with non-relational data on Azure** | **Describe considerations for working with non-relational data on Azure** | No change |
| Describe capabilities of Azure storage | Describe capabilities of Azure storage | No change |
| Describe capabilities and features of Azure Cosmos DB | Describe capabilities and features of Azure Cosmos DB | No change |
| **Describe an analytics workload on Azure** | **Describe an analytics workload on Azure** | No change |
| Describe common elements of large-scale analytics | Describe common elements of large-scale analytics | No change |
| Describe consideration for real-time data analytics | Describe consideration for real-time data analytics | No change |
| Describe data visualization in Microsoft Power BI | Describe data visualization in Microsoft Power BI | No change |

## Skills measured prior to November 2, 2023

### Audience profile

This exam is an opportunity to demonstrate knowledge of core data concepts and related Microsoft Azure data services. Candidates for this exam should have familiarity with DP-900’s self-paced or instructor-led learning material.

This exam is intended for candidates beginning to work with data in the cloud.

Candidates should be familiar with the concepts of relational and non-relational data, and different types of data workloads such as transactional or analytical.

Azure Data Fundamentals can be used to prepare for other Azure role-based certifications like Azure Database Administrator Associate or Azure Data Engineer Associate, but it is not a prerequisite for any of them.

### Skills at a glance

* Describe core data concepts (25–30%)
* Identify considerations for relational data on Azure (20–25%)
* Describe considerations for working with non-relational data on Azure (15–20%)
* Describe an analytics workload on Azure (25–30%)

### Describe core data concepts (25–30%)

#### Describe ways to represent data

* Describe features of structured data
* Describe features of semi-structured
* Describe features of unstructured data

#### Identify options for data storage

* Describe common formats for data files
* Describe types of databases

#### Describe common data workloads

* Describe features of transactional workloads
* Describe features of analytical workloads

#### Identify roles and responsibilities for data workloads

* Describe responsibilities for database administrators
* Describe responsibilities for data engineers
* Describe responsibilities for data analysts

### Identify considerations for relational data on Azure (20–25%)

#### Describe relational concepts

* Identify features of relational data
* Describe normalization and why it is used
* Identify common structured query language (SQL) statements
* Identify common database objects

#### Describe relational Azure data services

* Describe the Azure SQL family of products including Azure SQL Database, Azure SQL
* Managed Instance, and SQL Server on Azure Virtual Machines
* Identify Azure database services for open-source database systems

### Describe considerations for working with non-relational data on Azure (15–20%)

#### Describe capabilities of Azure storage

* Describe Azure Blob storage
* Describe Azure File storage
* Describe Azure Table storage

#### Describe capabilities and features of Azure Cosmos DB

* Identify use cases for Azure Cosmos DB
* Describe Azure Cosmos DB APIs

### Describe an analytics workload on Azure (25–30%)

#### Describe common elements of large-scale analytics

* Describe considerations for data ingestion and processing
* Describe options for analytical data stores
* Describe Azure services for data warehousing, including Azure Synapse Analytics, Azure Databricks, Azure HDInsight, and Azure Data Factory

#### Describe consideration for real-time data analytics

* Describe the difference between batch and streaming data
* Describe technologies for real-time analytics including Azure Stream Analytics, Azure Synapse Data Explorer, and Spark Structured Streaming

#### Describe data visualization in Microsoft Power BI

* Identify capabilities of Power BI
* Describe features of data models in Power BI
* Identify appropriate visualizations for data