QUESTIONS & *ANSWERS*

verified

DBCLOUD EXAMS

EXPERIENICIA Y TECNOLOGÍA

DBCLOUD CERTIFIED

Technology Provider: Microsoft
Exam: Data Engineering on Microsoft Azure
Number of questions in the database: 385
Exam Version: OCT.2026

Topic 1 - Question Set 1

Question #1*Topic 1*

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] [smallint] NULL
- B. [ManagerEmployeeKey] [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

Respuesta por mayor votación de la comunidad

https://dbcloud.academy/

C (100%)

Question #2*Topic 1*

You have an Azure Synapse workspace named MyWorkspace that contains an Apache Spark database named mytestdb.
You run the following command in an Azure Synapse Analytics Spark pool in MyWorkspace.
CREATE TABLE mytestdb.myParquetTable(
EmployeeID int,
EmployeeName string,
EmployeeStartDate date)

USING Parquet -
You then use Spark to insert a row into mytestdb.myParquetTable. The row contains the following data.

EmployeeName	EmployeeID	EmployeeStartDate
Alice	24	2020-01-25

One minute later, you execute the following query from a serverless SQL pool in MyWorkspace.

SELECT EmployeeID -
FROM mytestdb.dbo.myParquetTable
WHERE EmployeeName = 'Alice';
What will be returned by the query?

- A. 24
- B. an error
- C. a null value

Correct Answer: *B*

Respuesta por mayor votación de la comunidad

B (59%)

A (41%)

Question #3*Topic 1*

DRAG DROP -
You have a table named SalesFact in an enterprise data warehouse in Azure Synapse Analytics. SalesFact contains sales data from the past 36 months and has the following characteristics:
☞ Is partitioned by month
☞ Contains one billion rows
☞ Has clustered columnstore index
At the beginning of each month, you need to remove data from SalesFact that is older than 36 months as quickly as possible.
Which three actions should you perform in sequence in a stored procedure? To answer, move the

https://dbcloud.academy/

appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:

Actions	Answer Area
Switch the partition containing the stale data from SalesFact to SalesFact_Work.	
Truncate the partition containing the stale data.	
Drop the SalesFact_Work table.	
Create an empty table named SalesFact_Work that has the same schema as SalesFact.	
Execute a DELETE statement where the value in the Date column is more than 36 months ago.	
Copy the data to a new table by using CREATE TABLE AS SELECT (CTAS).	

Correct Answer:

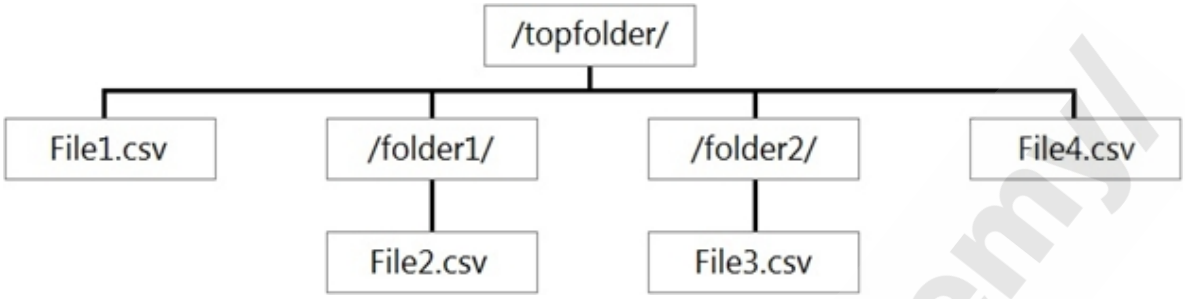
Actions	Answer Area
Switch the partition containing the stale data from SalesFact to SalesFact_Work.	Create an empty table named SalesFact_Work that has the same schema as SalesFact.
Truncate the partition containing the stale data.	Switch the partition containing the stale data from SalesFact to SalesFact_Work.
Drop the SalesFact_Work table.	Drop the SalesFact_Work table.
Create an empty table named SalesFact_Work that has the same schema as SalesFact.	
Execute a DELETE statement where the value in the Date column is more than 36 months ago.	
Copy the data to a new table by using CREATE TABLE AS SELECT (CTAS).	

Step 1: Create an empty table named SalesFact_work that has the same schema as SalesFact.
Step 2: Switch the partition containing the stale data from SalesFact to SalesFact_Work.
SQL Data Warehouse supports partition splitting, merging, and switching. To switch partitions between two tables, you must ensure that the partitions align on their respective boundaries and that the table definitions match.
Loading data into partitions with partition switching is a convenient way stage new data in a table that is not visible to users the switch in the new data.
Step 3: Drop the SalesFact_Work table.
Reference:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables-partition

Question #4*Topic 1*

https://dbcloud.academy/

You have files and folders in Azure Data Lake Storage Gen2 for an Azure Synapse workspace as shown in the following exhibit.



You create an external table named ExtTable that has LOCATION='/topfolder/'.
When you query ExtTable by using an Azure Synapse Analytics serverless SQL pool, which files are returned?

- A. File2.csv and File3.csv only
- B. File1.csv and File4.csv only
- C. File1.csv, File2.csv, File3.csv, and File4.csv
- D. File1.csv only

Correct Answer: *B*

Respuesta por mayor votación de la comunidad

B (73%)

C (25%)

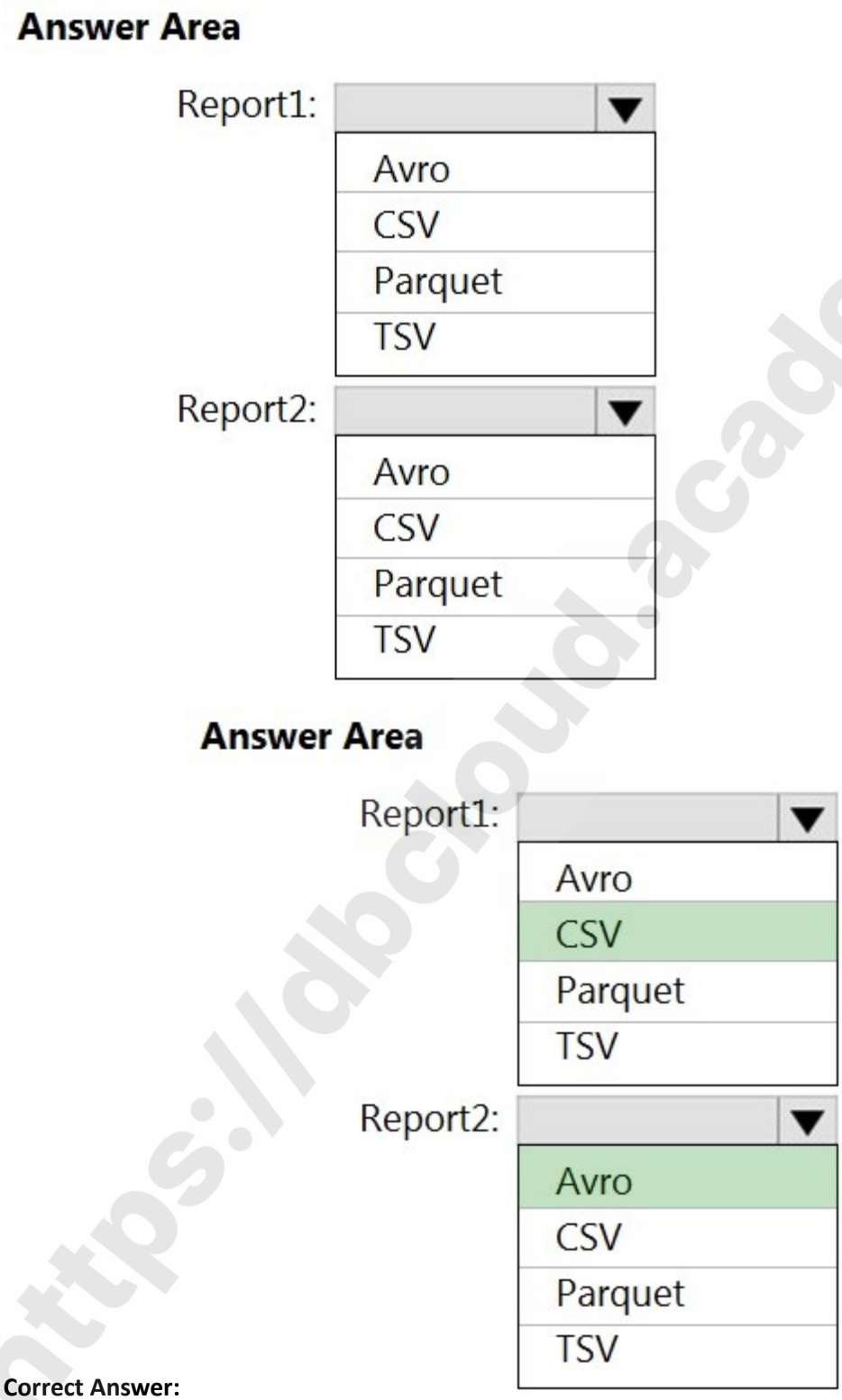
2%

Question #5*Topic 1*

HOTSPOT -
You are planning the deployment of Azure Data Lake Storage Gen2.
You have the following two reports that will access the data lake:
☞ Report1: Reads three columns from a file that contains 50 columns.
☞ Report2: Queries a single record based on a timestamp.
You need to recommend in which format to store the data in the data lake to support the reports.
The solution must minimize read times.
What should you recommend for each report? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

https://dbcloud.academy/

Hot Area:



Correct Answer:
Report1: CSV -
CSV: The destination writes records as delimited data.

https://dbcloud.academy/

Report2: AVRO -
AVRO supports timestamps.
Not Parquet, TSV: Not options for Azure Data Lake Storage Gen2.
Reference:
https://streamsets.com/documentation/datacollector/latest/help/datacollector/UserGuide/Destinations/ADLS-G2-D.html

Question #6*Topic 1*

You are designing the folder structure for an Azure Data Lake Storage Gen2 container.
Users will query data by using a variety of services including Azure Databricks and Azure Synapse Analytics serverless SQL pools. The data will be secured by subject area. Most queries will include data from the current year or current month.
Which folder structure should you recommend to support fast queries and simplified folder security?

- A. /{SubjectArea}/{DataSource}/{DD}/{MM}/{YYYY}/{FileData}_{YYYY}_{MM}_{DD}.csv
- B. /{DD}/{MM}/{YYYY}/{SubjectArea}/{DataSource}/{FileData}_{YYYY}_{MM}_{DD}.csv
- C. /{YYYY}/{MM}/{DD}/{SubjectArea}/{DataSource}/{FileData}_{YYYY}_{MM}_{DD}.csv
- D. /{SubjectArea}/{DataSource}/{YYYY}/{MM}/{DD}/{FileData}_{YYYY}_{MM}_{DD}.csv

Correct Answer: *D*
There's an important reason to put the date at the end of the directory structure. If you want to lock down certain regions or subject matters to users/groups, then you can easily do so with the POSIX permissions. Otherwise, if there was a need to restrict a certain security group to viewing just the UK data or certain planes, with the date structure in front a separate permission would be required for numerous directories under every hour directory. Additionally, having the date structure in front would exponentially increase the number of directories as time went on.
Note: In IoT workloads, there can be a great deal of data being landed in the data store that spans across numerous products, devices, organizations, and customers. It's important to pre-plan the directory layout for organization, security, and efficient processing of the data for down-stream consumers. A general template to consider might be the following layout:
{Region}/{SubjectMatter(s)}/{yyyy}/{mm}/{dd}/{hh}/

Respuesta por mayor votación de la comunidad

D (100%)

Question #7*Topic 1*

HOTSPOT -
You need to output files from Azure Data Factory.
Which file format should you use for each type of output? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

https://dbcloud.academy/