DUMPSQUEEN

Designing and Implementing Enterprise-Scale Analytics Solutions Using Microsoft Azure and Microsoft Power BI

Microsoft DP-500

Version Demo

Total Demo Questions: 10

Total Premium Questions: 113

Buy Premium PDF

https://dumpsqueen.com

support@dumpsqueen.com

dumpsqueen.com

Topic Break Down

Topic	No. of Questions
Topic 1, Litware, Inc.	6
Topic 2, Contoso, Ltd	8
Topic 3, Misc. Questions	99
Total	113

QUESTION NO: 1

Which two possible tools can you use to identify what causes the report to render slowly? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Synapse Studio

B. DAX Studio

C. Azure Data Studio

D. Performance analyzer in Power BI Desktop

ANSWER: B D

Explanation:

Some users indicate that the visuals in Power BI reports are slow to render when making filter selections.

B: You can investigate a slow query in a Power BI report using DAX Studio, looking at the query plan and the server timings.

D: Use Power BI Desktop Performance Analyzer to optimize the report or model.

Reference: https://www.sqlbi.com/tv/analyzing-a-slow-report-query-in-dax-studio/

https://docs.microsoft.com/en-us/power-bi/guidance/report-performance-troubleshoot

QUESTION NO: 2

You are optimizing a dataflow in a Power BI Premium capacity. The dataflow performs multiple joins. You need to reduce the load time of the dataflow.

Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Reduce the memory assigned to the dataflows.

B. Execute non-foldable operations before foldable operations.

C. Execute foldable operations before non-foldable operations.

D. Place the ingestion operations and transformation operations in a single dataflow.

E. Place the ingestion operations and transformation operations in separate dataflows.

ANSWER: C E

Explanation:

Using the compute engine to improve performance

Take the following steps to enable workloads trigger the compute engine, and always improve performance:

For computed and linked entities in the same workspace:

Ensure you perform the operations that fold, such as merges, joins, conversion, and others.

For ingestion focus on getting the data into the storage as fast as possible, using filters only if they reduce the overall dataset size. It's best practice to keep your transformation logic separate from this step, and allow the engine to focus on the initial gathering of ingredients. Next, separate your transformation and business logic into a separate dataflow in the same workspace, using linked or computed entities; doing so allows for the engine to activate and accelerate your computations. In our analogy, it's like food preparation in the kitchen: food preparation is typically a separate and distinct step from gathering your raw ingredients, and a pre-requisite for putting the food in the oven. Similarly, your logic needs to be prepared separately before it can take advantage of the compute engine.

Reference: https://docs.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-premium-workload-configuration

QUESTION NO: 3

You have a deployment pipeline for a Power BI workspace. The workspace contains two datasets that use import storage mode.

A database administrator reports a drastic increase in the number of queries sent from the Power BI service to an Azure SQL database since the creation of the deployment pipeline.

An investigation into the issue identifies the following:

You need to recommend a solution to reduce the size of the queries sent to the database when the dataset changes are published to development, test, or production.

What should you recommend?

A. From Capacity settings in the Power BI Admin portal, reduce the Max Intermediate Row Set Count setting.

B. Configure the dataset to use a composite model that has a DirectQuery connection to the fact table.

C. Enable the large dataset storage format for workspace.

D. From Capacity settings in the Power BI Admin portal, increase the Max Intermediate Row Set Count setting.

ANSWER: B

Explanation:

A composite model in Power BI means part of your model can be a DirectQuery connection to a data source (for example, SQL Server database), and another part as Import Data (for example, an Excel file). Previously, when you used DirectQuery, you couldn't even add another data source into the model.

DirectQuery and Import Data have different advantages.

Now the Composite Model combines the good things of both Import and DirectQuery into one model. Using the Composite Model, you can work with big data tables using DirectQuery, and still import smaller tables using Import Data.

Reference: https://radacad.com/composite-model-directquery-and-import-data-combined-evolution-begins-in-power-bi

QUESTION NO: 4

How should you configure the Power BI dataset refresh for the dbo.SalesTransactions table?

A. an incremental refresh of Product where the ModifiedDate value is during the last three days.

B. an incremental refresh of dbo.SalesTransactions where the SalesDate value is during the last three days.

C. a full refresh of all the tables

D. an incremental refresh of dbo.SalesTransactions where the SalesDate value is during the last hour.

ANSWER: B

Explanation:

The sales data in SQLDW is updated every 30 minutes. Records in dbo.SalesTransactions are updated in SQLDW up to three days after being created. The records do NOT change after three days.

Overview

Contoso, Ltd. is a company that sells enriched financial data to a variety of external customers.

Contoso has a main office in Los Angeles and two branch offices in New York and Seattle.

Data Infrastructure

Contoso has a 50-TB data warehouse that uses an instance of SQL Server on Azure Virtual Machines.

The data warehouse populates an Azure Synapse Analytics workspace that is accessed by the external customers. Currently, the customers can access alt the data.

Contoso has one Power BI workspace named FinData that contains a single dataset. The dataset contains financial data from around the world. The workspace is used by 10 internal users and one external customer. The dataset has the following two data sources: the data warehouse and the Synapse Analytics serverless SQL pool.

Users frequently query the Synapse Analytics workspace by using Transact-SQL.

User Problems

Contoso identifies the following user issues:

• Some users indicate that the visuals in Power BI reports are slow to render when making filter selections.

• Users indicate that queries against the serverless SQL pool fail occasionally because the size of tempdb has been exceeded.

• Users indicate that the data in Power BI reports is stale. You discover that the refresh process of the Power BI model occasionally times out

Planned Changes

Contoso plans to implement the following changes:

• Into the existing Power BI dataset, integrate an external data source that is accessible by using the REST API.

• Build a new dataset in the FinData workspace by using data from the Synapse Analytics dedicated SQL pool.

• Provide all the customers with their own Power BI workspace to create their own reports. Each workspace will use the new dataset in the FinData workspace.

• Implement subscription levels for the customers. Each subscription level will provide access to specific rows of financial data.

• Deploy prebuilt datasets to Power BI to simplify the query experience of the customers.

• Provide internal users with the ability to incorporate machine learning models loaded to the dedicated SQL pool.

QUESTION NO: 5 - (DRAG DROP)

You are configuring Azure Synapse Analytics pools to support the Azure Active Directory groups shown in the following table.

Name	Requirement
Group1	Analyze data to create and train machine learning models in Synapse Analytics.
Group2	Execute complex queries with multiple joins against relational data. Results will be exported by using PolyBase.
Group3	Query and load data from Apache Parquet files stored in Azure Data Lake Storage Gen2. Costs must be based on the amount of data processed.

Which type of pool should each group use? To answer, drag the appropriate pool types to the groups. Each pool type may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Pool Types	Answer Area	
Apache Spark pool	Group1:	Pool Type
Dedicated SQL pool	Group2:	Pool Type
Serverless SQL pool	Group3:	Pool Type

ANSWER:

Explanation:

Box 1: Apache Spark pool

An Apache Spark pool provides open-source big data compute capabilities. After you've created an Apache Spark pool in your Synapse workspace, data can be loaded, modeled, processed, and distributed for faster analytic insight.

Box 2: Dedicated SQL Pool

Dedicated SQL Pool - Data is stored in relational tables

Box 3: Serverless SQL pool

Serverless SQL pool - Cost is incurred for the data processed per query

Reference: https://docs.microsoft.com/en-us/azure/synapse-analytics/quickstart-create-apache-spark-pool-portal

https://www.royalcyber.com/blog/data-services/dedicated-sql-pool-vs-serverless-sql/

QUESTION NO: 6

You have new security and governance protocols for Power BI reports and datasets. The new protocols must meet the following requirements.

• New reports can be embedded only in locations that require authentication.

• Live connections are permitted only for workspaces that use Premium capacity datasets.

Which three actions should you recommend performing in the Power BI Admin portal? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. From Tenant settings, disable Allow XMLA endpoints and Analyze in Excel with on-premises datasets.

B. From the Premium per user settings, set XMLA Endpoint to Off.

C. From Embed Codes, delete all the codes.

D. From Capacity settings, set XMLA Endpoint to Read Write.

E. From Tenant settings, set Publish to web to Disable.

ANSWER: A D E

Explanation:

Reference: https://docs.microsoft.com/en-us/power-bi/enterprise/service-premium-connect-tools

https://powerbi.microsoft.com/en-us/blog/power-bi-february-service-update

QUESTION NO: 7

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are using an Azure Synapse Analytics serverless SQL pool to query a collection of Apache Parquet files by using automatic schema inference. The files contain more than 40 million rows of UTF-8-encoded business names, survey names, and participant counts. The database is configured to use the default collation.

The queries use open row set and infer the schema shown in the following table.

name	system_type_name	max_length
businessName	varchar(8000)	8000
surveyName	varchar(8000)	8000
participants	int	4

You need to recommend changes to the queries to reduce I/O reads and tempdb usage.

Solution: You recommend defining an external table for the Parquet files and updating the query to use the table

Does this meet the goal?

A. Yes

B. No

ANSWER: B

QUESTION NO: 8 - (HOTSPOT)

You manage a dataset that contains the two data sources as shown in the following table.

Data source	Type of data	Privacy level
Azure SQL database	Sensitive company data	Private
Microsoft SharePoint folder	Non-sensitive company data	Private

When you attempt to refresh the dataset in powerbi.com, you receive the following error message: "[Unable to combine data] Add Columns is accessing data sources that have privacy levels which cannot be used together. Please rebuild this data combination."

You discover that the dataset contains queries that fold data from the SharePoint folder to the Azure SQL database.

You need to resolve the error. The solution must provide the highest privacy possible.

Which privacy level should you select for each data source? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Azure SQL database:
| Organizational |
| Private |
| Public |

SharePoint folder:
| Organizational |
| Private |
| Public |

ANSWER:

Azure SQL database:
| Organizational |
| Private |
| Public |

SharePoint folder:
| Organizational |
| Private |
| Public |

Explanation:

Box 1: Private

This Formula.Firewall error is the result of Power Query's Data Privacy Firewall (aka the Firewall)

Note: Folding is a term that refers to converting expressions in M (such as filters, renames, joins, and so on) into operations against a raw data source (such as SQL, OData, and so on).

Box 2: Organizational

Organizational Limits the visibility of a data source to a trusted group of people. It is isolated from all Public data sources, but is visible to other Organizational data sources. A common example is a Microsoft Word document on an intranet SharePoint site with permissions enabled for a trusted group.

QUESTION NO: 9 - (DRAG DROP)

You have an Azure Synapse Analytics serverless SQL pool.

You need to return a list of files and the number of rows in each file.

How should you complete the Transact-SQL statement? To answer, drag the appropriate values to the targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Values

| APPROX_COUNT_DISTINCT |
| COUNT_BIG |
| OPENDATASOURCE |
| OPENJSON |
| OPENQUERY |
| OPENROWSET |

Answer Area

```
SELECT
    asa.filename() AS [filename]
    ,                          (*) AS [rows]
FROM

    BULK 'parquet/production/year=2017/month=9/*.parquet',
    DATA_SOURCE = 'DataLake1',
    FORMAT= 'PARQUET'
) asa

GROUP BY [filename]
```

ANSWER:

Values

| APPROX_COUNT_DISTINCT |
| COUNT_BIG |
| OPENDATASOURCE |
| OPENJSON |
| OPENQUERY |
| OPENROWSET |

Answer Area

```
SELECT
    asa.filename() AS [filename]
    , APPROX_COUNT_DISTINCT (*) AS [rows]
FROM
          OPENROWSET

    BULK 'parquet/production/year=2017/month=9/*.parquet',
    DATA_SOURCE = 'DataLake1',
    FORMAT= 'PARQUET'
) asa

GROUP BY [filename]
```

Explanation:

Box 1: APPROX_COUNT_DISTINCT

The APPROX_COUNT_DISTINCT function returns the approximate number of unique non-null values in a group.

Box 2: OPENROWSET

OPENROWSET function in Synapse SQL reads the content of the file(s) from a data source. The data source is an Azure storage account and it can be explicitly referenced in the OPENROWSET function or can be dynamically inferred from URL of the files that you want to read. The OPENROWSET function can optionally contain a DATA_SOURCE parameter to specify the data source that contains files.

The OPENROWSET function can be referenced in the FROM clause of a query as if it were a table name OPENROWSET. It supports bulk operations through a built-in BULK provider that enables data from a file to be read and returned as a rowset.

Reference: https://docs.microsoft.com/en-us/sql/t-sql/functions/approx-count-distinct-transact-sql

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-openrowset

QUESTION NO: 10 - (HOTSPOT)

You need to populate the CustomersWithProductScore table.

How should you complete the stored procedure? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
DECLARE @model                                    = (
                          ┌──────────────────┐
    SELECT model          │ BIT              │
                          │ FLOAT            │
    FROM MLModel          │ NVARCHAR(1000)   │
                          │ VARBINARY(max)   │
    WHERE model_name = 'PredictPurchase'      ┘

    );

INSERT INTO CustomersWithProductScore (

    CustomerID

    ,CustomerEmail

    ,ProductID

    ,ProductName

    ,Score

    )

SELECT d.CustomerID

    ,d.CustomerEmail

    ,d.ProductID

    ,d.ProductName
                                              ┌──────────────────────────────┐  AS d)
    ,p.score                                  │ dbo.Customer                 │
                                              │ dbo.CustomerPurchases        │
FROM PREDICT(MODEL = @model, DATA =           │ dbo.CustomersWithProductScore│
                                              │ dbo.Product                  │
    WITH (score FLOAT) AS p;                  └──────────────────────────────┘
```

ANSWER:

Answer Area

```
DECLARE @model                        = (

    SELECT model         │ BIT              │
                         │ FLOAT            │
    FROM MLModel         │ NVARCHAR(1000)   │
                         │ VARBINARY(max)   │
    WHERE model_name = 'PredictPurchase

    );

INSERT INTO CustomersWithProductScore (

    CustomerID

    ,CustomerEmail

    ,ProductID

    ,ProductName

    ,Score

    )

SELECT d.CustomerID

    ,d.CustomerEmail

    ,d.ProductID

    ,d.ProductName
                                              AS d)
    ,p.score                 │ dbo.Customer                      │
                             │ dbo.CustomerPurchases             │
FROM PREDICT(MODEL = @model, DATA = │ dbo.CustomersWithProductScore │
                             │ dbo.Product                       │
    WITH (score FLOAT) AS p;
```

Explanation:

```
DECLARE @model  [        ▼] = (
                 BIT
                 FLOAT
                 NVARCHAR(1000)
                 VARBINARY(max)
        SELECT model
        FROM MLModel
        WHERE model_name = 'PredictPurchase'
        );
INSERT INTO CustomersWithProductScore (
        CustomerID
      , CustomerEmail
      , ProductID
      , ProductName
      , Score
      )
SELECT d.CustomerID
     , d.CustomerEmail
     , d.ProductID
     , d.ProductName
     , p.score
FROM PREDICT(MODEL = @model, DATA = [          ▼] AS d)
                                     dbo.Customer
                                     dbo.CustomerPurchases
                                     dbo.CustomersWithProductScore
                                     dbo.Product

        WITH (score FLOAT) AS p;
```

Box 1: FLOAT

Identify which customers should receive promotional emails based on their likelihood of purchasing promoted products.

FLOT is used in the last statement of the code: WITH (score FLOAT) as p;

From syntax: MODEL

The MODEL parameter is used to specify the model used for scoring or prediction. The model is specified as a variable or a literal or a scalar expression.

Box 2: dbo.CustomerWithProductScore

Identify which customers should receive promotional emails based on their likelihood of purchasing promoted products.

Only table CustomerWithProductScore has the required filed score.

From the syntax:

DATA

The DATA parameter is used to specify the data used for scoring or prediction. Data is specified in the form of a table source in the query. Table source can be a table, table alias, CTE alias, view, or table-valued function.

Reference: https://docs.microsoft.com/en-us/sql/t-sql/queries/predict-transact-sql