MICROSOFT DP-203

Data Engineering on Microsoft Azure Certification Questions & Answers

Get Instant Access to Vital Exam Acing Materials | Study Guide | Sample Questions | Practice Test

DP-203

Microsoft Certified - Azure Data Engineer Associate

40-60 Questions Exam – 700 / 1000 Cut Score – Duration of 150 minutes




EDUSUM

#1 Online Certification Guide

Table of Contents:

Discover More about the DP-203 Certification 2

DP-203 Data Engineering on Microsoft Azure Certification Details: 2

DP-203 Syllabus:.....3

Design and Implement Data Storage (40-45%) 3
Design and Develop Data Processing (25-30%) 4
Design and Implement Data Security (10-15%) 6
Monitor and Optimize Data Storage and Data Processing (10-15%) 7

Broaden Your Knowledge with Microsoft DP-203 Sample Questions: 8

Avail the Study Guide to Pass DP-203 Data Engineering on Microsoft Azure Exam: 12

Career Benefits: 12

Discover More about the DP-203 Certification

Are you interested in passing the Microsoft DP-203 exam? First discover, who benefits from the DP-203 certification. The DP-203 is suitable for a candidate if he wants to learn about Microsoft Azure. Passing the DP-203 exam earns you the Microsoft Certified - Azure Data Engineer Associate title.

While preparing for the DP-203 exam, many candidates struggle to get the necessary materials. But do not worry; your struggling days are over. The DP-203 PDF contains some of the most valuable preparation tips and the details and instant access to useful **DP-203 study materials just at one click.**

DP-203 Data Engineering on Microsoft Azure Certification Details:

Exam Name	Microsoft Certified - Azure Data Engineer Associate
Exam Code	DP-203
Exam Price	$165 (USD)
Duration	150 mins
Number of Questions	40-60
Passing Score	700 / 1000
Books / Training	**DP-203T00: Data Engineering on Microsoft Azure**
Schedule Exam	**Pearson VUE**
Sample Questions	**Data Engineering on Microsoft Azure Sample Questions**
Practice Exam	**Microsoft DP-203 Certification Practice Exam**

DP-203 Syllabus:

Topic	Details
Design and Implement Data Storage (40-45%)	
Design a data storage structure	- design an Azure Data Lake solution - recommend file types for storage - recommend file types for analytical queries - design for efficient querying - design for data pruning - design a folder structure that represents the levels of data transformation - design a distribution strategy - design a data archiving solution
Design a partition strategy	- design a partition strategy for files - design a partition strategy for analytical workloads - design a partition strategy for efficiency/performance - design a partition strategy for Azure Synapse Analytics - identify when partitioning is needed in Azure Data Lake Storage Gen2
Design the serving layer	- design star schemas - design slowly changing dimensions - design a dimensional hierarchy - design a solution for temporal data - design for incremental loading - design analytical stores - design metastores in Azure Synapse Analytics and Azure Databricks
Implement physical data storage structures	- implement compression - implement partitioning - implement sharding - implement different table geometries with Azure Synapse Analytics pools - implement data redundancy

Topic	Details
	- implement distributions - implement data archiving
Implement logical data structures	- build a temporal data solution - build a slowly changing dimension - build a logical folder structure - build external tables - implement file and folder structures for efficient querying and data pruning
Implement the serving layer	- deliver data in a relational star schema - deliver data in Parquet files - maintain metadata - implement a dimensional hierarchy
Design and Develop Data Processing (25-30%)	
Ingest and transform data	- transform data by using Apache Spark - transform data by using Transact-SQL - transform data by using Data Factory - transform data by using Azure Synapse Pipelines - transform data by using Stream Analytics - cleanse data - split data - shred JSON - encode and decode data - configure error handling for the transformation - normalize and denormalize values - transform data by using Scala - perform data exploratory analysis
Design and develop a batch processing solution	- develop batch processing solutions by using Data Factory, Data Lake, Spark, Azure Synapse Pipelines, PolyBase, and Azure Databricks - create data pipelines - design and implement incremental data loads - design and develop slowly changing dimensions - handle security and compliance requirements - scale resources

Topic	Details
	- configure the batch size - design and create tests for data pipelines - integrate Jupyter/Python notebooks into a data pipeline - handle duplicate data - handle missing data - handle late-arriving data - upsert data - regress to a previous state - design and configure exception handling - configure batch retention - design a batch processing solution - debug Spark jobs by using the Spark UI
Design and develop a stream processing solution	- develop a stream processing solution by using Stream Analytics, Azure Databricks, and Azure Event Hubs - process data by using Spark structured streaming - monitor for performance and functional regressions - design and create windowed aggregates - handle schema drift - process time series data - process across partitions - process within one partition - configure checkpoints/watermarking during processing - scale resources - design and create tests for data pipelines - optimize pipelines for analytical or transactional purposes - handle interruptions - design and configure exception handling - upsert data - replay archived stream data - design a stream processing solution

Topic	Details
Manage batches and pipelines	- trigger batches - handle failed batch loads - validate batch loads - manage data pipelines in Data Factory/Synapse Pipelines - schedule data pipelines in Data Factory/Synapse Pipelines - implement version control for pipeline artifacts - manage Spark jobs in a pipeline
Design and Implement Data Security (10-15%)	
Design security for data policies and standards	- design data encryption for data at rest and in transit - design a data auditing strategy - design a data masking strategy - design for data privacy - design a data retention policy - design to purge data based on business requirements - design Azure role-based access control (Azure RBAC) and POSIX-like Access Control List (ACL) for Data Lake Storage Gen2 - design row-level and column-level security
Implement data security	- implement data masking - encrypt data at rest and in motion - implement row-level and column-level security - implement Azure RBAC - implement POSIX-like ACLs for Data Lake Storage Gen2 - implement a data retention policy - implement a data auditing strategy - manage identities, keys, and secrets across different data platform technologies - implement secure endpoints (private and public) - implement resource tokens in Azure Databricks - load a DataFrame with sensitive information

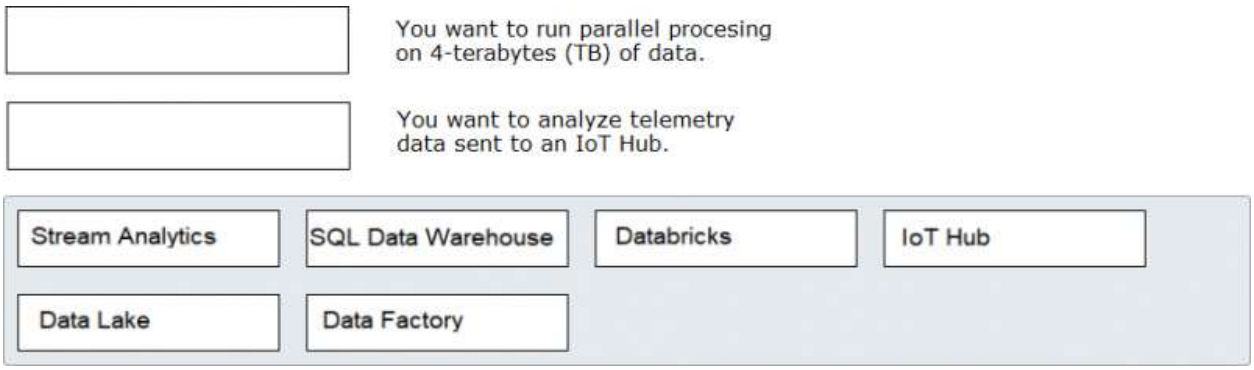
Topic	Details
	- write encrypted data to tables or Parquet files - manage sensitive information
Monitor and Optimize Data Storage and Data Processing (10-15%)	
Monitor data storage and data processing	- implement logging used by Azure Monitor - configure monitoring services - measure performance of data movement - monitor and update statistics about data across a system - monitor data pipeline performance - measure query performance - monitor cluster performance - understand custom logging options - schedule and monitor pipeline tests - interpret Azure Monitor metrics and logs - interpret a Spark directed acyclic graph (DAG)
Optimize and troubleshoot data storage and data processing	- compact small files - rewrite user-defined functions (UDFs) - handle skew in data - handle data spill - tune shuffle partitions - find shuffling in a pipeline - optimize resource management - tune queries by using indexers - tune queries by using cache - optimize pipelines for analytical or transactional purposes - optimize pipeline for descriptive versus analytical workloads - troubleshoot a failed spark job - troubleshoot a failed pipeline run

Broaden Your Knowledge with Microsoft DP-203 Sample Questions:

Question: 1

You are designing a big data streaming solution. You need to choose the most appropriate resource for different scenarios. Which Azure resource should you choose?

To answer, drag the appropriate resource to each scenario. A resource may be used once, more than once, or not at all.



a) SQL Data Warehouse
Stream Analytics
b) SQL Data Warehouse
Databricks
c) Stream Analytics
Data Lake
d) IoT Hub
Data Factory

Answer: a

Question: 2

Which offering provides scale-out parallel processing and dramatically accelerates performance of analytics clusters when integrated with the IBM Flash System?

a) IBM Cloud Object Storage
b) IBM Spectrum Accelerate
c) IBM Spectrum Scale
d) IBM Spectrum Connect

Answer: c

Question: 3

You are a data engineer for an Azure SQL Database. You write the following SQL statements:

CREATE TABLE Customer (

CustomerID int IDENTITY PRIMARY KEY,

GivenName varchar(100) MASKED WITH (FUNCTION = 'partial(2,"XX",0)') NULL,

SurName varchar(100) NOT NULL,

Phone varchar(12) MASKED WITH (FUNCTION = 'default()')

INSERT Customer (GivenName, SurName, Phone) VALUES ('Sammy', 'Jack', '555.111.2222');

SELECT * FROM Customer;

You need to determine what is returned by the SELECT query. What data is returned?

a) 1 SaXX Jack XXX.XXX.2222
b) 1 XXXX Jack XXX.XXX.XXXX
c) 1 xx Jack XXX.XXX.2222
d) 1 SaXX Jack xxxx

Answer: d

Question: 4

A company has an Azure SQL data warehouse. They want to use PolyBase to retrieve data from an Azure Blob storage account and ingest into the Azure SQL data warehouse. The files are stored in parquet format. The data needs to be loaded into a table called lead2pass_sales.

Which of the following actions need to be performed to implement this requirement?

(Choose 4)

a) Create an external file format that would map to the parquet-based files
b) Load the data into a staging table
c) Create an external table called lead2pass_sales_details
d) Create an external data source for the Azure Blob storage account
e) Create a master key on the database
f) Configure Polybase to use the Azure Blob storage account

Answer: b, c, d, e

Question: 5

A company purchases IoT devices to monitor manufacturing machinery. The company uses an Azure IoT Hub to communicate with the IoT devices. The company must be able to monitor the devices in real-time.
You need to design the solution. What should you recommend?

a) Azure Data Factory instance using Azure Portal
b) Azure Analysis Services using Microsoft Visual Studio
c) Azure Stream Analytics Edge application using Microsoft Visual Studio
d) Azure Data Factory instance using Microsoft Visual Studio

Answer: c

Question: 6

A company manages several on-premises Microsoft SQL Server databases. You need to migrate the databases to Microsoft Azure by using a backup process of Microsoft SQL Server.
Which data technology should you use?

a) Azure SQL Database single database
b) Azure SQL Data Warehouse
c) Azure Cosmos DB
d) Azure SQL Database Managed Instance
e) HDInsight Spark cluster

Answer: d

Question: 7

A company is planning on creating an Azure SQL database to support a mission critical application. The application needs to be highly available and not have any performance degradation during maintenance windows.
Which of the following technologies can be used to implement this solution?
(Choose 3)

a) Premium Service Tier
b) Virtual Machine Scale Sets
c) Basic Service Tier
d) SQL Data Sync
e) Always On Availability Groups
f) Zone-redundant configuration

Answer: a, e, f

Question: 8

A company has a SaaS solution that uses Azure SQL Database with elastic pools. The solution contains a dedicated database for each customer organization. Customer organizations have peak usage at different periods during the year. You need to implement the Azure SQL Database elastic pool to minimize cost. Which option or options should you configure?

a) Number of transactions only
b) eDTUs per database only
c) Number of databases only
d) CPU usage only
e) eDTUs and max data size

Answer: e

Question: 9

The data engineering team manages Azure HDInsight clusters. The team spends a large amount of time creating and destroying clusters daily because most of the data pipeline process runs in minutes.
You need to implement a solution that deploys multiple HDInsight clusters with minimal effort. What should you implement?

a) Azure Databricks
b) Azure Traffic Manager
c) Azure Resource Manager templates
d) Ambari web user interface

Answer: c

Question: 10

An in-house team is developing a new application. The design document specifies that data should be represented using nodes and relationships in graph structures. Individual data elements are relatively small.
You need to recommend an appropriate data storage solution. Which solution should you recommend?

a) Azure Storage Blobs
b) Cosmos DB
c) Azure Data Lake Store
d) HBase in HDInsight

Answer: b

Avail the Study Guide to Pass DP-203 Data Engineering on Microsoft Azure Exam:

- Find out about the DP-203 syllabus topics. Visiting the official site offers an idea about the exam structure and other important study resources. Going through the syllabus topics help to plan the exam in an organized manner.
- Once you are done exploring the **DP-203 syllabus**, it is time to plan for studying and covering the syllabus topics from the core. Chalk out the best plan for yourself to cover each part of the syllabus in a hassle-free manner.
- A study schedule helps you to stay calm throughout your exam preparation. It should contain your materials and thoughts like study hours, number of topics for daily studying mentioned on it. The best bet to clear the exam is to follow your schedule rigorously.
- The candidate should not miss out on the scope to learn from the DP-203 training. Joining the Microsoft provided training for DP-203 exam helps a candidate to strengthen his practical knowledge base from the certification.
- Learning about the probable questions and gaining knowledge regarding the exam structure helps a lot. Go through the **DP-203 sample questions** and boost your knowledge
- Make yourself a pro through online practicing the syllabus topics. DP-203 practice tests would guide you on your strengths and weaknesses regarding the syllabus topics. Through rigorous practicing, you can improve the weaker sections too. Learn well about time management during exam and become confident gradually with practice tests.

Career Benefits:

- Passing the DP-203 exam, helps a candidate to prosper highly in his career. Having the certification on the resume adds to the candidate's benefit and helps to get the best opportunities.

Here Is the Trusted Practice Test for the DP-203 Certification

EduSum.Com is here with all the necessary details regarding the DP-203 exam. We provide authentic practice tests for the DP-203 exam. What do you gain from these practice tests? You get to experience the real exam-like questions made by industry experts and get a scope to improve your performance in the actual exam. Rely on EduSum.Com for rigorous, unlimited two-month attempts on the **DP-203 practice tests**, and gradually build your confidence. Rigorous practice made many aspirants successful and made their journey easy towards grabbing the Microsoft Certified - Azure Data Engineer Associate.

Start Online Practice of DP-203 Exam by visiting URL
https://www.edusum.com/microsoft/dp-203-data-engineering-microsoft-azure