# DP-203 Q&As

Data Engineering on Microsoft Azure

## Pass Microsoft DP-203 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

**https://www.leads4pass.com/dp-203.html**

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

**QUESTION 1**

You are designing an enterprise data warehouse in Azure Synapse Analytics that will contain a table named Customers. Customers will contain credit card information.

You need to recommend a solution to provide salespeople with the ability to view all the entries in Customers. The solution must prevent all the salespeople from viewing or inferring the credit card information.

What should you include in the recommendation?

A. data masking

B. Always Encrypted

C. column-level security

D. row-level security

Correct Answer: C

Column-level security simplifies the design and coding of security in your application, allowing you to restrict column access to protect sensitive data.

Reference: https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/column-level-security

**QUESTION 2**

You have a table in an Azure Synapse Analytics dedicated SQL pool. The table was created by using the following Transact-SQL statement.

```
CREATE TABLE [dbo].[DimEmployee](
   [EmployeeKey] [int] IDENTITY(1,1) NOT NULL,
   [EmployeeID] [int] NOT NULL,
   [FirstName] [varchar](100) NOT NULL,
   [LastName] [varchar](100) NOT NULL,
   [JobTitle] [varchar](100) NULL,
   [LastHireDate] [date] NULL,
   [StreetAddress] [varchar](500) NOT NULL,
   [City] [varchar](200) NOT NULL,
   [StateProvince] [varchar](50) NOT NULL,
   [Portalcode] [varchar](10) NOT NULL
   )
```

You need to alter the table to meet the following requirements:

Ensure that users can identify the current manager of employees.

Support creating an employee reporting hierarchy for your entire company.

Provide fast lookup of the managers\\' attributes such as name and job title.

Which column should you add to the table?

A. [ManagerEmployeeID] [int] NULL

B. [ManagerEmployeeID] [smallint] NULL

C. [ManagerEmployeeKey] [int] NULL

D. [ManagerName] [varchar](200) NULL

Correct Answer: C

We need an extra column to identify the Manager. Use the data type as the EmployeeKey column, an int column.

Reference: https://docs.microsoft.com/en-us/analysis-services/tabular-models/hierarchies-ssas-tabular

**QUESTION 3**

You are designing a star schema for a dataset that contains records of online orders. Each record includes an order date, an order due date, and an order ship date. You need to ensure that the design provides the fastest query times of the records when querying for arbitrary date ranges and aggregating by fiscal calendar attributes. Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Create a date dimension table that has a DateTime key.

B. Use built-in SQL functions to extract date attributes.

C. Create a date dimension table that has an integer key in the format of YYYYMMDD.

D. In the fact table, use integer columns for the date fields.

E. Use DateTime columns for the date fields.

Correct Answer: BD

**QUESTION 4**

You are designing an Azure Databricks table. The table will ingest an average of 20 million streaming events per day.

You need to persist the events in the table for use in incremental load pipeline jobs in Azure Databricks. The solution must minimize storage costs and incremental load times.

What should you include in the solution?

A. Partition by DateTime fields.

B. Sink to Azure Queue storage.

C. Include a watermark column.

D. Use a JSON format for physical data storage.

Correct Answer: B

A. Partition by DataTime field

Each partition will generate a file. Loading latency may reduce, but feel storage cost will increase because generate more folders and files for different partition. Is it right???

B. Sink to Azure Queue Storage.

Read this document. Spark table files are stored in DBFS. Mount Azure Blob storage containers to Databricks File System (DBFS). If The Databricks ABS-AQS provides these two benefits, sounds like it is a correct answer.

https://docs.microsoft.com/en-us/azure/databricks/spark/latest/structured-streaming/aqs

C. Include a watermark column: For sure it is not correct

Watermarks define how long your aggregate should wait around for data delay.

D. User a Json format for physical data storage. - ???

Don\\'t find any documents to compare physical data storage of JSON, CSV, and Parquet.

The Databricks ABS-AQS connector uses Azure Queue Storage (AQS) to provide an optimized file source that lets you find new files written to an Azure Blob storage (ABS) container without repeatedly listing all of the files. This provides two major advantages:

1.

Lower latency: no need to list nested directory structures on ABS, which is slow and resource intensive.

2.

Lower costs: no more costly LIST API requests made to ABS.

Reference: https://docs.microsoft.com/en-us/azure/databricks/spark/latest/structured-streaming/aqs

---

**QUESTION 5**

You have two Azure Blob Storage accounts named account1 and account2.

You plan to create an Azure Data Factory pipeline that will use scheduled intervals to replicate newly created or modified blobs from account1 to account2.

You need to recommend a solution to implement the pipeline. The solution must meet the following requirements:

1.

Ensure that the pipeline only copies blobs that were created or modified since the most recent replication event.

2.

Minimize the effort to create the pipeline. What should you recommend?

A. Run the Copy Data tool and select Metadata-driven copy task.

B. Create a pipeline that contains a Data Flow activity.

C. Create a pipeline that contains a flowlet.

D. Run the Copy Data tool and select Built-in copy task.

Correct Answer: A

Build large-scale data copy pipelines with metadata-driven approach in copy data tool

When you want to copy huge amounts of objects (for example, thousands of tables) or load data from large variety of sources, the appropriate approach is to input the name list of the objects with required copy behaviors in a control table,

and then use parameterized pipelines to read the same from the control table and apply them to the jobs accordingly. By doing so, you can maintain (for example, add/remove) the objects list to be copied easily by just updating the object

names in control table instead of redeploying the pipelines. What\\'s more, you will have single place to easily check which objects copied by which pipelines/triggers with defined copy behaviors.

Copy data tool in ADF eases the journey of building such metadata driven data copy pipelines. After you go through an intuitive flow from a wizard-based experience, the tool can generate parameterized pipelines and SQL scripts for you to

create external control tables accordingly. After you run the generated scripts to create the control table in your SQL database, your pipelines will read the metadata from the control table and apply them on the copy jobs automatically.

Incorrect:

Not C: A flowlet is a reusable container of activities that can be created from an existing mapping data flow or started from scratch. By reusing patterns you can prevent logic duplication and apply the same logic across many mapping data

flows.

With flowlets you can create logic to do things such as address cleaning or string trimming. You can then map the input and outputs to columns in the calling data flow for a dynamic code reuse experience.

Reference:

https://learn.microsoft.com/en-us/azure/data-factory/copy-data-tool-metadata-driven