DP-600: Implementing Analytics Solutions Using Microsoft Fabric (beta)

Case study -
This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.
To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.
At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -
To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -
Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

Existing Environment -

Identity Environment -
Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

Data Environment -
Contoso has the following data environment:
The Sales division uses a Microsoft Power BI Premium capacity.
The semantic model of the Online Sales department includes a fact table named Orders that uses Import made. In the system of origin, the OrderID value represents the sequence in which orders are created.
The Research department uses an on-premises, third-party data warehousing product.
Fabric is enabled for contoso.com.
An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.
A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

Requirements -

Planned Changes -
Contoso plans to make the following changes:
Enable support for Fabric in the Power BI Premium capacity used by the Sales division.
Make all the data for the Sales division and the Research division available in Fabric.
For the Research division, create two Fabric workspaces named Productline1ws and Productine2ws.
In Productline1ws, create a lakehouse named Lakehouse1.
In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

Data Analytics Requirements -
Contoso identifies the following data analytics requirements:
All the workspaces for the Sales division and the Research division must support all Fabric experiences.
The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.
The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name.
For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and

shortcuts by using SQL endpoints.
For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.
All the semantic models and reports for the Research division must use version control that supports branching.

Data Preparation Requirements -
Contoso identifies the following data preparation requirements:
The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.
All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

Semantic Model Requirements -
Contoso identifies the following requirements for implementing and managing semantic models:
The number of rows added to the Orders table during refreshes must be minimized.
The semantic models in the Research division workspaces must use Direct Lake mode.

General Requirements -
Contoso identifies the following high-level requirements that must be considered for all solutions:
Follow the principle of least privilege when applicable.
Minimize implementation and maintenance effort when possible.
You need to ensure that Contoso can use version control to meet the data analytics requirements and the general requirements.
What should you do?

A. Store at the semantic models and reports in Data Lake Gen2 storage.

B. Modify the settings of the Research workspaces to use a GitHub repository.

C. Modify the settings of the Research division workspaces to use an Azure Repos repository.

D. Store all the semantic models and reports in Microsoft OneDrive.

Correct Answer: *B*

Community vote distribution

C (88%) 13%

SamuComqi 4 days, 6 hours ago

Selected Answer: C

C. Modify the settings of the Research division workspaces to use an Azure Repos repository.

Only Git currently supported (https://learn.microsoft.com/en-us/fabric/cicd/git-integration/intro-to-git-integration).
upvoted 1 times

Momoanwar 4 days, 16 hours ago

Selected Answer: C

Azure Repos
upvoted 1 times

David_Webb 6 days, 8 hours ago

Selected Answer: C

C as only Git in Azure Repos is supported currently.
https://learn.microsoft.com/en-us/fabric/cicd/git-integration/intro-to-git-integration#considerations-and-limitations
upvoted 2 times

rmeng 1 week ago

Selected Answer: B

option B: Modify the settings of the Research workspaces to use a GitHub repository
upvoted 1 times

Nicofr 1 week, 2 days ago

Selected Answer: C

"Currently, only Git in Azure Repos is supported." https://learn.microsoft.com/en-us/fabric/cicd/git-integration/intro-to-git-integration#considerations-and-limitations
upvoted 2 times

theseon 1 week, 6 days ago

Selected Answer: C

Why not Azure Repos?

upvoted 1 times

PraAg 1 week, 6 days ago

Correct

upvoted 1 times

Question #2 *Topic 1*

HOTSPOT -

Case study -
This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.
To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.
At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -
To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -
Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

Existing Environment -

Identity Environment -
Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

Data Environment -
Contoso has the following data environment:
The Sales division uses a Microsoft Power BI Premium capacity.
The semantic model of the Online Sales department includes a fact table named Orders that uses Import made. In the system of origin, the OrderID value represents the sequence in which orders are created.
The Research department uses an on-premises, third-party data warehousing product.
Fabric is enabled for contoso.com.
An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.
A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

Requirements -

Planned Changes -
Contoso plans to make the following changes:
Enable support for Fabric in the Power BI Premium capacity used by the Sales division.
Make all the data for the Sales division and the Research division available in Fabric.
For the Research division, create two Fabric workspaces named Productline1ws and Productine2ws.
In Productline1ws, create a lakehouse named Lakehouse1.
In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

Data Analytics Requirements -
Contoso identifies the following data analytics requirements:
All the workspaces for the Sales division and the Research division must support all Fabric experiences.
The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.

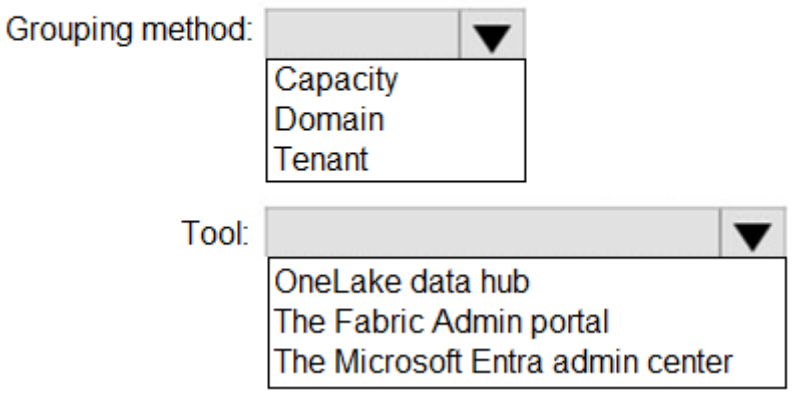
The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name.
For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and shortcuts by using SQL endpoints.
For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.
All the semantic models and reports for the Research division must use version control that supports branching.

Data Preparation Requirements -
Contoso identifies the following data preparation requirements:
The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.
All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

Semantic Model Requirements -
Contoso identifies the following requirements for implementing and managing semantic models:
The number of rows added to the Orders table during refreshes must be minimized.
The semantic models in the Research division workspaces must use Direct Lake mode.

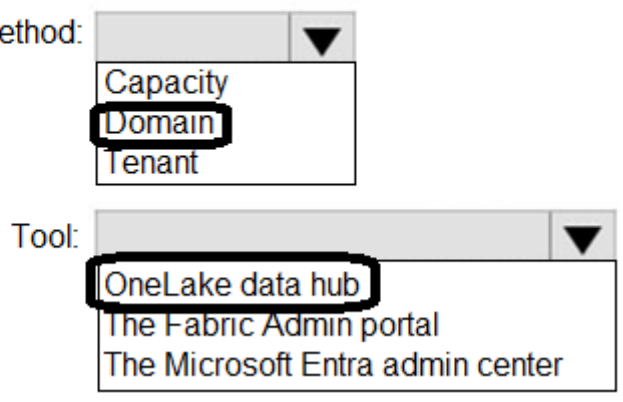
General Requirements -
Contoso identifies the following high-level requirements that must be considered for all solutions:
Follow the principle of least privilege when applicable.
Minimize implementation and maintenance effort when possible.
You need to recommend a solution to group the Research division workspaces.
What should you include in the recommendation? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area



Answer Area

Correct Answer:



SamuComqi 4 days, 6 hours ago
- Domain
- The Fabric Admin Portal

https://learn.microsoft.com/en-us/fabric/governance/domains#configure-domain-settings
upvoted 2 times

Momoanwar 4 days, 16 hours ago
Domain\Fabric Admin Portal
upvoted 2 times

David_Webb 6 days, 5 hours ago
Group method: Domain
Tool: The Fabric Admin Portal

The tool here should mean the tool to implement the grouping solution. Thus, the answer should be the Fabric admin portal instead.
https://learn.microsoft.com/en-us/fabric/governance/domains#create-a-domain

upvoted 2 times

Bharat 5 days, 18 hours ago

Thank you for the link. I stand corrected.

upvoted 1 times

praisekim 6 days, 12 hours ago

I think theseon is right
One Lake Data Hub is used to filter data based on domain, but It doesn't create the domain or group data
When you want to make a domain, you should use Admin portal

upvoted 1 times

Bharat 1 week ago

One Lake Data Hub is the right answer according to the documentation link that you have provided.

upvoted 1 times

rmeng 1 week ago

Domain: Group the Research division workspaces based on their departmental context.
One Lake Data Hub: Use One Lake Data Hub to filter and organize the Research division workspaces effectively.
Explanation:

Domain allows you to group workspaces based on their purpose or business context. In this case, grouping by department (Research division) aligns with the requirement.
One Lake Data Hub provides the necessary filtering capabilities for organizing workspaces based on department names.

upvoted 1 times

theseon 1 week, 6 days ago

Should the tool not be Fabric Admin Portal?

https://learn.microsoft.com/en-us/fabric/governance/domains#configure-domain-settings

upvoted 3 times

Case study -
This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.
To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.
At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -
To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -
Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

Existing Environment -

Identity Environment -
Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

Data Environment -
Contoso has the following data environment:
The Sales division uses a Microsoft Power BI Premium capacity.
The semantic model of the Online Sales department includes a fact table named Orders that uses Import made. In the system of origin, the OrderID value represents the sequence in which orders are created.
The Research department uses an on-premises, third-party data warehousing product.
Fabric is enabled for contoso.com.
An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.
A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

Requirements -

Planned Changes -
Contoso plans to make the following changes:
Enable support for Fabric in the Power BI Premium capacity used by the Sales division.
Make all the data for the Sales division and the Research division available in Fabric.
For the Research division, create two Fabric workspaces named Productline1ws and Productine2ws.
In Productline1ws, create a lakehouse named Lakehouse1.
In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

Data Analytics Requirements -
Contoso identifies the following data analytics requirements:
All the workspaces for the Sales division and the Research division must support all Fabric experiences.
The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.
The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name.
For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and

shortcuts by using SQL endpoints.
For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.
All the semantic models and reports for the Research division must use version control that supports branching.

Data Preparation Requirements -
Contoso identifies the following data preparation requirements:
The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.
All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

Semantic Model Requirements -
Contoso identifies the following requirements for implementing and managing semantic models:
The number of rows added to the Orders table during refreshes must be minimized.
The semantic models in the Research division workspaces must use Direct Lake mode.

General Requirements -
Contoso identifies the following high-level requirements that must be considered for all solutions:
Follow the principle of least privilege when applicable.
Minimize implementation and maintenance effort when possible.
You need to refresh the Orders table of the Online Sales department. The solution must meet the semantic model requirements.
What should you include in the solution?

A. an Azure Data Factory pipeline that executes a Stored procedure activity to retrieve the maximum value of the OrderID column in the destination lakehouse

B. an Azure Data Factory pipeline that executes a Stored procedure activity to retrieve the minimum value of the OrderID column in the destination lakehouse

C. an Azure Data Factory pipeline that executes a dataflow to retrieve the minimum value of the OrderID column in the destination lakehouse

D. an Azure Data Factory pipeline that executes a dataflow to retrieve the maximum value of the OrderID column in the destination lakehouse

Correct Answer: *D*

Community vote distribution

D (100%)

SamuComqi 4 days, 6 hours ago

D. an Azure Data Factory pipeline that executes a dataflow to retrieve the maximum value of the OrderID column in the destination lakehouse.
A dataflow can be used to retrieve the max OrderID number (stored in the destination table - OrderID is a sequencial number). This number can be used to set from which row data must be added to the destination table (implementing an incremental load).
upvoted 1 times

Momoanwar 4 days, 16 hours ago

Selected Answer: D
Max with dataflow
upvoted 2 times

theseon 1 week, 6 days ago

Selected Answer: D
we need to retrieve the maximum OrderID in the destination table to minimize the number of rows added during refresh. this would be an incremental load. can be done with data flows
upvoted 2 times

Case study -
This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.
To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.
At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -
To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -
Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

Existing Environment -

Identity Environment -
Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

Data Environment -
Contoso has the following data environment:
The Sales division uses a Microsoft Power BI Premium capacity.
The semantic model of the Online Sales department includes a fact table named Orders that uses Import made. In the system of origin, the OrderID value represents the sequence in which orders are created.
The Research department uses an on-premises, third-party data warehousing product.
Fabric is enabled for contoso.com.
An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.
A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

Requirements -

Planned Changes -
Contoso plans to make the following changes:
Enable support for Fabric in the Power BI Premium capacity used by the Sales division.
Make all the data for the Sales division and the Research division available in Fabric.
For the Research division, create two Fabric workspaces named Productline1ws and Productine2ws.
In Productline1ws, create a lakehouse named Lakehouse1.
In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

Data Analytics Requirements -
Contoso identifies the following data analytics requirements:
All the workspaces for the Sales division and the Research division must support all Fabric experiences.
The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.
The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name.
For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and

shortcuts by using SQL endpoints.
For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.
All the semantic models and reports for the Research division must use version control that supports branching.

Data Preparation Requirements -
Contoso identifies the following data preparation requirements:
The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.
All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

Semantic Model Requirements -
Contoso identifies the following requirements for implementing and managing semantic models:
The number of rows added to the Orders table during refreshes must be minimized.
The semantic models in the Research division workspaces must use Direct Lake mode.

General Requirements -
Contoso identifies the following high-level requirements that must be considered for all solutions:
Follow the principle of least privilege when applicable.
Minimize implementation and maintenance effort when possible.

Which syntax should you use in a notebook to access the Research division data for Productline1?

A. spark.read.format("delta").load("Tables/productline1/ResearchProduct")

B. spark.sql("SELECT * FROM Lakehouse1.ResearchProduct ")

C. external_table('Tables/ResearchProduct)

D. external_table(ResearchProduct)

Correct Answer: *A*

Community vote distribution

B (80%) A (20%)

David_Webb 1 day, 12 hours ago

Selected Answer: B

The correct answer is B.
The folder hierarchy of Tables in Lakehouse is incorrect for A.

upvoted 2 times

Jeff_Zhu 2 days, 10 hours ago

Selected Answer: B

The correct answer is B.
A should correct as :
df = spark.read.format("delta").load("Tables/ResearchProduct")

The Table folder could not handle two hierarchical

upvoted 2 times

SamuComqi 4 days, 6 hours ago

B. spark.sql("SELECT * FROM Lakehouse1.ResearchProduct ")

The syntax of C and D is correct for KQL databases (incorrect in this use-case). When the shortcut is created, no additional folders have been added to the Tables section, therefore answer A is incorrect. Once created, the line of answer B can be used to access data correctly.

https://learn.microsoft.com/en-us/fabric/onelake/onelake-shortcuts

upvoted 1 times

Momoanwar 4 days, 15 hours ago

Selected Answer: A

I think A. We get data from shirtcut

upvoted 1 times

HOTSPOT -

Case study -
This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.
To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.
At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -
To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -
Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

Existing Environment -

Fabric Environment -
Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

Available Data -
Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

Survey -

Question -

Response -
For each survey submitted, the following occurs:
One row is added to the Survey table.
One row is added to the Response table for each question in the survey.
The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score.
Customers can submit a survey after each purchase.

User Problems -
The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.
Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic

models, but the logic does NOT always match across implementations.

Requirements -

Planned Changes -
Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity
The following three workspaces will be created:
AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store
DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake
DataSciPOC: Will contain all the notebooks and reports created by the data scientists
The following will be created in the AnalyticsPOC workspace:
A data store (type to be decided)

A custom semantic model -

A default semantic model -

Interactive reports -
The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.
All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

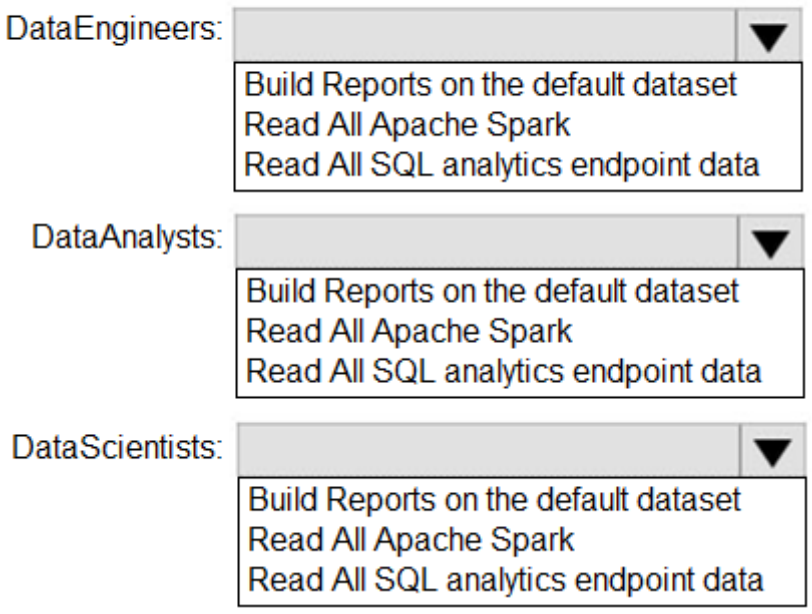
Technical Requirements -
The data store must support the following:
Read access by using T-SQL or Python
Semi-structured and unstructured data
Row-level security (RLS) for users executing T-SQL queries
Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.
Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model
The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model
The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.
The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SOL. queries and in the default semantic model. The following logic must be used:
List prices that are less than or equal to 50 are in the low pricing group.
List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.
List prices that are greater than 1,000 are in the high pricing group.

Security Requirements -
Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.
Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:
Fabric administrators will be the workspace administrators.
The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.
The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.
The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook
The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.
The date dimension must be available to all users of the data store.
The principle of least privilege must be followed.
Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:
FabricAdmins: Fabric administrators

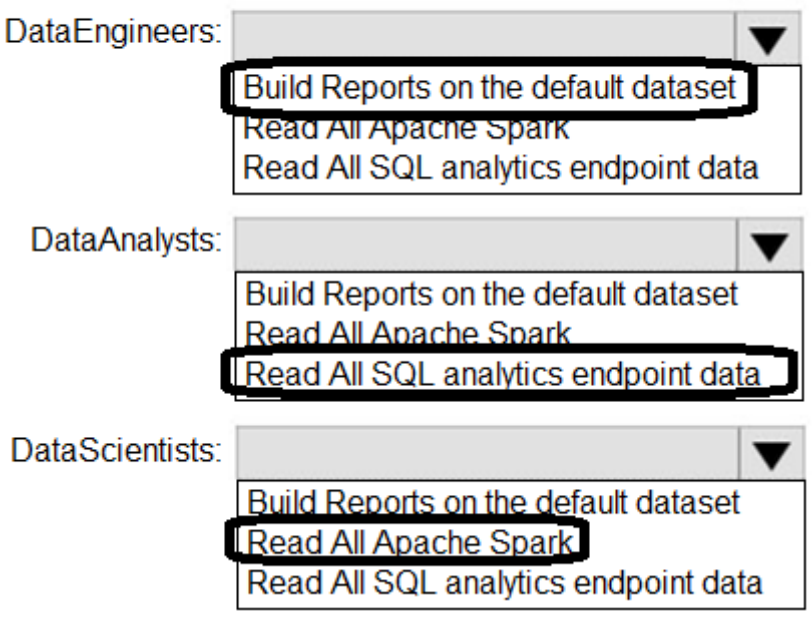
AnalyticsTeam: All the members of the analytics team
DataAnalysts: The data analysts on the analytics team
DataScientists: The data scientists on the analytics team
DataEngineers: The data engineers on the analytics team
AnalyticsEngineers: The analytics engineers on the analytics team

Report Requirements -
The data analysts must create a customer satisfaction report that meets the following requirements:
Enables a user to select a product to filter customer survey responses to only those who have purchased that product.
Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected dat.
Shows data as soon as the data is updated in the data store.
Ensures that the report and the semantic model only contain data from the current and previous year.
Ensures that the report respects any table-level security specified in the source data store.
Minimizes the execution time of report queries.
You need to assign permissions for the data store in the AnalyticsPOC workspace. The solution must meet the security requirements.
Which additional permissions should you assign when you share the data store? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area



Answer Area



Correct Answer:

David_Webb 1 day, 12 hours ago

DataEngineers: ReadAll Apache Spark
DataAnalyst: Build Reports on the default dataset
DataScientist: ReadAll Apache Spark

Data engineers will use Pyspark in notebooks to transform data from the Files folder in the Lakehouse.
Data analysts will build reports and dashboards from the prepared dataset.
Data scientists will use MLLib in notebooks to build models.
The "Read All SQL analytics endpoint data" should be for the AnalyticsEngineers. The analytics team has four types of members.

upvoted 1 times

wojciech_wie 2 days, 21 hours ago

Data Engineers = ReadAll Apache Spark
Data Analyst = Build Reports
Data Scientist = ReadAll Apache Spark

upvoted 1 times

vissu_settipally 3 days, 8 hours ago

Data Engineers = Read sql endpoints
Data Analyst = Build Reports
Data Scientist = They prefer Spark Always

upvoted 1 times

SamuComqi 4 days, 5 hours ago

Data Engineers: Read all SQL Analytics Endpoint data (use SQL to explore and create/modify tables, views, stored procedures).
Data Analysts: Build reports on the default dataset (using Power BI).
Data Scientists: Read all Apache Spark (they will use Notebooks to analyze data and apply ML models).

upvoted 1 times

Momoanwar 4 days, 15 hours ago

Engineers = spark
Analyst = report
Scientist = Endpoint

upvoted 1 times

Bharat 1 week ago

Here is my take on it:
Data Engineers: Read all Apache Spark - because they need to be able to work with Spark for Data curation.
Data Analysts: Build Reports on the default dataset - because they are report builders
Data Scientists: Read SQL Endpoints - They leverage curated data (by engineers) to do predictive analytics.
Let me know what you think.

upvoted 3 times

Nicofr 1 week, 1 day ago

For Data Engineers: Read SQL Endpoints
For Data Analysts: Build Reports on the default dataset
For Data Scientists: Read all Apache Spark

upvoted 2 times

Nicofr 1 week, 1 day ago

Maybe we can have multiple options for each as it is in Fabric :
For Data Engineers: Read SQL Endpoints and Read all Apache Spark
For Data Analysts: Build Reports on the default dataset and Read SQL Endpoints
For Data Scientists: Read all Apache Spark

upvoted 2 times

XiltroX 1 week, 2 days ago

For Data Engineers: Read all Apache Spark
For Data Analysts: Read SQL Endpoints
For Data Scientists: the last one remaining.

upvoted 1 times

HOTSPOT -

Case study -
This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.
To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.
At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -
To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -
Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

Existing Environment -

Fabric Environment -
Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

Available Data -
Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

Survey -

Question -

Response -
For each survey submitted, the following occurs:
One row is added to the Survey table.
One row is added to the Response table for each question in the survey.
The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score.
Customers can submit a survey after each purchase.

User Problems -
The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.
Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic

models, but the logic does NOT always match across implementations.

Requirements -

Planned Changes -
Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity
The following three workspaces will be created:
AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store
DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake
DataSciPOC: Will contain all the notebooks and reports created by the data scientists
The following will be created in the AnalyticsPOC workspace:
A data store (type to be decided)

A custom semantic model -

A default semantic model -

Interactive reports -
The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.
All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

Technical Requirements -
The data store must support the following:
Read access by using T-SQL or Python
Semi-structured and unstructured data
Row-level security (RLS) for users executing T-SQL queries
Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.
Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model
The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model
The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.
The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SOL. queries and in the default semantic model. The following logic must be used:
List prices that are less than or equal to 50 are in the low pricing group.
List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.
List prices that are greater than 1,000 are in the high pricing group.

Security Requirements -
Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.
Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:
Fabric administrators will be the workspace administrators.
The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.
The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.
The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook
The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.
The date dimension must be available to all users of the data store.
The principle of least privilege must be followed.
Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:
FabricAdmins: Fabric administrators

AnalyticsTeam: All the members of the analytics team
DataAnalysts: The data analysts on the analytics team
DataScientists: The data scientists on the analytics team
DataEngineers: The data engineers on the analytics team
AnalyticsEngineers: The analytics engineers on the analytics team

Report Requirements -
The data analysts must create a customer satisfaction report that meets the following requirements:
Enables a user to select a product to filter customer survey responses to only those who have purchased that product.
Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected dat.
Shows data as soon as the data is updated in the data store.
Ensures that the report and the semantic model only contain data from the current and previous year.
Ensures that the report respects any table-level security specified in the source data store.
Minimizes the execution time of report queries.
You need to create a DAX measure to calculate the average overall satisfaction score.
How should you complete the DAX code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

```
Rolling 12 Overall Satisfaction =
VAR NumberOfMonths = 12
VAR LastCurrentDate = MAX ( 'Date' [Date] )
VAR Period = DATESINPERIOD ( 'Date' [Date], LastCurrentDate,
- NumberOfMonths, MONTH )
VAR Result =
  CALCULATE (
```

[▼]
AVERAGE('Survey'[Response Value]),
AVERAGEA('Question'[Question Text]),
AVERAGEX(VALUES('Survey'[Customer Key]),

[▼]
LastCurrentDate,
NumberOfMonths,
Period,

```
    'Survey Question'[Question Title] = "Overall Satisfaction"
  )
  RETURN
    Result
```

Answer Area

```
Rolling 12 Overall Satisfaction =
VAR NumberOfMonths = 12
VAR LastCurrentDate = MAX ( 'Date' [Date] )
VAR Period = DATESINPERIOD ( 'Date' [Date], LastCurrentDate,
- NumberOfMonths, MONTH )
VAR Result =
  CALCULATE (
```

Correct Answer:

[▼]
AVERAGE('Survey'[Response Value]),
AVERAGEA('Question'[Question Text]),
AVERAGEX(VALUES('Survey'[Customer Key]),

[▼]
LastCurrentDate,
NumberOfMonths,
Period,

```
    'Survey Question'[Question Title] = "Overall Satisfaction"
  )
  RETURN
    Result
```

theseon Highly Voted 1 week, 6 days ago

Iwould say:
AVERAGE('Survery'[Respone Value]) and Period

I do not unterstand why we would use the CustomerKey for Average Calculation.

upvoted 6 times

SamuComqi Most Recent 4 days, 5 hours ago

For the first part, "AVERAGE('Survey'[Response Value])" because the second option uses a text column as argument, and the third option is not relevant in this context (no need to perform row calculations).
For the second part, "Period": data is filtered to compute the average in the last 12 months (interval defined in the variable).

upvoted 1 times

Momoanwar 4 days, 15 hours ago

Average and Period

upvoted 2 times

Nicknamefordiscussions69 1 week, 1 day ago

Response Value and Period

upvoted 1 times

HOTSPOT -

Case study -
This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.
To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.
At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -
To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -
Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

Existing Environment -

Fabric Environment -
Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

Available Data -
Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

Survey -

Question -

Response -
For each survey submitted, the following occurs:
One row is added to the Survey table.
One row is added to the Response table for each question in the survey.
The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score.
Customers can submit a survey after each purchase.

User Problems -
The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.
Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic

models, but the logic does NOT always match across implementations.

Requirements -

Planned Changes -
Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity
The following three workspaces will be created:
AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store
DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake
DataSciPOC: Will contain all the notebooks and reports created by the data scientists
The following will be created in the AnalyticsPOC workspace:
A data store (type to be decided)

A custom semantic model -

A default semantic model -

Interactive reports -
The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.
All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

Technical Requirements -
The data store must support the following:
Read access by using T-SQL or Python
Semi-structured and unstructured data
Row-level security (RLS) for users executing T-SQL queries
Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.
Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model
The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model
The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.
The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SOL. queries and in the default semantic model. The following logic must be used:
List prices that are less than or equal to 50 are in the low pricing group.
List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.
List prices that are greater than 1,000 are in the high pricing group.

Security Requirements -
Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.
Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:
Fabric administrators will be the workspace administrators.
The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.
The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.
The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook
The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.
The date dimension must be available to all users of the data store.
The principle of least privilege must be followed.
Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:
FabricAdmins: Fabric administrators

AnalyticsTeam: All the members of the analytics team
DataAnalysts: The data analysts on the analytics team
DataScientists: The data scientists on the analytics team
DataEngineers: The data engineers on the analytics team
AnalyticsEngineers: The analytics engineers on the analytics team

Report Requirements -
The data analysts must create a customer satisfaction report that meets the following requirements:
Enables a user to select a product to filter customer survey responses to only those who have purchased that product.
Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected dat.
Shows data as soon as the data is updated in the data store.
Ensures that the report and the semantic model only contain data from the current and previous year.
Ensures that the report respects any table-level security specified in the source data store.
Minimizes the execution time of report queries.
You need to resolve the issue with the pricing group classification.
How should you complete the T-SQL statement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

```
CREATE [ FUNCTION / PROCEDURE / TABLE / VIEW ] [dbo].[ProductsWithPricingGroup]
AS

SELECT ProductId,
     ProductName,
  ProductCategory,
  ListPrice,
  [ CASE / COALESCE / IIF / SET ]
   WHEN ListPrice <= 50 THEN 'low'
   [ WHEN(ListPrice > 50 AND ListPrice < 1000) THEN 'medium' /
     WHEN(ListPrice >= 50 AND ListPrice < 1000) THEN 'medium' /
     WHEN ListPrice BETWEEN 50 AND 1000) THEN 'medium' ]
   WHEN ListPrice > 1000 THEN 'high'
  END AS PricingGroup
  FROM dbo.Products
```

Answer Area

```
CREATE [ FUNCTION | PROCEDURE | TABLE | VIEW ] [dbo].[ProductsWithPricingGroup]
AS

SELECT ProductId,
     ProductName,
  ProductCategory,
  ListPrice,
  [ CASE | COALESCE | IIF | SET ]
    WHEN ListPrice <= 50 THEN 'low'
    [ WHEN(ListPrice > 50 AND ListPrice < 1000) THEN 'medium' |
      WHEN(ListPrice >= 50 AND ListPrice < 1000) THEN 'medium' |
      WHEN ListPrice BETWEEN 50 AND 1000) THEN 'medium' ]
    WHEN ListPrice > 1000 THEN 'high'
 END AS PricingGroup
FROM dbo.Products
```

Correct Answer: CASE; WHEN ListPrice BETWEEN 50 AND 1000) THEN 'medium'

SamuComqi 4 days, 5 hours ago

* VIEW: from an existing table.
* CASE: correct syntax before the WHENs.
* WHEN ListPrice BETWEEN 50 AND 1000 THEN 'medium': the other two options miss value 1000; on the other hand, the BETWEEN includes both 50 and 1000.

upvoted 1 times

Momoanwar 4 days, 15 hours ago

View
Case
Between

upvoted 2 times

Bharat 1 week ago

It has to be BETWEEN because the other two choices will miss 1000

upvoted 1 times

Nicofr 1 week ago

View => "Shows data as soon as the data is updated in the data store."
CASE
BETWEEN(50,1000) => the when clause for =50 is done in the first WHEN so it will never go in the second when if value is 50

upvoted 1 times

theseon 1 week, 6 days ago

1 and 2 correct but should 3 not be BETWEEN 51 AND 1000 since the BETWEEN command is inclusive

upvoted 2 times

Question #8 *Topic 1*

Case study -
This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.
To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.
At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -
To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -
Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

Existing Environment -

Fabric Environment -
Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

Available Data -
Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

Survey -

Question -

Response -
For each survey submitted, the following occurs:
One row is added to the Survey table.
One row is added to the Response table for each question in the survey.
The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score.
Customers can submit a survey after each purchase.

User Problems -
The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.
Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic models, but the logic does NOT always match across implementations.

Requirements -

Planned Changes -
Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity
The following three workspaces will be created:
AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store
DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake
DataSciPOC: Will contain all the notebooks and reports created by the data scientists
The following will be created in the AnalyticsPOC workspace:
A data store (type to be decided)

A custom semantic model -

A default semantic model -

Interactive reports -
The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.
All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

Technical Requirements -
The data store must support the following:
Read access by using T-SQL or Python
Semi-structured and unstructured data
Row-level security (RLS) for users executing T-SQL queries
Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.
Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model
The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model
The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.
The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SOL. queries and in the default semantic model. The following logic must be used:
List prices that are less than or equal to 50 are in the low pricing group.
List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.
List prices that are greater than 1,000 are in the high pricing group.

Security Requirements -
Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.
Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:
Fabric administrators will be the workspace administrators.
The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.
The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.
The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook
The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.
The date dimension must be available to all users of the data store.
The principle of least privilege must be followed.
Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:
FabricAdmins: Fabric administrators
AnalyticsTeam: All the members of the analytics team
DataAnalysts: The data analysts on the analytics team

DataScientists: The data scientists on the analytics team
DataEngineers: The data engineers on the analytics team
AnalyticsEngineers: The analytics engineers on the analytics team

Report Requirements -
The data analysts must create a customer satisfaction report that meets the following requirements:
Enables a user to select a product to filter customer survey responses to only those who have purchased that product.
Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected dat.
Shows data as soon as the data is updated in the data store.
Ensures that the report and the semantic model only contain data from the current and previous year.
Ensures that the report respects any table-level security specified in the source data store.
Minimizes the execution time of report queries.
What should you recommend using to ingest the customer data into the data store in the AnalyticsPOC workspace?

A. a stored procedure

B. a pipeline that contains a KQL activity

C. a Spark notebook

D. a dataflow

Correct Answer: *D*

Community vote distribution

D (100%)

David_Webb 1 day, 5 hours ago

Selected Answer: D

In the Interactive reports requirement, it stated, "Whenever possible, the data engineers will use low-code tools for data ingestion".

upvoted 1 times

SamuComqi 4 days, 5 hours ago

Selected Answer: D

D. a dataflow

Even though the text reads "Data will be loaded without transformation in one area of the AnalyticsPOC data store": in general, dataflows are used when data transformations are involved after ingestion. As suggested by user BHARAT, the Copy Activity should be the optimal solution.

upvoted 2 times

Momoanwar 4 days, 15 hours ago

Selected Answer: D

D see Bharat comment.

upvoted 1 times

Bharat 1 week ago

Ideally, It should be the COPY activity of the pipeline, but that is not given as a choice

upvoted 1 times

Nicofr 1 week, 2 days ago

Selected Answer: D

"Whenever possible, the data engineers will use low-code tools for data ingestion."

upvoted 2 times

theseon 1 week, 6 days ago

Selected Answer: D

Spark Notebook is also possible but i would say D is correct:
"Data will be loaded without transformation in one area of the AnalyticsPOC data store"

upvoted 1 times

Question #9 *Topic 1*

Case study -
This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.
To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.
At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -
To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -
Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

Existing Environment -

Fabric Environment -
Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

Available Data -
Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

Survey -

Question -

Response -
For each survey submitted, the following occurs:
One row is added to the Survey table.
One row is added to the Response table for each question in the survey.
The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score.
Customers can submit a survey after each purchase.

User Problems -
The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.
Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic models, but the logic does NOT always match across implementations.

Requirements -

Planned Changes -
Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity
The following three workspaces will be created:
AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store
DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake
DataSciPOC: Will contain all the notebooks and reports created by the data scientists
The following will be created in the AnalyticsPOC workspace:
A data store (type to be decided)

A custom semantic model -

A default semantic model -

Interactive reports -
The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.
All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

Technical Requirements -
The data store must support the following:
Read access by using T-SQL or Python
Semi-structured and unstructured data
Row-level security (RLS) for users executing T-SQL queries
Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.
Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model
The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model
The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.
The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SOL. queries and in the default semantic model. The following logic must be used:
List prices that are less than or equal to 50 are in the low pricing group.
List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.
List prices that are greater than 1,000 are in the high pricing group.

Security Requirements -
Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.
Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:
Fabric administrators will be the workspace administrators.
The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.
The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.
The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook
The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.
The date dimension must be available to all users of the data store.
The principle of least privilege must be followed.
Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:
FabricAdmins: Fabric administrators
AnalyticsTeam: All the members of the analytics team
DataAnalysts: The data analysts on the analytics team

DataScientists: The data scientists on the analytics team
DataEngineers: The data engineers on the analytics team
AnalyticsEngineers: The analytics engineers on the analytics team

Report Requirements -
The data analysts must create a customer satisfaction report that meets the following requirements:
Enables a user to select a product to filter customer survey responses to only those who have purchased that product.
Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected dat.
Shows data as soon as the data is updated in the data store.
Ensures that the report and the semantic model only contain data from the current and previous year.
Ensures that the report respects any table-level security specified in the source data store.
Minimizes the execution time of report queries.
Which type of data store should you recommend in the AnalyticsPOC workspace?

A. a data lake

B. a warehouse

C. a lakehouse

D. an external Hive metastore

Correct Answer: *C*

Community vote distribution

C (100%)

David_Webb 1 day, 4 hours ago

In the technical requirement, it stated "Semi-structured and unstructured data" for the AnalyticsPOC data store. Thus, it must be a lakehouse.

upvoted 1 times

SamuComqi 4 days, 5 hours ago

Selected Answer: C

C. a lakehouse

The data store must handle semi-structured and unstructured data, therefore a Lakehouse should be the optimal solution supporting read access with T-SQL and Python.

upvoted 1 times

Momoanwar 4 days, 15 hours ago

Selected Answer: C

Analytic=lakehouse

upvoted 1 times

Nicofr 1 week, 2 days ago

Selected Answer: C

"Read access by using T-SQL or Python
Semi-structured and unstructured data"

upvoted 2 times

theseon 1 week, 6 days ago

Selected Answer: C

C. a lakehouse

upvoted 2 times

You have a Fabric warehouse that contains a table named Staging.Sales. Staging.Sales contains the following columns.

Name	Data type	Nullable
ProductID	Integer	No
ProductName	Varchar(30)	No
SalesDate	Datetime2(6)	No
WholesalePrice	Decimal(18,2)	Yes
Amount	Decimal(18,2)	Yes

You need to write a T-SQL query that will return data for the year 2023 that displays ProductID and ProductName and has a summarized Amount that is higher than 10,000.

Which query should you use?

A.
```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR,SaleDate) = '2023'
GROUP BY ProductID, ProductName
HAVING SUM(Amount) > 10000
```

B.
```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
GROUP BY ProductID, ProductName
HAVING DATEPART(YEAR,SaleDate) = '2023' AND SUM(Amount) > 10000
```

C.
```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR,SaleDate) = '2023' AND SUM(Amount) > 10000
```

D.
```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR,SaleDate) = '2023'
GROUP BY ProductID, ProductName
HAVING TotalAmount > 10000
```

Correct Answer: *A*

Community vote distribution

A (100%)

David_Webb 1 day, 4 hours ago

Selected Answer: A

The answer is A, no-brainer.

upvoted 1 times

SamuComqi 4 days, 5 hours ago

Selected Answer: A

A. SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR, SaleDate) = '2023'
GROUP BY ProductID, ProductName
HAVING SUM(Amount) > 10000

Selected data is first filtered by Year, then grouped by ProductID and ProductName to compute the TotalAmount. Finally, only SUM(Amount) cane be used after HAVING (not the alias).

upvoted 1 times

Momoanwar 4 days, 15 hours ago

Selected Answer: A

Where to filter year data
Having to filter summerized data
Alias like TotalAmount not work in having

upvoted 1 times

objecto 1 week, 5 days ago

Selected Answer: A

TotalAmount can not be used with HAVING. You must use SUM(Amount)

upvoted 4 times

theseon 1 week, 6 days ago

Selected Answer: A

Summarized Amount by ProductID and ProductName -> Group BY
Above 10.000 -> HAVING SUM(Amount) ...
upvoted 3 times

Question #11 *Topic 1*

HOTSPOT -
You have a data warehouse that contains a table named Stage.Customers. Stage.Customers contains all the customer record updates from a customer relationship management (CRM) system. There can be multiple updates per customer.
You need to write a T-SQL query that will return the customer ID, name. postal code, and the last updated time of the most recent row for each customer ID.
How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

```
WITH CUSTOMERBASE AS (
     SELECT [CustomerID]
     ,[CustomerName]
     ,[PostalCode]
     ,[LastUpdated]
     ,x = [ LAST_Value() / NTILE() / ROW_NUMBER() ] OVER (PARTITION BY CustomerID
          ORDER BY LastUpdated DESC)
FROM [LakehousePOC].[dbo].[CustomerChanges]
     )
     SELECT CustomerID, CustomerName, PostalCode, LastUpdated
     FROM CUSTOMERBASE
     [ Having Max(LastUpdated) = 1 / WHERE LastUpdated = Max(LastUpdated) / WHERE X = 1 ]
```

Correct Answer: *Box 1: ROW_NUMBER()*
Box 2: WHERE X = 1

David_Webb 1 day, 4 hours ago
First drop-down box: ROW_NUMBER()
Second drop-down box: WHERE X = 1
As ORDER BY LastUpdated DESC was used, the first row will be the most recent row.
upvoted 1 times

Momoanwar 4 days, 15 hours ago
Row_number
X=1

Row_number give row position and start from 1
upvoted 2 times

R3D_ENGINEER 1 week, 2 days ago
The correct query is:
WITH CUSTOMERBASE AS (
SELECT CustomerID, CustomerName, PostalCode, LastUpdated,
ROW_NUMBER() OVER(PARTITION BY CustomerID ORDER BY LastUpdated DESC) as X
FROM LakehousePOC.dbo.CustomerChanges
)

SELECT CustomerID, CustomerName, PostalCode, LastUpdated
FROM CUSTOMERBASE
WHERE X = 1
upvoted 3 times

HOTSPOT -

You have a Fabric tenant.

You plan to create a Fabric notebook that will use Spark DataFrames to generate Microsoft Power BI visuals.

You run the following code.

```
from powerbiclient import QuickVisualize, get_dataset_config, Report

PBI_visualize = QuickVisualize(get_dataset_config(df))
PBI_visualize
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
The code embeds an existing Power BI report.	○	○
The code creates a Power BI report.	○	○
The code displays a summary of the DateFrame.	○	○

Correct Answer:

Answer Area

Statements	Yes	No
The code embeds an existing Power BI report.	○	◉
The code creates a Power BI report.	◉	○
The code displays a summary of the DateFrame.	◉	○

Momoanwar 4 days, 15 hours ago

I think : No yes no
To display summary its describe() or sumarry()

upvoted 1 times

wojciech_wie 4 days, 22 hours ago

https://learn.microsoft.com/en-us/power-bi/create-reports/jupyter-quick-report
1) No - based on dataframe (df)
2) Yes
3) Yes/No? they need to be more specific about it, because the code does not explicitly display a summary of the DataFrame, but PowerBI report based on DataFrame

upvoted 1 times

Question #13 *Topic 1*

You are the administrator of a Fabric workspace that contains a lakehouse named Lakehouse1. Lakehouse1 contains the following tables:
Table1: A Delta table created by using a shortcut
Table2: An external table created by using Spark

Table3: A managed table -
You plan to connect to Lakehouse1 by using its SQL endpoint.
What will you be able to do after connecting to Lakehouse1?

A. Read Table3.

B. Update the data Table3.

C. Read Table2.

D. Update the data in Table1.

Correct Answer: *D*

Community vote distribution

A (83%) Other

conie Highly Voted 1 week, 4 days ago

Selected Answer: A

B & D is out as you can't update a table in lakehouse using SQL endpoint as this is read only. You will need to use spark or dataflows.
C is out because when you create external table using spark, you can see the table from the lakehouse but you can't see the table from SQL endpoint let alone ready.

A is the answer, as I was able to see and read a managed table using SQL Endpoint

upvoted 7 times

BrandonPerks Most Recent 4 days, 1 hour ago

I can confirm it is A. I manually went to my fabric tenant to investigate.

upvoted 1 times

Momoanwar 4 days, 15 hours ago

Selected Answer: A

End point dont modify and not read external

upvoted 1 times

wojciech_wie 4 days, 22 hours ago

Answer A is correct
Using SQL Endpoint we can only READ tables, so B & D is out
Moreover for now we can't read external tables using SQL enpoint (https://community.fabric.microsoft.com/t5/General-Discussion/Fabric-SQL-end-point-not-showing-external-delta-tables/m-p/3475969)

upvoted 1 times

IshtarSQL 1 week, 5 days ago

Selected Answer: D

D is correct because you can update a Delta table.

upvoted 1 times

theseon 1 week, 6 days ago

Selected Answer: C

C
"The SQL analytics endpoint operates in read-only mode over lakehouse delta tables"
https://learn.microsoft.com/en-us/fabric/data-engineering/lakehouse-sql-analytics-endpoint

upvoted 1 times

Fermd 1 week, 6 days ago

Selected Answer: A

The right answer is A. A managed table, it is stored within the Fabric storage and becomes immediately accessible through the SQL endpoint upon connection.

D is not right becouse tables created using shortcuts might not be immediately accessible through the SQL endpoint for updates.

upvoted 2 times

Question #14 *Topic 1*

You have a Fabric tenant that contains a warehouse.
You use a dataflow to load a new dataset from OneLake to the warehouse.
You need to add a PowerQuery step to identify the maximum values for the numeric columns.
Which function should you include in the step?

A. Table.MaxN

B. Table.Max

C. Table.Range

D. Table.Profile

Correct Answer: *B*

Community vote distribution

B (50%) D (25%) A (25%)

lengzhai 3 days, 3 hours ago

Selected Answer: D

D. Table.Profile

https://learn.microsoft.com/en-us/powerquery-m/table-profile

upvoted 1 times

Momoanwar 4 days, 15 hours ago

Selected Answer: B

Ambigus i will sayed B.

upvoted 1 times

wojciech_wie 4 days, 22 hours ago

I think it should be C Table.Profile - for EACH COLUMN maximum
https://learn.microsoft.com/en-us/powerquery-m/table-profile

Table.Max return 1 row for SPECIFIC COLUMN
https://learn.microsoft.com/en-us/powerquery-m/table-max
Table.MaxN works similar like Max but we have additional condion
https://learn.microsoft.com/en-us/powerquery-m/table-maxn
Table.Range https://learn.microsoft.com/en-us/powerquery-m/table-range

upvoted 3 times

Nicofr 1 week, 2 days ago

Selected Answer: A

"the maximum valueS" it should be MaxN (https://learn.microsoft.com/en-us/powerquery-m/table-maxn)

upvoted 1 times

DataWeetom 1 week, 3 days ago

Selected Answer: B

https://learn.microsoft.com/en-us/powerquery-m/list-max

Voted B

upvoted 1 times

Question #15 *Topic 1*

You have a Fabric tenant that contains a machine learning model registered in a Fabric workspace.
You need to use the model to generate predictions by using the PREDICT function in a Fabric notebook.
Which two languages can you use to perform model scoring? Each correct answer presents a complete solution.
NOTE: Each correct answer is worth one point.

A. T-SQL

B. DAX

C. Spark SQL

D. PySpark

Correct Answer: *CD*

Community vote distribution

CD (100%)

David_Webb 1 day, 3 hours ago

Selected Answer: CD

Using Fabric notebook, thus must be C and D.

upvoted 1 times

BrandonPerks 4 days ago

The mention of Fabric Notebook, gives the hint to using Spark. Thus I went with CD

upvoted 1 times

Momoanwar 4 days, 15 hours ago

Selected Answer: CD

C & d in notebook

upvoted 2 times

wojciech_wie 4 days, 22 hours ago

CD - https://learn.microsoft.com/en-us/azure/synapse-analytics/machine-learning/tutorial-score-model-predict-spark-pool

upvoted 1 times

olavrab8 1 week, 2 days ago

Selected Answer: CD

Answer CD Correct
T-SQL Cannot be used, nor DAX

upvoted 2 times

Question #16 *Topic 1*

You are analyzing the data in a Fabric notebook.
You have a Spark DataFrame assigned to a variable named df.
You need to use the Chart view in the notebook to explore the data manually.
Which function should you run to make the data available in the Chart view?

A. displayHTML

B. show

C. write

D. display

Correct Answer: *D*

Community vote distribution

D (100%)

Momoanwar 4 days, 15 hours ago

Selected Answer: D

Display allow to see chart and inspect statistiques

upvoted 1 times

wojciech_wie 4 days, 22 hours ago

D is correct

upvoted 1 times

Question #17 *Topic 1*

You have a Fabric tenant that contains a Microsoft Power BI report named Report1. Report1 includes a Python visual.
Data displayed by the visual is grouped automatically and duplicate rows are NOT displayed.
You need all rows to appear in the visual.
What should you do?

A. Reference the columns in the Python code by index.

B. Modify the Sort Column By property for all columns.

C. Add a unique field to each row.

D. Modify the Summarize By property for all columns.

Correct Answer: *A*

Community vote distribution

D (100%)

Momoanwar 4 days, 15 hours ago

Selected Answer: D

Juste disable summurize in column on visual

upvoted 1 times

olavrab8 1 week, 2 days ago

Selected Answer: D

Summarization should be none

upvoted 2 times

Danialmellfoye 1 week, 3 days ago

Selected Answer: D

Simply right-click the column in visualization pane and choose "Don't summarize" option

upvoted 1 times

IshtarSQL 1 week, 5 days ago

A. Final Answer:
def find_indices(l, value):
return [
index for index, item in enumerate(l)
if item == value
]

upvoted 1 times

Fermd 1 week, 6 days ago

Selected Answer: D

The right answer is D: By setting the "Summarize By" property to "None" for all columns, you disable automatic aggregation and ensure all rows, including duplicates, are displayed in the Python visual.

upvoted 4 times

DRAG DROP -

You have a Fabric tenant that contains a semantic model. The model contains data about retail stores.

You need to write a DAX query that will be executed by using the XMLA endpoint. The query must return a table of stores that have opened since December 1, 2023.

How should you complete the DAX expression? To answer, drag the appropriate values to the correct 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

DEFINE

EVALUATE

FILTER

SUMMARIZE

TABLE

Answer Area

```
[          ]
VAR_SalesSince =
     DATE (2023, 12, 01)
[          ]
FILTER(
  [          ](Store, Store[Name], Store[OpenDate]),
   Store[OpenDate] >= _SalesSince
)
```

Correct Answer:

Answer Area

```
DEFINE
VAR_SalesSince =
     DATE (2023, 12, 01)
EVALUATE
FILTER(
  TABLE (Store, Store[Name], Store[OpenDate]),
   Store[OpenDate] >= _SalesSince
)
```

Jeff_Zhu 2 days, 13 hours ago

DEFINE
EVALUATE
SUMMARIZE

upvoted 1 times

Momoanwar 4 days, 15 hours ago

Define - evaluate - summarize

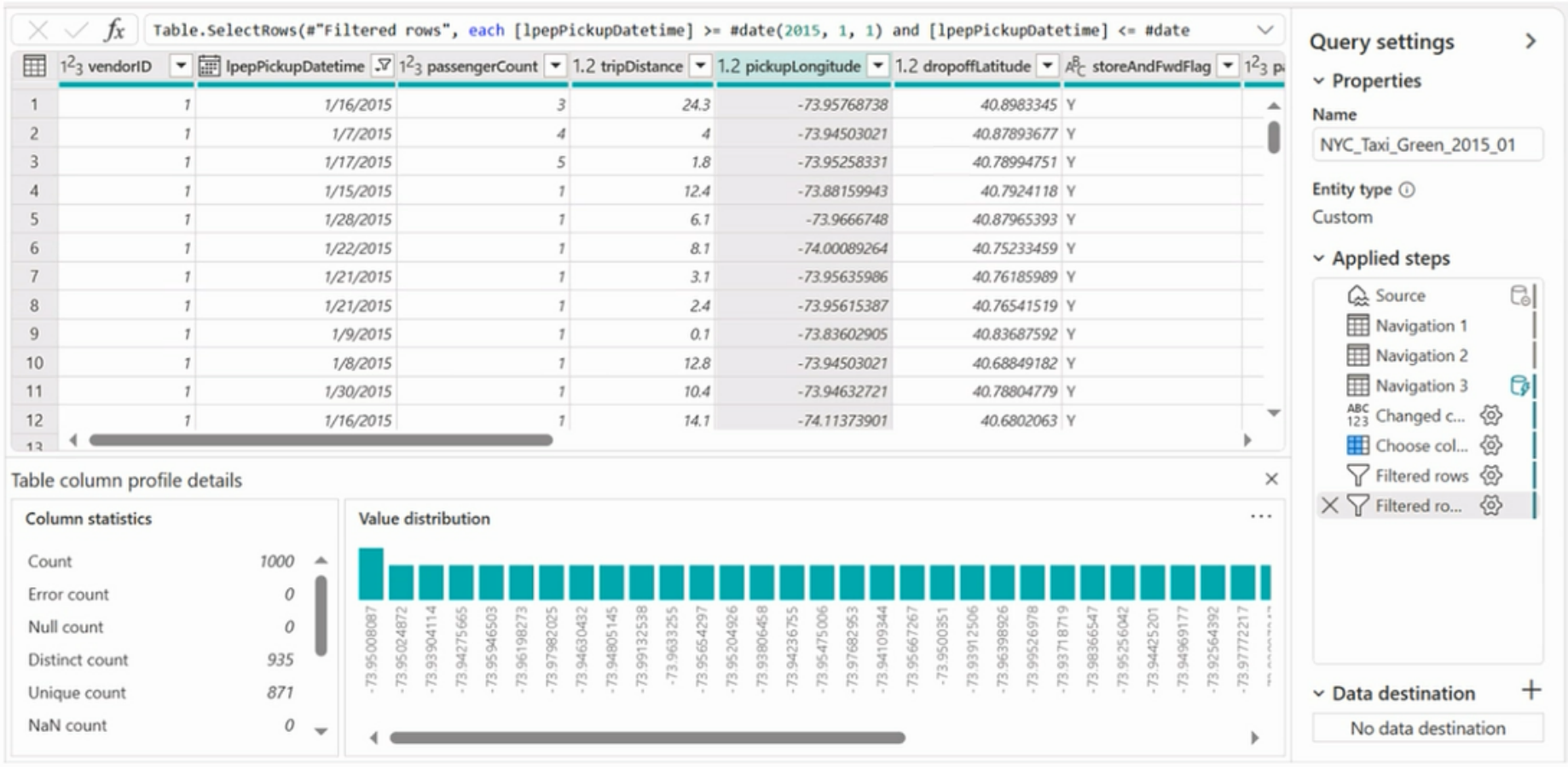
upvoted 2 times

wojciech_wie 4 days, 22 hours ago

1) DEFINE
2) EVALUATE
3) SUMMARIZE

upvoted 2 times

You have a Fabric workspace named Workspace1 that contains a dataflow named Dataflow1. Dataflow1 has a query that returns 2,000 rows. You view the query in Power Query as shown in the following exhibit.



What can you identify about the pickupLongitude column?

A. The column has duplicate values.

B. All the table rows are profiled.

C. The column has missing values.

D. There are 935 values that occur only once.

Correct Answer: *A*

Community vote distribution

A (100%)

Momoanwar 4 days, 15 hours ago

Selected Answer: A

Its A
Only one column selected here
No informations about missing values
Distinct count not mean exist only once

upvoted 2 times

nmosq 1 week, 4 days ago

Selected Answer: A

Answer A.

B - Not all the rows are profiled in the sample (only 1000 of 2000)
C- From the column statistics, you don't have any missing values in the sample
D- The values that occur only once are 871 (unique count)

upvoted 3 times

wojciech_wie 4 days, 22 hours ago

correct

upvoted 1 times

Question #20 *Topic 1*

You have a Fabric tenant named Tenant1 that contains a workspace named WS1. WS1 uses a capacity named C1 and contains a dataset named DS1.
You need to ensure read-write access to DS1 is available by using XMLA endpoint.
What should be modified first?

A. the DS1 settings

B. the WS1 settings

C. the C1 settings

D. the Tenant1 settings

Correct Answer: *C*

Community vote distribution

C (100%)

Momoanwar 4 days, 15 hours ago

Selected Answer: C

First check on capacity level

upvoted 1 times

wojciech_wie 4 days, 22 hours ago

C1
https://learn.microsoft.com/en-us/power-bi/enterprise/service-premium-connect-tools

upvoted 1 times

Nerd_Remo 1 week ago

C1 is Correct.

upvoted 1 times

Question #21 *Topic 1*

You have a Fabric tenant that contains a workspace named Workspace1. Workspace1 is assigned to a Fabric capacity.
You need to recommend a solution to provide users with the ability to create and publish custom Direct Lake semantic models by using external tools. The solution must follow the principle of least privilege.
Which three actions in the Fabric Admin portal should you include in the recommendation? Each correct answer presents part of the solution.
NOTE: Each correct answer is worth one point.

A. From the Tenant settings, set Allow XMLA Endpoints and Analyze in Excel with on-premises datasets to Enabled.

B. From the Tenant settings, set Allow Azure Active Directory guest users to access Microsoft Fabric to Enabled.

C. From the Tenant settings, select Users can edit data model in the Power BI service.

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

E. From the Tenant settings, set Users can create Fabric items to Enabled.

F. From the Tenant settings, enable Publish to Web.

Correct Answer: *ACD*

Community vote distribution

ACD (100%)

Momoanwar 4 days, 15 hours ago

Selected Answer: ACD

Correct ACD
B an d for external users
E to enable fabric

upvoted 1 times

Momoanwar 4 days, 15 hours ago

I mean B and F

upvoted 1 times

Question #22 *Topic 1*

You are creating a semantic model in Microsoft Power BI Desktop.
You plan to make bulk changes to the model by using the Tabular Model Definition Language (TMDL) extension for Microsoft Visual Studio Code.
You need to save the semantic model to a file.
Which file format should you use?

A. PBIP

B. PBIX

C. PBIT

D. PBIDS

Correct Answer: *B*

Community vote distribution

A (60%) C (40%)

Jeff_Zhu 2 days, 2 hours ago

Selected Answer: A

The answer is A
The PBIP will create one file and two folders, PBIP.Dataset contains definition folder that is use to host the .tmdl files

upvoted 2 times

lengzhai 3 days, 2 hours ago

Selected Answer: C

https://learn.microsoft.com/en-us/power-bi/create-reports/desktop-templates

upvoted 1 times

Momoanwar 4 days, 14 hours ago

Selected Answer: A

its A.
Pbit iS only template.
For source control use pbip it also generate bim file for TMDL.

upvoted 1 times

IshtarSQL 1 week, 5 days ago

Selected Answer: C

".pbit" file is a Power BI Desktop template file. It includes both the data model and any reports or visuals you've created.

upvoted 1 times

IshtarSQL 1 week, 5 days ago

My first thought would be BIM, but that is not an option. ?

upvoted 1 times

HOTSPOT -

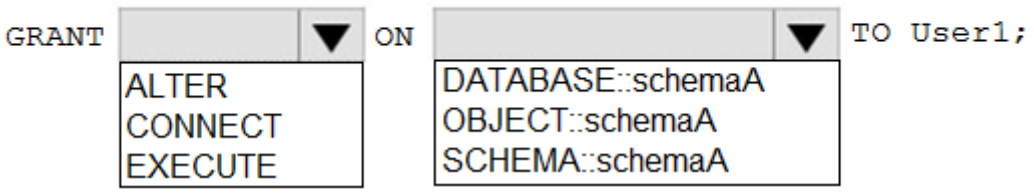
You have a Fabric tenant that contains a warehouse named Warehouse1. Warehouse1 contains three schemas named schemaA, schemaB, and schemaC.

You need to ensure that a user named User1 can truncate tables in schemaA only.

How should you complete the T-SQL statement? To answer, select the appropriate options in the answer area.

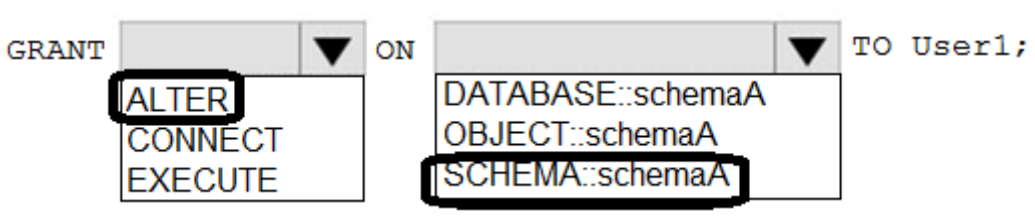
NOTE: Each correct selection is worth one point.

Answer Area



Answer Area

Correct Answer:



Momoanwar 4 days, 14 hours ago

Correct

upvoted 1 times

Question #24 *Topic 1*

You plan to deploy Microsoft Power BI items by using Fabric deployment pipelines. You have a deployment pipeline that contains three stages named Development, Test, and Production. A workspace is assigned to each stage.
You need to provide Power BI developers with access to the pipeline. The solution must meet the following requirements:
Ensure that the developers can deploy items to the workspaces for Development and Test.
Prevent the developers from deploying items to the workspace for Production.
Follow the principle of least privilege.
Which three levels of access should you assign to the developers? Each correct answer presents part of the solution.
NOTE: Each correct answer is worth one point.

A. Build permission to the production semantic models

B. Admin access to the deployment pipeline

C. Viewer access to the Development and Test workspaces

D. Viewer access to the Production workspace

E. Contributor access to the Development and Test workspaces

F. Contributor access to the Production workspace

Correct Answer: *ADE*

Community vote distribution

ADE (100%)

Momoanwar 4 days, 14 hours ago

Selected Answer: ADE

I think its correct if developper can read data on production.
Dont need b and c with E
F is too high

upvoted 1 times

Question #25 *Topic 1*

You have a Fabric workspace that contains a DirectQuery semantic model. The model queries a data source that has 500 million rows.
You have a Microsoft Power Bi report named Report1 that uses the model. Report1 contains visuals on multiple pages.
You need to reduce the query execution time for the visuals on all the pages.
What are two features that you can use? Each correct answer presents a complete solution,
NOTE: Each correct answer is worth one point.

A. user-defined aggregations

B. automatic aggregation

C. query caching

D. OneLake integration

Correct Answer: *BD*

Community vote distribution

AB (40%) AC (40%) BD (20%)

Momoanwar 5 days, 12 hours ago

Selected Answer: AB

D: onelake integration not for Direct Query
C: only at loading for first page

So AV

upvoted 2 times

BrandonPerks 3 days, 23 hours ago

Agreed to AB.
Both UDA's and AA optimize direct query performance. One just requires more manual work and in depth knowledge data modelling and query optimization techniques (UDA), whereas the other makes simplifies this process through the use of ML algorithms (AA).

upvoted 2 times

Momoanwar 5 days, 12 hours ago

I mean AB*

upvoted 1 times

Fermd 1 week, 2 days ago

Selected Answer: AC

A. User-defined aggregations (UDAs) allow you to pre-aggregate specific calculations directly in the semantic model. This reduces the amount of data that needs to be retrieved from the source each time a visual requires the calculation, significantly improving query execution time.
C. Power BI Desktop enables query caching for DirectQuery models. This stores frequently used queries on the client machine, eliminating the need to re-send them to the source data for subsequent interactions.

upvoted 2 times

Nicofr 1 week, 2 days ago

Selected Answer: BD

https://learn.microsoft.com/en-us/power-bi/enterprise/aggregations-auto
https://learn.microsoft.com/en-us/power-bi/enterprise/onelake-integration-overview

upvoted 1 times

Question #26 *Topic 1*

You have a Fabric tenant that contains 30 CSV files in OneLake. The files are updated daily.
You create a Microsoft Power BI semantic model named Model1 that uses the CSV files as a data source. You configure incremental refresh for Model1 and publish the model to a Premium capacity in the Fabric tenant.
When you initiate a refresh of Model1, the refresh fails after running out of resources.
What is a possible cause of the failure?

A. Query folding is occurring.

B. Only refresh complete days is selected.

C. XMLA Endpoint is set to Read Only.

D. Query folding is NOT occurring.

E. The delta type of the column used to partition the data has changed.

Correct Answer: *D*

Community vote distribution

D (33%) C (33%) E (33%)

Momoanwar 5 days, 12 hours ago

Selected Answer: D

https://learn.microsoft.com/en-us/power-bi/connect-data/incremental-refresh-troubleshoot#problem-loading-data-takes-too-long

upvoted 1 times

Nicofr 1 week, 1 day ago

Selected Answer: E

https://learn.microsoft.com/en-us/power-bi/connect-data/incremental-refresh-troubleshoot#problem-loading-data-takes-too-long

upvoted 1 times

Fermd 1 week, 2 days ago

Selected Answer: C

C. XMLA Endpoint is set to Read Only: If the XMLA endpoint for the Premium capacity is set to Read Only, any attempt to update or refresh the model through this endpoint, including incremental refresh, would fail. This configuration directly explains the resource exhaustion during a refresh operation as the read-only mode wouldn't allow the necessary updates to occur.

upvoted 1 times

Question #27 *Topic 1*

You have a Fabric tenant that uses a Microsoft Power BI Premium capacity.
You need to enable scale-out for a semantic model.
What should you do first?

A. At the semantic model level, set Large dataset storage format to Off.

B. At the tenant level, set Create and use Metrics to Enabled.

C. At the semantic model level, set Large dataset storage format to On.

D. At the tenant level, set Data Activator to Enabled.

Correct Answer: *C*

Community vote distribution

C (100%)

Momoanwar 4 days, 14 hours ago

Selected Answer: C

Correct

upvoted 1 times

Nicofr 1 week, 2 days ago

Selected Answer: C

https://learn.microsoft.com/en-us/power-bi/enterprise/service-premium-scale-out-configure

upvoted 2 times

Question #28 *Topic 1*

You have a Fabric tenant that contains a warehouse. The warehouse uses row-level security (RLS).
You create a Direct Lake semantic model that uses the Delta tables and RLS of the warehouse.
When users interact with a report built from the model, which mode will be used by the DAX queries?

A. DirectQuery

B. Dual

C. Direct Lake

D. Import

Correct Answer: *D*

Community vote distribution

A (100%)

Momoanwar 4 days, 14 hours ago

Selected Answer: A

Dax and fallback its direct query

upvoted 1 times

wojciech_wie 4 days, 21 hours ago

A. Direct Query "Row-level security only applies to queries on a Warehouse or SQL analytics endpoint in Fabric. Power BI queries on a warehouse in Direct Lake mode will fall back to Direct Query mode to abide by row-level security."
https://learn.microsoft.com/en-us/fabric/data-warehouse/row-level-security

upvoted 1 times

Dali2908 1 week, 2 days ago

Answer is A - DirectQuery.

upvoted 2 times

IshtarSQL 1 week, 5 days ago

Known issues and limitations
Currently, Direct Lake models can only contain tables and views from a single Lakehouse or Data Warehouse. However, tables in the model based on T-SQL-based views cannot be queried in Direct Lake mode. DAX queries that use these model tables fall back to DirectQuery mode.
https://learn.microsoft.com/en-us/power-bi/enterprise/directlake-overview

upvoted 1 times

IshtarSQL 1 week, 5 days ago

Selected Answer: A

When users interact with a report built from a Direct Lake semantic model, which leverages row-level security (RLS) and Delta tables from a warehouse, the DAX queries will operate in DirectQuery mode.

upvoted 2 times

Question #29 *Topic 1*

You have a Fabric tenant that contains a complex semantic model. The model is based on a star schema and contains many tables, including a fact table named Sales.
You need to create a diagram of the model. The diagram must contain only the Sales table and related tables.
What should you use from Microsoft Power BI Desktop?

A. data categories

B. Data view

C. Model view

D. DAX query view

Correct Answer: *C*

Community vote distribution

C (100%)

SamuComqi 3 days, 6 hours ago

Selected Answer: C

C. Model view

In the Model view, it is possible to analyze the semantic model and create new layouts.

upvoted 1 times

Momoanwar 4 days, 14 hours ago

Selected Answer: C

Model = model view

upvoted 1 times

Question #30 *Topic 1*

You have a Fabric tenant that contains a semantic model. The model uses Direct Lake mode.
You suspect that some DAX queries load unnecessary columns into memory.
You need to identify the frequently used columns that are loaded into memory.
What are two ways to achieve the goal? Each correct answer presents a complete solution.
NOTE: Each correct answer is worth one point.

A. Use the Analyze in Excel feature.

B. Use the Vertipaq Analyzer tool.

C. Query the $System.DISCOVER_STORAGE_TABLE_COLUMN_SEGMENTS dynamic management view (DMV).

D. Query the DISCOVER_MEMORYGRANT dynamic management view (DMV).

Correct Answer: *AC*

Community vote distribution

BC (100%)

Momoanwar 5 days, 12 hours ago

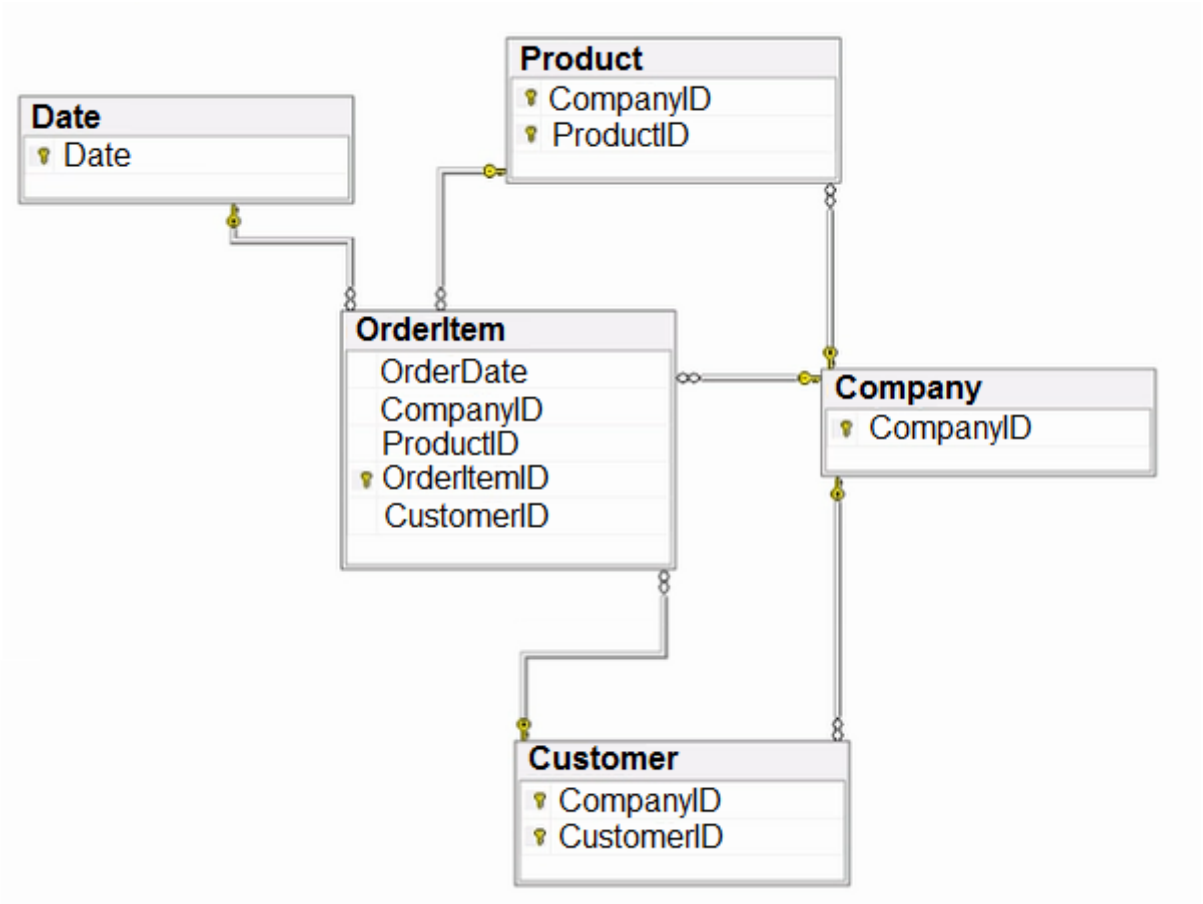
Selected Answer: BC

I think BC.
A is only tobread data and D only memory allocations

upvoted 2 times

HOTSPOT -

You have the source data model shown in the following exhibit.



The primary keys of the tables are indicated by a key symbol beside the columns involved in each key.

You need to create a dimensional data model that will enable the analysis of order items by date, product, and customer.

What should you include in the solution? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

The relationship between OrderItem and Product must be based on:

- The ProductID column
- Both the CompanyID and the ProductID columns
- A new key that combines the CompanyID and ProductID columns

The Company entity must be:

- Omitted
- Denormalized into the Product entity only
- Denormalized into the Customer and Product entities

Correct Answer:

Answer Area

The relationship between OrderItem and Product must be based on:

- The ProductID column
- Both the CompanyID and the ProductID columns
- A new key that combines the CompanyID and ProductID columns

The Company entity must be:

- Omitted
- Denormalized into the Product entity only
- Denormalized into the Customer and Product entities

Momoanwar 4 days, 14 hours ago

Correct

upvoted 1 times

Question #32 *Topic 1*

You have a Fabric tenant that contains a semantic model named Model1. Model1 uses Import mode. Model1 contains a table named Orders. Orders has 100 million rows and the following fields.

Name	Data type	Description
OrderID	Integer	Column imported from the source
OrderDateTime	Date/Time	Column imported from the source
Quantity	Integer	Column imported from the source
Price	Decimal	Column imported from the source
TotalSalesAmount	Decimal	Calculated column that multiplies Quantity and Price
TotalQuantity	Integer	Measure

You need to reduce the memory used by Model1 and the time it takes to refresh the model.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

A. Split OrderDateTime into separate date and time columns.

B. Replace TotalQuantity with a calculated column.

C. Convert Quantity into the Text data type.

D. Replace TotalSalesAmount with a measure.

Correct Answer: *BD*

Community vote distribution

AD (100%)

Momoanwar 5 days, 11 hours ago

Selected Answer: AD

Its AD

upvoted 2 times

Momoanwar 5 days, 11 hours ago

A : Best practice
D : measure better than column

upvoted 2 times

Nicofr 1 week ago

Selected Answer: AD

A should compress the memory size
D should reduce the memory usage but

upvoted 2 times

fabric1 1 week, 2 days ago

I was under the impression that A should be correct due to the fact that separate date and time column achieve higher columnar redundancy and allow better data compression. Whereas solution B, an additional calculated column, would inflate the memory usage. Measure are not stored in memory and would therefore be favorable concerning the stated objective of reducing memory used and minimizing refresh times.

upvoted 1 times

Question #33 *Topic 1*

You have a Fabric tenant that contains a semantic model.
You need to prevent report creators from populating visuals by using implicit measures.
What are two tools that you can use to achieve the goal? Each correct answer presents a complete solution.
NOTE: Each correct answer is worth one point.

A. Microsoft Power BI Desktop

B. Tabular Editor

C. Microsoft SQL Server Management Studio (SSMS)

D. DAX Studio

Correct Answer: *AC*

Community vote distribution

AB (100%)

Momoanwar 4 days, 14 hours ago

Selected Answer: AB

Ssms have nothing to do here and dax studio not alter model

upvoted 2 times

IshtarSQL 1 week, 5 days ago

Selected Answer: AB

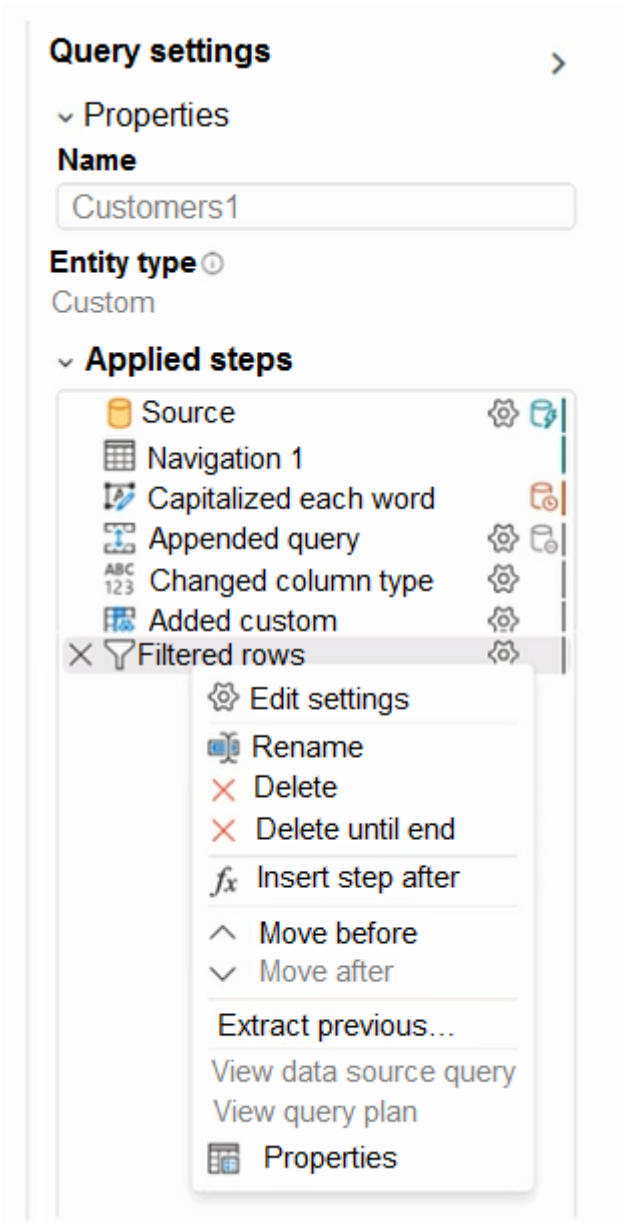
AB: To prevent report creators from populating visuals using implicit measures in a Power BI semantic model within a Fabric tenant, you can utilize the following tools:
1. Tabular Editor:
2. Power BI Desktop (Data Model View):

upvoted 2 times

HOTSPOT -

You have a Fabric tenant that contains two lakehouses.

You are building a dataflow that will combine data from the lakehouses. The applied steps from one of the queries in the dataflow is shown in the following exhibit.



Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.

NOTE: Each correct selection is worth one point.

Answer Area

[Answer choice] of the transformation steps in the query will fold. [▼]
- All
- None
- Some

The Added custom step will be performed in **[answer choice]**. [▼]
- each lakehouse's query engine
- the Microsoft Power Query engine
- the source lakehouse query engine

Correct Answer:

Answer Area

[Answer choice] of the transformation steps in the query will fold. [▼]
- All
- None
- **Some**

The Added custom step will be performed in **[answer choice]**. [▼]
- each lakehouse's query engine
- **the Microsoft Power Query engine**
- the source lakehouse query engine

SamuComqi 4 days, 5 hours ago

The red icon suggests where query folding is interrupted, therefore only SOME steps are folded. All the other steps after the red icon will not fold and will be executed by the POWER QUERY engine.

upvoted 1 times

Momoanwar 5 days, 11 hours ago

1: Some.blue icone fold and red dont fold
2 : on power query. all process after the step that dont fold are on power query

upvoted 3 times

Question #35 *Topic 1*

You have a Fabric tenant that contains a lakehouse named Lakehouse'. Lakehouse1 contains a table named Tablet.
You are creating a new data pipeline.
You plan to copy external data to Table'. The schema of the external data changes regularly.
You need the copy operation to meet the following requirements:
Replace Table1 with the schema of the external data.
Replace all the data in Table1 with the rows in the external data.
You add a Copy data activity to the pipeline.
What should you do for the Copy data activity?

A. From the Source tab, add additional columns.

B. From the Destination tab, set Table action to Overwrite.

C. From the Settings tab, select Enable staging.

D. From the Source tab, select Enable partition discovery.

E. From the Source tab, select Recursively.

Correct Answer: *B*

Community vote distribution

B (100%)

Momoanwar 4 days, 14 hours ago

Selected Answer: B

Replace all with overwrite

upvoted 2 times

IshtarSQL 1 week, 5 days ago

Selected Answer: B

Correct: Enable "Truncate table" option: This option truncates the target table before copying data, ensuring that all existing data is replaced with the new data from the external source.

upvoted 1 times

Question #36 *Topic 1*

You have a Fabric tenant that contains a lakehouse.
You plan to query sales data files by using the SQL endpoint. The files will be in an Amazon Simple Storage Service (Amazon S3) storage bucket.
You need to recommend which file format to use and where to create a shortcut.
Which two actions should you include in the recommendation? Each correct answer presents part of the solution.
NOTE: Each correct answer is worth one point.

A. Create a shortcut in the Files section.

B. Use the Parquet format

C. Use the CSV format.

D. Create a shortcut in the Tables section.

E. Use the delta format.

Correct Answer: *CD*

Community vote distribution

BD (60%) DE (40%)

Momoanwar 4 days, 13 hours ago

Selected Answer: BD

I think its BD.
Delta is not file format, lakehouse will read parquet files and create delta table.

upvoted 1 times

Nicofr 1 week, 1 day ago

Selected Answer: DE

delta format like the default format for Fabric

upvoted 2 times

Bharat 4 days, 18 hours ago

Two things make the Parquet format a better choice: 1. the question says Query, and 2. It is a shortcut to S3 therefore you don't need the bells and whistles of the Delta format.

upvoted 2 times

IshtarSQL 1 week, 5 days ago

Selected Answer: BD

B: You should use a columnar file format such as Parquet or ORC (Optimized Row Columnar). These formats are highly optimized for analytical queries and provide efficient storage and query performance.
D: In the Tables section of your lakehouse, you define virtual tables that represent external data sources. These virtual tables can be backed by data stored externally in formats such as Parquet or ORC in Amazon S3.

upvoted 2 times

Question #37 *Topic 1*

You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a subfolder named Subfolder1 that contains CSV files.
You need to convert the CSV files into the delta format that has V-Order optimization enabled.
What should you do from Lakehouse explorer?

A. Use the Load to Tables feature.

B. Create a new shortcut in the Files section.

C. Create a new shortcut in the Tables section.

D. Use the Optimize feature.

Correct Answer: *A*

Community vote distribution

A (86%) 14%

Sanji931 2 days, 14 hours ago

Selected Answer: A

With "Load to tables" : tables are always loaded using the Delta Lake table format with V-Order optimization enabled.
https://learn.microsoft.com/en-us/fabric/data-engineering/load-to-tables#load-to-table-capabilities-overview

upvoted 2 times

Momoanwar 4 days, 13 hours ago

Selected Answer: A

Correct

upvoted 2 times

Nicofr 1 week, 1 day ago

Selected Answer: A

"Load to Tables" functionality, the Optimize check mark is set by default.

upvoted 1 times

fabric1 1 week, 2 days ago

Selected Answer: A

Hi IshtarSQL,
the "Optimize" feature is only applicable on already existing tables and cannot convert CSV files as far as I know. When loading a CSV file as table using the "Load to Tables" functionality, the Optimize check mark is set by default. Therefore, A should be correct.
Cheers fabric1

upvoted 1 times

IshtarSQL 1 week, 5 days ago

Selected Answer: D

The "New Optimize" feature in the Lakehouse Explorer to convert CSV files into Delta format with V-Order optimization enabled.
The "New Optimize" feature allows you to optimize your data in Delta Lake format, including enabling V-Order optimization. V-Order optimization improves query performance by organizing data according to the values of frequently queried columns.

upvoted 1 times

Question #38 *Topic 1*

You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains an unpartitioned table named Table1.
You plan to copy data to Table1 and partition the table based on a date column in the source data.
You create a Copy activity to copy the data to Table1.
You need to specify the partition column in the Destination settings of the Copy activity.
What should you do first?

A. From the Destination tab, set Mode to Append.

B. From the Destination tab, select the partition column.

C. From the Source tab, select Enable partition discovery.

D. From the Destination tabs, set Mode to Overwrite.

Correct Answer: *D*

Community vote distribution

D (67%) C (33%)

lengzhai 2 days, 23 hours ago

Selected Answer: D

D is correct. Partition is available when Overwrite is checked

upvoted 1 times

SamuComqi 4 days, 5 hours ago

Selected Answer: D

D. From the Destination tabs, set Mode to Overwrite.

When setting up the Copy Activity, you need to choose the Overwrite mode to make the partition option appear (not visibile in Append mode).

upvoted 1 times

Momoanwar 5 days, 8 hours ago

Selected Answer: C

I think its C. For partition in pipeline we have to specify column in source

upvoted 1 times

Bharat 4 days, 18 hours ago

The table is unpartitioned to begin with.

upvoted 1 times

Question #39 *Topic 1*

HOTSPOT -
You have a Fabric tenant that contains a warehouse named Warehouse1. Warehouse1 contains a fact table named FactSales that has one billion rows.
You run the following T-SQL statement.
CREATE TABLE test.FactSales AS CLONE OF Dbo.FactSales;
For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
A replica of `dbo.Sales` is created in the test schema by copying the metadata only.	○	○
Additional schema changes to `dbo.FactSales` will also apply to `test.FactSales`.	○	○
Additional schema changes to `dbo.FactSales` will also apply to `test.FactSales`.	○	○

Correct Answer:

Answer Area

Statements	Yes	No
A replica of `dbo.Sales` is created in the test schema by copying the metadata only.	●	○
Additional schema changes to `dbo.FactSales` will also apply to `test.FactSales`.	○	●
Additional schema changes to `dbo.FactSales` will also apply to `test.FactSales`.	○	●

SamuComqi 4 days, 5 hours ago
Y - N - N

The AS CLONE AS creates a replica of the original table by copying the metadata (no data).
The two copies are independent therefore any changes will not be inherited.

Source: https://learn.microsoft.com/en-us/fabric/data-warehouse/clone-table
upvoted 2 times

Question #40 *Topic 1*

You have source data in a folder on a local computer.
You need to create a solution that will use Fabric to populate a data store. The solution must meet the following requirements:
Support the use of dataflows to load and append data to the data store.
Ensure that Delta tables are V-Order optimized and compacted automatically.
Which type of data store should you use?

A. a lakehouse

B. an Azure SQL database

C. a warehouse

D. a KQL database

Correct Answer: *C*

Community vote distribution

A (100%)

Momoanwar 4 days, 13 hours ago

Selected Answer: A

Delta table... = Lakehouse

upvoted 1 times

IshtarSQL 1 week, 5 days ago

Selected Answer: A

To meet the requirements of supporting dataflows to load and append data to the data store while ensuring that Delta tables are V-Order optimized and compacted automatically, you should use a lakehouse in Fabric as your solution.

upvoted 3 times

Question #41 *Topic 1*

HOTSPOT -

You have a Fabric tenant that contains a lakehouse.

You are using a Fabric notebook to save a large DataFrame by using the following code. df.write.partitionBy("year", "month", "day").mode("overwrite").parquet("Files/SalesOrder")

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
The results will form a hierarchy of folders for each partition key.	○	○
The resulting file partitions can be read in parallel across multiple nodes.	○	○
The resulting file partitions will use file compression.	○	○

Correct Answer:

Answer Area

Statements	Yes	No
The results will form a hierarchy of folders for each partition key.	●	○
The resulting file partitions can be read in parallel across multiple nodes.	●	○
The resulting file partitions will use file compression.	○	●

Momoanwar 4 days, 13 hours ago

I think yes yes yes.
Parquet= compression

upvoted 1 times

You have a Fabric workspace named Workspace1 that contains a data flow named Dataflow1 contains a query that returns the data shown in the following exhibit.

You need to transform the data columns into attribute-value pairs, where columns become rows.

You select the VendorID column.

Which transformation should you select from the context menu of the VendorID column?

A. Group by

B. Unpivot columns

C. Unpivot other columns

D. Split column

E. Remove other columns

Correct Answer: *C*

Community vote distribution

C (100%)

Momoanwar 4 days, 13 hours ago

Selected Answer: C

Correct

upvoted 2 times

Question #43 *Topic 1*

You have a Fabric tenant that contains a data pipeline.
You need to ensure that the pipeline runs every four hours on Mondays and Fridays.
To what should you set Repeat for the schedule?

A. Daily

B. By the minute

C. Weekly

D. Hourly

Correct Answer: *D*

Community vote distribution

C (83%) A (17%)

Sanji931 2 days, 13 hours ago

Selected Answer: C

Answer C : Weekly.
The only way to do this is to set the schedule to ''Weekly'', set the days on Monday and Friday and add manually 6 Time of 4 hour intervals.

upvoted 2 times

Momoanwar 4 days, 13 hours ago

Selected Answer: C

Weekly

upvoted 1 times

Nicofr 1 week ago

Selected Answer: C

Weekly allow to choose the days and the time

upvoted 1 times

fabric1 1 week, 2 days ago

Selected Answer: C

Hi IshtarSQL,
the selection of individual week days is only possible in the "Weekly" schedule option. Therefore, C is correct here.

upvoted 1 times

IshtarSQL 1 week, 5 days ago

Selected Answer: A

To ensure that the pipeline runs every four hours on Mondays and Fridays, you should set the "Repeat" frequency for the schedule to "Daily" and set the "Interval" to 4 hours. Then, you can specify the days of the week when the pipeline should run by selecting only Mondays and Fridays.

upvoted 1 times

Question #44 *Topic 1*

You have a Fabric tenant that contains a warehouse.
Several times a day, the performance of all warehouse queries degrades. You suspect that Fabric is throttling the compute used by the warehouse.
What should you use to identify whether throttling is occurring?

A. the Capacity settings

B. the Monitoring hub

C. dynamic management views (DMVs)

D. the Microsoft Fabric Capacity Metrics app

Correct Answer: *D*

Community vote distribution

D (100%)

Momoanwar 4 days, 13 hours ago

Selected Answer: D

Microsoft Fabric Capacity Metrics app for trottling

upvoted 1 times

HOTSPOT -

You have a Fabric workspace that uses the default Spark starter pool and runtime version 1.2.

You plan to read a CSV file named Sales_raw.csv in a lakehouse, select columns, and save the data as a Delta table to the managed area of the lakehouse. Sales_raw.csv contains 12 columns.

You have the following code.

```
from pyspark.sql.functions import year

(spark
 .red
 .format("csv")
 .option("heade", 'true')
 .load("Files/sales_raw.csv")
 .select('SalesOrderNumber', 'OrderDate', 'CustomerName',
'UnitPrice')
 .withColumn("Year", year("OrderDate"))
 .write
 .partitionBy('Year')
 .saveAsTable("sales")
)
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
The Spark engine will read only the 'SalesOrderNumber', 'OrderDate', 'CustomerName', 'UnitPrice' columns from Sales_raw.csv.	○	○
Removing the partition will reduce the execution time of the query.	○	○
Adding `inferSchema= 'true'` to the options will increase the execution time of the query.	○	○

Answer Area

	Statements	Yes	No
Correct Answer:	The Spark engine will read only the 'SalesOrderNumber', 'OrderDate', 'CustomerName', 'UnitPrice' columns from Sales_raw.csv.	◉	○
	Removing the partition will reduce the execution time of the query.	○	◉
	Adding `inferSchema= 'true'` to the options will increase the execution time of the query.	◉	○

Momoanwar 4 days, 13 hours ago

Its read not red.
This question is ambiguous would say : no no yes.
For the point 1 : with case sensitivity sales_raw is not Sales_raw

upvoted 1 times

Question #46 *Topic 1*

You have a Fabric tenant that contains a warehouse.
A user discovers that a report that usually takes two minutes to render has been running for 45 minutes and has still not rendered.
You need to identify what is preventing the report query from completing.
Which dynamic management view (DMV) should you use?

A. sys.dm_exec_requests

B. sys.dm_exec_sessions

C. sys.dm_exec_connections

D. sys.dm_pdw_exec_requests

Correct Answer: *D*

Community vote distribution

A (100%)

Sanji931 2 days, 12 hours ago

Selected Answer: A

Answer is A.
https://learn.microsoft.com/en-us/fabric/data-warehouse/monitor-using-dmv

upvoted 1 times

DRAG DROP -

You are creating a data flow in Fabric to ingest data from an Azure SQL database by using a T-SQL statement.

You need to ensure that any foldable Power Query transformation steps are processed by the Microsoft SQL Server engine.

How should you complete the code? To answer, drag the appropriate values to the correct 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

- EnableFolding
- NativeQuery
- Optimize
- Record
- StopFolding
- Table
- Value

Answer Area

```
let
     Source = Sql.Databases(
          "server.database.windows.net"
     ),
     Database = Source{[Name = "db"]}[Data],
     Query = [          ] . [          ] (
     Database,
          " SELECT * FROM customer WHERE country IN ('USA', 'UK')",
          null,
          [          ] = true]
     )
in
     Query
```

Correct Answer:

Answer Area

```
let
     Source = Sql.Databases(
          "server.database.windows.net"
     ),
     Database = Source{[Name = "db"]}[Data],
     Query = [ Value ] . [ NativeQuery ] (
     Database,
          " SELECT * FROM customer WHERE country IN ('USA', 'UK')",
          null,
          [ EnableFolding ] = true]
     )
in
     Query
```

SamuComqi 4 days, 5 hours ago

* Value
* NativeQuery
* EnableFolding

Source: https://learn.microsoft.com/en-us/power-query/native-query-folding

upvoted 1 times

Momoanwar 4 days, 19 hours ago

Correct

upvoted 1 times

DRAG DROP -

You have a Fabric tenant that contains a lakehouse named Lakehouse1.

Readings from 100 IoT devices are appended to a Delta table in Lakehouse1. Each set of readings is approximately 25 KB. Approximately 10 GB of data is received daily.

All the table and SparkSession settings are set to the default.

You discover that queries are slow to execute. In addition, the lakehouse storage contains data and log files that are no longer used.

You need to remove the files that are no longer used and combine small files into larger files with a target size of 1 GB per file.

What should you do? To answer, drag the appropriate actions to the correct requirements. Each action 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.

Actions

- Set the autoCompact table setting.
- Set the optimizeWrite table setting.
- Run the VACUUM command on a schedule.
- Set the autoCompact SparkSession setting.
- Run the OPTIMIZE command on a schedule.
- Set the parallelDelete SparkSession setting.

Answer Area

Remove the files: [ ]

Combine the files: [ ]

Answer Area

Correct Answer:

Remove the files: Run the VACUUM command on a schedule.

Combine the files: Run the OPTIMIZE command on a schedule.

SamuComqi 3 days, 22 hours ago

VACUUM: to remove old files no longer referenced.
OPTIMIZE: to create fewer files with a larger size.

Sources:
* https://learn.microsoft.com/en-us/fabric/data-engineering/delta-optimization-and-v-order?tabs=sparksql
* VACUUM: https://docs.delta.io/latest/delta-utility.html#-delta-vacuum
* OPTIMIZE: https://docs.delta.io/latest/optimizations-oss.html

upvoted 2 times

Momoanwar 4 days, 18 hours ago

Correct :
OPTIMIZE Improves query performance by optimizing file sizes. See Compact data files with optimize on Delta Lake.
VACUUM Reduces storage costs by deleting data files no longer referenced by the table. See Remove unused data files with vacuum.

upvoted 2 times

Question #49 *Topic 1*

You need to create a data loading pattern for a Type 1 slowly changing dimension (SCD).
Which two actions should you include in the process? Each correct answer presents part of the solution.
NOTE: Each correct answer is worth one point.

A. Update rows when the non-key attributes have changed.

B. Insert new rows when the natural key exists in the dimension table, and the non-key attribute values have changed.

C. Update the effective end date of rows when the non-key attribute values have changed.

D. Insert new records when the natural key is a new value in the table.

Correct Answer: *AC*

Community vote distribution

AD (100%)

Momoanwar 4 days, 18 hours ago

Selected Answer: AD

No history for scd1

upvoted 1 times

fabric1 1 week, 2 days ago

Selected Answer: AD

Type 1 SCD does not preserve history, therefore no end dates for table entries exists. A and D are correct.

upvoted 2 times

Question #50 *Topic 1*

HOTSPOT -

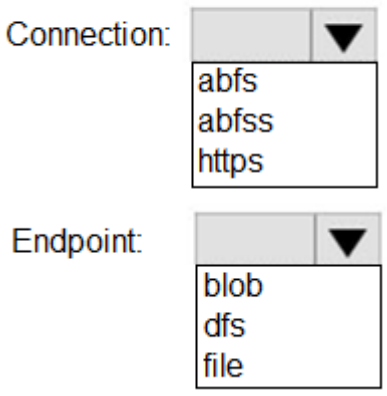
You have a Fabric workspace named Workspace1 and an Azure Data Lake Storage Gen2 account named storage1. Workspace1 contains a lakehouse named Lakehouse1.

You need to create a shortcut to storage1 in Lakehouse1.

Which connection and endpoint should you specify? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area



Correct Answer: *Box 1: abfss -*

Box 2: dfs

Momoanwar 4 days, 18 hours ago

Correct path exemple :
abfss://Dev@onelake.dfs.fabric.microsoft.com/lakehouse1...

upvoted 1 times

IshtarSQL 1 week, 5 days ago

abfs, dfs

upvoted 2 times

Momoanwar 4 days, 18 hours ago

Access Azure storage
Once you have properly configured credentials to access your Azure storage container, you can interact with resources in the storage account using URIs. Databricks recommends using the abfss driver for greater security.

Python

Copy
spark.read.load("abfss://<container-name>@<storage-account-name>.dfs.core.windows.net/<path-to-data>")
https://learn.microsoft.com/en-us/azure/databricks/connect/storage/azure-storage

upvoted 1 times

Question #51 *Topic 1*

You are analyzing customer purchases in a Fabric notebook by using PySpark.
You have the following DataFrames:
transactions: Contains five columns named transaction_id, customer_id, product_id, amount, and date and has 10 million rows, with each row representing a transaction. customers: Contains customer details in 1,000 rows and three columns named customer_id, name, and country.
You need to join the DataFrames on the customer_id column. The solution must minimize data shuffling.
You write the following code.
from pyspark.sql import functions as F
results =
Which code should you run to populate the results DataFrame?

A. transactions.join(F.broadcast(customers), transactions.customer_id == customers.customer_id)

B. transactions.join(customers, transactions.customer_id == customers.customer_id).distinct()

C. transactions.join(customers, transactions.customer_id == customers.customer_id)

D. transactions.crossJoin(customers).where(transactions.customer_id == customers.customer_id)

Correct Answer: *A*

Community vote distribution

A (100%)

SamuComqi 4 days, 5 hours ago

Selected Answer: A

A. transactions.join(F.broadcast(customers), transactions.customer_id == customers.customer_id)

Optimized method to perform a join between a very large table and a smaller one.

Source: https://sparkbyexamples.com/spark/broadcast-join-in-spark/"

upvoted 1 times

Momoanwar 4 days, 18 hours ago

Selected Answer: A

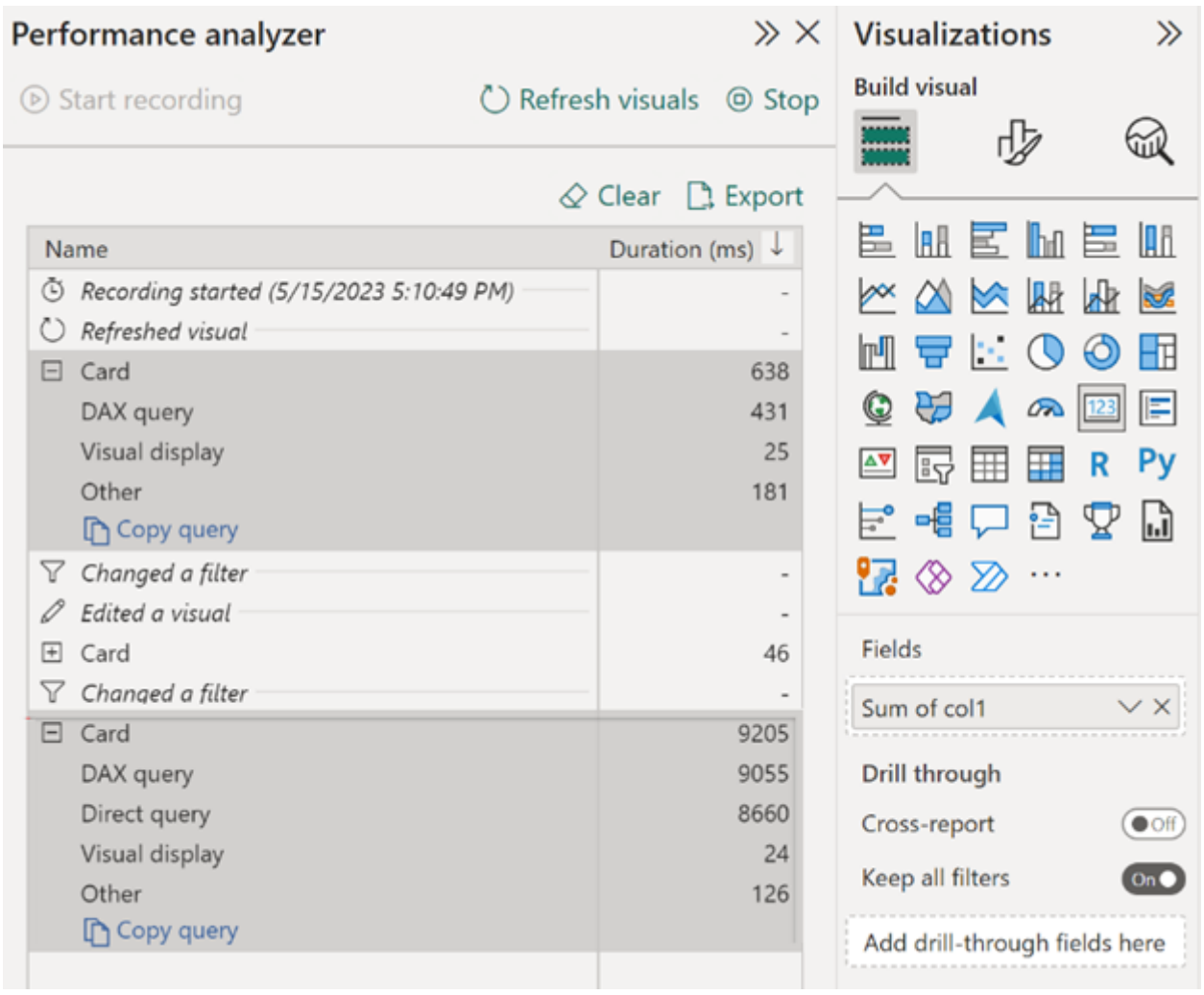
In Apache Spark, broadcasting refers to an optimization technique for join operations. When you join two DataFrames or RDDs and one of them is significantly smaller than the other, Spark can "broadcast" the smaller table to all nodes in the cluster. This approach avoids the need for network shuffles for each row of the larger table, significantly reducing the execution time of the join operation.

upvoted 1 times

HOTSPOT -

You have a Microsoft Power BI report and a semantic model that uses Direct Lake mode.

From Power BI Desktop, you open Performance analyzer as shown in the following exhibit.



Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.

NOTE: Each correct selection is worth one point.

Answer Area

The Direct Lake fallback behavior is set to [▼]
- Automatic
- DirectLakeOnly
- DirectQueryOnly

The query for the table visual is executed by using [▼]
- the composite model
- Direct Lake
- Direct Query

Answer Area

Correct Answer:

The Direct Lake fallback behavior is set to [▼]
- Automatic
- DirectLakeOnly
- DirectQueryOnly

The query for the table visual is executed by using [▼]
- the composite model
- Direct Lake
- Direct Query

Momoanwar 4 days, 17 hours ago

I think automatic and direct lake : we have fallback direct query for card and no informations about table visual so its direct lake.

upvoted 1 times

Question #53 *Topic 1*

HOTSPOT -

You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a table named Nyctaxi_raw. Nyctaxi_row contains the following table:

Name	Data type
pickupDateTime	Timestamp
passengerCount	Integer
fareAmount	Double
paymentType	String
tipAmount	Double

You create a Fabric notebook and attach it to Lakehouse1.

You need to use PySpark code to transform the data. The solution must meet the following requirements:

Add a column named pickupDate that will contain only the date portion of pickupDateTime.

Filter the DataFrame to include only rows where fareAmount is a positive number that is less than 100.

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

NOTE: Each correct selection is worth one point.

Answer Area

```
df = spark.read.format("delta").load("Tables/nyctaxi_raw")

df2= [ ▼ ] ("pickupDate", df['tpepPickupDateTime']
      df.columns
      df.select
      df.withColumn
      df.withColumnsRenamed

[ ▼ ] \
.alias('date')
.cast('date')
.cast('pickupDate')
.getfield('date')

      [ ▼ ]
      .filter("fareAmount > 0 AND farAmount < 100")
      .filter(col("fareAmount").contains("1..100"))
      .when(df.fareAmount > 0 AND fareAmount < 100)
      .where(df.fareAmount.isin([1,100]))
```

Answer Area

```
df = spark.read.format("delta").load("Tables/nyctaxi_raw")

df2= [ ▼ ] ("pickupDate", df['tpepPickupDateTime']
      df.columns
      df.select
      df.withColumn
      df.withColumnsRenamed
```

Correct Answer:

```
[ ▼ ] \
.alias('date')
.cast('date')
.cast('pickupDate')
.getfield('date')

      [ ▼ ]
      .filter("fareAmount > 0 AND farAmount < 100")
      .filter(col("fareAmount").contains("1..100"))
      .when(df.fareAmount > 0 AND fareAmount < 100)
      .where(df.fareAmount.isin([1,100]))
```

Momoanwar 4 days, 16 hours ago

We need to add column not rename existing column. Here is the correct answer :
df.withColumn('pickupDate', df['pickupDateTime'].cast(DateType())) \
.filter("fareAmount > 0 AND fareAmount < 100")

upvoted 1 times

Question #54 *Topic 1*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

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 have a Fabric tenant that contains a new semantic model in OneLake.

You use a Fabric notebook to read the data into a Spark DataFrame.

You need to evaluate the data to calculate the min, max, mean, and standard deviation values for all the string and numeric columns.

Solution: You use the following PySpark expression:

df.explain()

Does this meet the goal?

A. Yes

B. No

Correct Answer: *B*

Community vote distribution

B (100%)

SamuComqi 4 days, 4 hours ago

Selected Answer: B

The correct syntax is df.describe().

Sources:
* describe --> https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.describe.html
* explain --> https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.explain.html

upvoted 1 times

SamuComqi 4 days, 4 hours ago

Also df.summary() is a valid solution.

Source ---> https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.summary.html

upvoted 1 times

Momoanwar 4 days, 16 hours ago

Selected Answer: B

No explain is for the execut plan

upvoted 1 times

Question #55 *Topic 1*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
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 have a Fabric tenant that contains a new semantic model in OneLake.
You use a Fabric notebook to read the data into a Spark DataFrame.
You need to evaluate the data to calculate the min, max, mean, and standard deviation values for all the string and numeric columns.
Solution: You use the following PySpark expression:
df.show()
Does this meet the goal?

A. Yes

B. No

Correct Answer: *B*

Community vote distribution

B (100%)

SamuComqi 4 days, 4 hours ago

Selected Answer: B

The correct syntax is df.describe().

Sources:
* describe --> https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.describe.html
* show --> https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.show.html

upvoted 1 times

SamuComqi 4 days, 4 hours ago

Also df.summary() is a valid solution.

Source ---> https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.summary.html

upvoted 1 times

Momoanwar 4 days, 16 hours ago

Selected Answer: B

No show is to display data

upvoted 1 times

Question #56 *Topic 1*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
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 have a Fabric tenant that contains a new semantic model in OneLake.
You use a Fabric notebook to read the data into a Spark DataFrame.
You need to evaluate the data to calculate the min, max, mean, and standard deviation values for all the string and numeric columns.
Solution: You use the following PySpark expression:
df.summary()
Does this meet the goal?

A. Yes

B. No

Correct Answer: *A*

Community vote distribution

A (100%)

SamuComqi 4 days, 4 hours ago

Selected Answer: A

Also df.describe() is a valid solution.

Sources:
* summary --> https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.summary.html
* describe --> https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.describe.html

upvoted 1 times

Momoanwar 4 days, 16 hours ago

Selected Answer: A

Correct

upvoted 1 times

Question #57 *Topic 1*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
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 have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a Delta table named Customer.
When you query Customer, you discover that the query is slow to execute. You suspect that maintenance was NOT performed on the table.
You need to identify whether maintenance tasks were performed on Customer.
Solution: You run the following Spark SQL statement:

DESCRIBE HISTORY customer -
Does this meet the goal?

A. Yes

B. No

Correct Answer: *A*

Community vote distribution

A (100%)

Momoanwar 4 days, 16 hours ago

Selected Answer: A

Correct

upvoted 1 times

Question #58 *Topic 1*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
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 have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a Delta table named Customer.
When you query Customer, you discover that the query is slow to execute. You suspect that maintenance was NOT performed on the table.
You need to identify whether maintenance tasks were performed on Customer.
Solution: You run the following Spark SQL statement:

REFRESH TABLE customer -
Does this meet the goal?

A. Yes

B. No

Correct Answer: *B*

Community vote distribution

B (100%)

Momoanwar 4 days, 16 hours ago

Selected Answer: B

Correct

upvoted 1 times

Question #59 *Topic 1*

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
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 have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a Delta table named Customer.
When you query Customer, you discover that the query is slow to execute. You suspect that maintenance was NOT performed on the table.
You need to identify whether maintenance tasks were performed on Customer.
Solution: You run the following Spark SQL statement:

EXPLAIN TABLE customer -
Does this meet the goal?

A. Yes

B. No

Correct Answer: *B*

Community vote distribution

B (100%)

Momoanwar 4 days, 16 hours ago

Selected Answer: B

Given answer is correct. Explain is for query execution plan

upvoted 1 times