K21Academy
Learn Cloud from Experts

Silver Microsoft Partner

Microsoft Azure Data Engineer Certification [DP-203]

25 SAMPLE EXAM QUESTIONS

[EDITION 01]

Question 1: Which data platform technology is a globally distributed, multimodel database that can perform queries in less than a second?

- Azure SQL Database
- Azure Cosmos DB
- Azure SQL Data Warehouse
- Azure SQL Managed Instance

Correct. Azure Cosmos DB is a globally distributed, multimodel database that can offer subsecond query performance.

Question 2: Which data store is the least expensive choice when you want to store data but don't need to query it?

- Azure Stream Analytics
- Azure Cosmos DB
- Azure Databricks
- Azure Storage

Correct. Azure Storage offers a massively scalable object store for data objects and file system services for the cloud. If you create a Blob storage account, you can't directly query the data.

Question 3: Which Azure service is the best choice to store documentation about a data source?

- Azure Data Factory
- Azure Blob Storage
- Azure Data Catalog
- Azure Data Lake Storage

Correct. Azure Data Catalog is a central place where an organization's users can contribute their knowledge. Together, they build a community of data sources that the organization owns.

Question 4: Which role works with Azure Cognitive Services, Cognitive Search, and the Bot Framework?

- A data engineer
- A data Analyst
- A data scientist
- An AI engineer

Correct. Artificial intelligence (AI) engineers work with AI services such as Cognitive Services, Cognitive Search, and the Bot Framework.

Question 5: Which Azure data platform is commonly used to process data in an ELT framework?

- Azure HDInsight
- Azure Data Factory
- Azure Databricks
- Azure Data Lake Storage

Correct. Azure Data Factory is a cloud-integration service that orchestrates the movement of data between various data stores.

Question 6: Which type of transactional database system would work best for product data?

- OLAP
- OLTP

Correct. OLTP systems support a large set of users, have quick response times, handle large volumes of data, are highly available, and are great for small or relatively simple transactions.

Question 7: . A JSON file is an example of which type of data?

- Structured
- Semi-structured
- Unstructured

Correct. Semi-structured data contains tags that make the organization and hierarchy of the data apparent.

Question 8: A video is an example of which type of data?

- Structured
- Relational
- Semi-structured
- Unstructured

Correct. Unstructured data is often delivered in files. A video may have an overall structure but the data that comprises the video itself is unstructured.

Question 9: The name of a storage account must be:

- Unique within the containing resource group.
- Unique within your Azure subscription.
- Globally unique.
-

Correct. The storage account name is used as part of the URI for API access, so it must be globally unique.

-
-

Question 10: In a typical project, when would you create your storage account(s)?

- At the beginning, during project setup.
- After deployment, when the project is running.
- Before creating a resource group.
- At the end, during resource cleanup.

Correct. Storage accounts are stable for the lifetime of a project. It's common to create them at the start of a project.

Question 11: Which of the following describes a good strategy for creating storage accounts and blob containers for your application?

- Create both your Azure Storage accounts and containers before deploying your application.
- Create Azure Storage accounts in your application as needed. Create the containers before deploying the application.
- Create Azure Storage accounts before deploying your app. Create containers in your application as needed.

Correct. Creating an Azure Storage account is an administrative activity and can be done prior to deploying an application. Container creation is lightweight and is often driven by run-time data which makes it a good activity to do in your application.

Question 12: Which of the following can be used to initialize the Blob Storage client library within an application?

- An Azure username and password.

  An Azure SQL user name and password.

  The Azure Storage account connection string.

  A globally-unique identifier (GUID) that represents the application.

  The Azure Storage account datacenter and location identifiers.

Correct. The correct answer is **The Azure Storage account connection string**.

The Azure Storage account connection string is a string that contains the information needed to connect

to an Azure Storage account. It includes the storage account name, account key, and endpoint suffix. The connection string can be found in the Azure Portal under the "Access Keys" section of the storage account.

Question 13: Which technology is typically used as a staging area in a modern data warehousing architecture?

- Azure Data Lake.
- Azure Synapse SQL Pools.
- Azure Cosmos DB

  Azure Synapse Spark Pools

Correct. Correct. Azure Data Lake Store Gen 2 is the technology that will be used to stage data before loading it into the various components of Azure Synapse Analytics.

Question 14: Which component enables you to perform code free transformations in Azure Synapse Analytics?

- Studio.
- Copy activity.
- Linked services
- Mapping data flow.

Correct. You can natively perform data transformations with Azure Data Factory code free using the Mapping Data Flow task.

Question 15: Which transformation in the Mapping Data Flow is used to routes data rows to different streams based on matching conditions?

- Lookup.
- Window
- Conditional Split.
- Select.

Correct. A Conditional Split transformation routes data rows to different streams based on matching conditions. The conditional split transformation is similar to a CASE decision structure in a programming language.

Question 16: Which transformation is used to load data into a data store or compute resource?

Window.

Aggregate

Source.

Sink.

Correct. A Sink transformation allows you to choose a dataset definition for the destination output data.

You can have as many sink transformations as your data flow requires.

Question 17: What distribution option would you use for a product dimension table that will contain 1,000 records in Synapse Analytics?

- DISTRIBUTION = ROUND_ROBIN.
- DISTRIBUTION = HASH([ProductId]).
- DISTRIBUTION = REPLICATE.

Correct. Replicate will result in a copy of the table on each compute node, which performs well with joins to the distributed fact table.

Question 18: What is the difference between a star schema and a snowflake schema?

- A star schema uses surrogate keys while a snowflake schema uses business keys.
- All dimensions in a star schema join directly to the fact table (denormalized) while some dimension tables in a snowflake schema are normalized.
- All dimensions in a star schema are normalized while all dimensions in a snowflake schema join directly to the fact table (denormalized).

Correct. A star schema is highly denormalized so that the fact table joins directly to dimension; a snowflake schema normalizes some dimensions into multiple tables such as DimProduct, DimProductSubcategory, and DimProductCategory.

Question 19: How does splitting source files help maintain good performance when loading into Synapse Analytics?

- optimized processing of smaller file sizes.
- Compute node to storage segment alignment
- Reduced possibility of data corruptions.

Correct. The correct answer is **optimized processing of smaller file sizes**.

When source files are split into smaller files, it allows Synapse Analytics to process them more efficiently. This is because Synapse Analytics can parallelize the load process across multiple compute nodes, and each compute node can only process one file at a time. If the source files are too large, then the load process will be bottlenecked by the compute nodes, and performance will suffer.

Question 20: Which Workload Management capability manages minimum and maximum resource allocations during peak periods?

Workload Isolation.

Workload Importance.

Workload Containment.

Correct. Workload Isolation assigns maximum and minimum usage values for varying resources under load. These adjustments can be done live without having to take the SQL Pool offline.

Question 21: Which T-SQL Statement loads data directly from Azure Storage?

- LOAD DATA.
- INSERT DATA.
- COPY
- INSERT FROM FILE.

Correct. The T-SQL COPY Statement reads data from Azure Blob Storage or the Azure Data Lake and inserts it into a table within the SQL Pool.

Question 22: Which Azure Data Factory component orchestrates a transformation job or runs a data movement command?

- Linked Services
- Pipeline
- Datasets
- Activities

Correct. Activities contains the transformation logic or the analysis commands of the Azure Data Factory's work.

Question 23: You are moving data from an Azure Data Lake Gen2 store to Azure Synapse Analytics. Which Azure Data Factory integration runtime would be used in a data copy activity?

- Azure-SSIS
- Azure
- Self-hosted

Correct. The correct answer is **Azure**.

The Azure integration runtime is the default integration runtime used by Azure Data Factory. It is a fully managed, serverless integration runtime that can be used to connect to a variety of data stores and compute services. The Azure integration runtime is the most appropriate choice for moving data from an Azure Data Lake Gen2 store to Azure Synapse Analytics

---

Question 24: In Azure Data Factory authoring tool, where would you find the Copy data activity?

- Move & Transform
- Batch Service

- Databricks
- HDInsight

Correct. The Move & Transform section contains activities that are specific to Azure Data Factory copying data and defining data flows.

-

Question 25: You want to ingest data from a SQL Server database hosted on an on-premises Windows Server. What integration runtime is required for Azure Data Factory to ingest data from the on-premises server?

Azure-SSIS Integration Runtime

Self-Hosted Integration Runtime

Azure Integration Runtime

Correct. A self-hosted integration runtime can run copy activities between a cloud data store and a data store in a private network. It also can dispatch transform activities against compute resources in an on-premises network or an Azure virtual network.

Question 26: By default, how long are the Azure Data Factory diagnostic logs retained for?

- 15 days
- 90 days
- 30 days
- 45 days

Correct. The Azure Data Factory diagnostic logs are retained for 45 days.

FREE CLASS

Register for our FREE Class To know about What is Azure Data Engineer role, most important Azure services you will master Like - ADLS Gen 2, Azure Databricks, Azure Synapse, Azure Data Factory, Azure Stream Analytics, Event hubs, IoT hubs, Azure Purview, etc., and what to study Including Hands-On labs you must perform to clear Azure Data Engineer Certification [DP-203] so that you can stay ahead in your career and earn a lot more a https://k21academy.com/dp20302

ABOUT AUTHOR

Atul Kumar Is An Oracle ACE, Author & Oracle Certified Cloud Architect With 20+ Years Experience.

He is helping his customer to become an expert in Microsoft Azure Data Engineer Certification [DP-203]

/oracleappsdba

/k21academy

/k21academy

/k21academy

/k21academy

Learn Cloud from Experts

contact@k21academy.com