

Study guide for Exam DP-203: Data Engineering on Microsoft Azure

QUICK NAVIGATION

Document purpose

About Exam DP-203: Data Engineering on Microsoft Azure

Skills measured

Certification journey

Exam overview

Objective domains

Additional study resources

Document purpose

As an attendee of the Exam Prep session for Exam **DP-203: Data Engineering on Microsoft Azure**, you can use this guide as a summary of the topics covered and to explore important links and additional resources. The information and materials found here can help you focus your studies as you prepare for the exam.

About Exam DP-203: Data Engineering on Microsoft Azure

[Exam DP-203](#) is required to earn the [Microsoft Certified: Azure Data Engineer Associate certification](#).

This exam measures your ability to accomplish the following technical tasks: design and implement data storage; design and develop data processing; design and implement data security; monitor and optimize data storage and data processing.

1 - Study guide for Exam DP-203

© Microsoft Corporation. All rights reserved.

Candidates for this certification have subject matter expertise integrating, transforming, and consolidating data from various structured and unstructured data systems into structures that are suitable for building analytics solutions. Candidates must have solid knowledge of data processing languages, such as SQL, Python, or Scala, and they need to understand parallel processing and data architecture patterns.

Skills measured

For the full list of the skills that the exam measures, along with the level of experience and expertise that you'll need as an exam candidate, check out the [Exam DP-203 skills outline](#).

Certification journey

For an overview of the journey to Microsoft Certification, including prerequisites (if any) and follow-up resources, explore [The journey to Microsoft Certified: Azure Data Engineer Associate](#).

Exam overview

For information on the exam, including the types of questions you may encounter, read [About Microsoft Certification exams](#).

Objective domains

This section itemizes the topics covered in the Exam Prep session and links to Microsoft documentation so you can review the topics in detail.

- [Design and implement data storage \(40–45%\)](#)
- [Design and develop data processing \(25–30%\)](#)
- [Design and implement data security \(10–15%\)](#)
- [Monitor and optimize data storage and data processing \(10–15%\)](#)

Design and implement data storage (40–45%)

Design a data storage structure

- [Overview of Azure Data Lake Storage Gen2](#)
- [Access tiers for Azure Blob Storage](#)
- [Right storage account for file requirements](#)
- [Query data using Azure Synapse Analytics](#)
- [Dynamic file pruning \(DFP\)](#)

Design a partition strategy

- [Create file partitions using Azure Synapse Studio](#)
- [Partitioning tables](#)
- [Distribution in tabular models](#)
- [Optimizing data warehouse query performance in Azure Synapse Analytics](#)
- [Partitioning tables in Dedicated SQL Pool](#)

Design the serving layer

- [Star schema](#)
- [Multidimensional schemas and data](#)
- [Temporal tables in Azure SQL Database](#)
- [Getting started with temporal tables in Azure SQL Database and Azure SQL Managed Instance](#)
- [Self-hosted integration runtime](#)
- [Manage the self-hosted integration runtime](#)
- [Analytical data store in Azure](#)
- [Usage of Apache Spark pools](#)
- [Shared metadata table in Azure Synapse Analytics](#)

Implement physical data storage structures

- [Data compression](#)
- [Partitioning tables in Azure Synapse](#)
- [Table distribution in Azure Synapse](#)
- [Change how a storage account is replicated](#)
- [Implement lifecycle management](#)

Implement logical data structures

- [Design Slowly Changing Dimensions](#)
- [Slowly Changing Dimension transformation](#)
- [Create external tables in Azure Synapse serverless SQL pools](#)
- [Create views in Azure Synapse server SQL pools](#)

Implement the serving layer

- [Implement a star schema in Azure SQL and Azure Synapse](#)
- [Load Parquet file data in an Azure Spark Notebook in Azure Synapse](#)
- [Create a serverless Apache Spark pool in Azure Synapse Analytics using web tools](#)
- [Flatten nested structures and explode arrays with Apache Spark in Azure Synapse](#)
- [Preserve metadata and ACLs using copy activity in Azure Data Factory](#)

Design and develop data processing (25–30%)

Ingest and transform data

- [Ingestion technologies](#)
- [Extract, transform, and load \(ETL\) data by using Azure Databricks](#)
- [Azure Stream Analytics workflow](#)
- [Methods to handle missing data](#)
- [Explore and fix data using Azure Synapse Analytics](#)
- [Query and transform JSON data with Apache Spark for Azure Synapse](#)
- [Column-family databases](#)
- [Transform data using Scala](#)
- [Prepare and transform data](#)

Design and develop a batch processing solution

- [Overview of modern data warehouse architecture](#)
- [Choosing a batch processing technology in Azure](#)
- [Manage source data files](#)
- [Create pipeline with copy activity in Azure Data Factory](#)
- [Azure Synapse Studio notebooks](#)
- [Dedupe rows and find nulls](#)
- [MERGE \(Transact-SQL\)](#)
- [Continuous integration and delivery for Azure Synapse workspace](#)
- [Handle SQL truncation error rows](#)
- [Backup and restore in Azure Synapse Dedicated SQL Pool](#)
- [Best practices for loading data in Azure Synapse Analytics](#)
- [Debugging Apache Spark jobs](#)

Design and develop a stream processing solution

- [Modern data warehouse architecture](#)
- [Stream processing: Azure Databricks and Azure Event Hubs](#)
- [Apache Spark structured streaming](#)
- [Monitor for performance efficiency](#)
- [How to work with windowing functions](#)
- [Schema drift](#)
- [Time handling considerations](#)
- [Checkpoints in Azure Stream Analytics jobs](#)
- [Late arriving data](#)
- [Watermarks in Azure Stream Analytics jobs](#)
- [Tune a Stream Analytics query to increase throughput for Stream Analytics jobs](#)
- [Repartitioning and parallelization](#)
- [Output error policy in Azure Stream Analytics](#)
- [Stream Analytics upsert](#)
- [Time progression](#)
- [Stream processing with Azure Stream Analytics](#)

Manage batches and pipelines

- [Data loading best practices](#)
- [Design and create tests for analytics in Azure Synapse](#)
- [Trigger batches using Azure Synapse](#)
- [Monitor your Azure Synapse workspace](#)

Design and implement data security (10–15%)

Design security for data policies and standards

- [Encryption at rest and in transit](#)
- [Data ingestion security considerations](#)
- [Configure authentication using Azure Key Vault](#)
- [Configure data lake security with access control lists \(ACLs\)](#)
- [Securing access to data](#)
- [Column-level security in SQL Data Warehouse](#)

Implement data security

- [Manage authorization: Column and row level](#)
- [Manage user permissions in Azure Synapse](#)
- [Retention policy on storage account](#)
- [Auditing for Azure SQL Database and Azure Synapse Analytics](#)
- [Network security options for Azure Synapse Analytics](#)
- [Generate resource token in Azure Databricks](#)
- [Dynamic data masking](#)
- [Data masking in Azure Synapse](#)
- [Security controls: Data protection](#)
- [Secure a Dedicated SQL Pool](#)

Monitor and optimize data storage and data processing (10–15%)

Monitor data storage and data processing

- [Monitor and alert Data Factory by using Azure Monitor](#)
- [Improving performance with workload management and materialized views](#)
- [Creating statistics to improve performance](#)
- [Cluster performance in Dedicated SQL Pools](#)
- [Collect custom logs with Log Analytics agent](#)
- [Monitor workspace pipeline runs using Azure Synapse Studio](#)
- [Deploying Apache Airflow in Azure to build and run data pipelines](#)

Optimize and troubleshoot data storage and data processing

- [Auto Optimize in Azure Databricks](#)
- [Modify user-defined functions](#)
- [Choosing a distributed column](#)
- [Handle data spillage](#)
- [Troubleshoot performance bottlenecks in Azure Databricks](#)
- [Azure Synapse Dedicated SQL Pool shuffle speed](#)
- [Create an Azure Synapse workspace using an Azure Resource Manager template \(ARM template\)](#)
- [Indexing Dedicated SQL Pool tables in Azure Synapse Analytics](#)
- [Performance tuning with result set caching](#)
- [Optimize Apache Spark jobs in Azure Synapse Analytics](#)
- [Troubleshoot library installation errors](#)
- [Debug pipelines using Azure Synapse Pipelines](#)

Additional study resources

In addition to the documentation listed in the previous sections, we offer several resources to help you prepare for the exam and to stay up to speed and engaged with the Azure community. These resources range from formal training to blogs and even interviews with Microsoft team members.

[Course DP-203T00: Data Engineering on Microsoft Azure](#)

Take a four-day instructor-led course that covers data engineering patterns and practices as they pertain to working with batch and real-time analytical solutions using Azure data platform technologies. The course combines lectures with practical, hands-on exercises.

[DP-203 learning paths](#)

Don't miss these free, self-paced online resources to help you gain the skills needed to earn your certification.

[Azure documentation](#)

Stay informed on the latest products, tools, and features, and get information on pricing, partners, support, solutions, and more.

[Azure Community Support](#)

Ask questions, get answers, and connect with Microsoft engineers and Azure community experts.

[Microsoft Learn Community Blog](#)

Get the latest information about certification tests and exam study groups.

[Channel 9](#)

Explore this community site for customers. It includes video channels, discussions, podcasts, screencasts, and interviews.

[Azure Tuesdays with Corey](#)

Corey Sanders answers your questions about Microsoft Azure.

[Azure Fridays](#)

Scott Hanselman, Partner Program Manager, speaks with Azure engineers as they demo capabilities and share insights.

[Microsoft Azure Blog](#)

Keep current on what's happening in Azure, including what's in preview and what's generally available, along with Azure news, updates, and much more.