**Exam DP-600: Implementing Analytics Solutions Using Microsoft Fabric - Results**

Return to review

Chart

Pie chart with 4 slices.

End of interactive chart.

Attempt 1

All questions

Top of Form

Question 1: **Correct**

Your company uses a Fabric data warehouse containing customer data and purchase transactions. You need to enrich this data with additional information, such as customer location based on IP addresses and product popularity scores derived from social media sentiment analysis.

Which of the following approaches is the MOST efficient and scalable way to achieve this data enrichment?

A. Manually update existing customer records with location data from a separate IP geolocation API.

B. Develop Azure Data Factory pipelines to periodically extract IP addresses and product IDs, perform external API calls for location and sentiment analysis, and then update the data warehouse tables.

C. Implement materialized views in the data warehouse, joining the existing customer data with pre-computed location and sentiment tables refreshed through external APIs.

D. Use Azure Databricks notebooks to enrich the data on-demand, joining external data sources and updating the data warehouse tables in real-time.

* **A**
* **C**
* **B**

**(Correct)**

* **D**

**Explanation**

Answer: B

Explanation:

A: Manually updating data is inefficient and prone to errors.

B: Azure Data Factory pipelines offer a scalable and automated solution. You can define pipelines that extract relevant data, perform API calls, and update the data warehouse tables based on schedules or triggers. This ensures regular enrichment without manual intervention.

C: Materialized views can improve query performance for frequently accessed enriched data, but they add complexity and storage overhead. They're not suitable for real-time updates.

D: Databricks notebooks are powerful for complex transformations, but managing them separately from the data pipeline adds complexity. Azure Data Factory pipelines provide a more integrated approach for scheduled or triggered data enrichment.

Bottom of Form

Top of Form

Question 2: **Correct**

You've designed a Fabric environment for a large retail company with various departments (e.g., marketing, finance, operations). Each department requires access to specific data sets and reports, while maintaining overall data governance.

Which combination of workspace and item-level access controls would best achieve this objective?

A. Create individual workspaces for each department with read-only access to shared reports.

B. Implement workspace roles with granular permissions for specific data sets within a single workspace.

C. Utilize item-level security based on data sensitivity labels to control access across different workspaces.

D. Combine workspace roles and item-level security for fine-grained control and data governance.

* **A**
* **C**
* **B**
* **D**

**(Correct)**

**Explanation**

The correct answer is D. Combine workspace roles and item-level security for fine-grained control and data governance. Here's why:

Individual workspaces: While providing isolation, this approach limits collaboration and increases maintenance overhead.

Workspace roles with granular permissions: Enables department-specific access within a single workspace, but lacks detailed data control.

Item-level security: Offers precise control over individual data items, but requires manual labeling and might be insufficient for broader access control.

Combination of roles and item-level security: Provides the most comprehensive approach. Workspace roles grant department-level access, while item-level security restricts access to specific data within those roles, ensuring data privacy and compliance.

Explanation of wrong answers:

A. Read-only access limits data exploration and analysis capabilities for departments.

B. Granular permissions within a single workspace might not address inter-departmental data sharing needs.

C. Item-level security alone doesn't manage overall workspace access and collaboration.

Bottom of Form

Top of Form

Question 3: **Incorrect**

You're building a Power BI report for a marketing agency analyzing campaign performance across different regions and product categories. You want to create a calculation group that dynamically displays the top-performing region based on a selected product category.

Which DAX calculation group best achieves this?

A.

**Code snippet**

1. TopRegion =
2. VAR SelectedCategory = SELECTEDVALUE('Products'[Category])
3. RETURN
4. CALCULATE(
5. MAX('Sales'[SalesAmount]),
6. FILTER('Sales', 'Sales'[ProductCategory] = SelectedCategory)
7. )

B.

**Code snippet**

1. TopRegion =
2. VAR SelectedCategory = SELECTEDVALUE('Products'[Category])
3. RETURN
4. CALCULATE(
5. MAX('Sales'[SalesAmount]),
6. FILTER('Sales', 'Sales'[ProductCategory] = SelectedCategory)
7. )

C.

**Code snippet**

1. TopRegion =
2. VAR SelectedCategory = SELECTEDVALUE('Products'[Category])
3. RETURN
4. CALCULATE(
5. MAX('Sales'[Region]),
6. FILTER('Sales', 'Sales'[ProductCategory] = SelectedCategory)
7. )

D.

**Code snippet**

1. TopRegion =
2. VAR SelectedCategory = SELECTEDVALUE('Products'[Category])
3. RETURN
4. CALCULATE(
5. MAX('Sales'[Region]),
6. FILTER('Sales', 'Sales'[ProductCategory] = CONCATENATE("Category:", SelectedCategory))
7. )

* **C**

**(Incorrect)**

* **D**
* **A**

**(Correct)**

* **B**

**Explanation**

The correct answer is A. VAR SelectedCategory = ...

Here's why:

A correctly uses a calculation group with variables and dynamic filtering based on the selected product category:

SelectedCategory variable stores the user-selected category.

The CALCULATE function filters the Sales table based on the selected category and then finds the region with the maximum sales amount.

B is missing the RETURN statement within the calculation group.

C finds the region with the highest sales amount for the selected category, but it doesn't necessarily correspond to the top-performing region in terms of sales.

D uses incorrect syntax for dynamically filtering based on the selected category. CONCATENATE is not necessary within the filter expression.

Explanation of wrong answers:

B: It's missing the RETURN statement within the calculation group, making it incomplete.

C: It finds the region with the highest sales amount, not necessarily the top-performing region within the chosen category.

D: It uses incorrect syntax for dynamically filtering based on the selected category. CONCATENATE is not needed in this scenario.

Bottom of Form

Top of Form

Question 4: **Incorrect**

Your team has created a Fabric data lake containing customer data from different sources. To enable secure access for various departments, you need to create separate shortcuts for each department, limiting them to specific customer subsets based on their needs.

How can you achieve this granular access control using Fabric shortcuts?

A. Create individual shortcuts with different names and point them to the same data lake folder, relying on user permissions for access control.

B. Implement Azure Data Factory to filter data based on department needs and create separate data sets with shortcuts pointing to them.

C. Utilize Azure Active Directory groups and assign different shortcut access permissions to each group based on department requirements.

D. Develop custom scripts to dynamically generate shortcuts based on user attributes and associated data subsets.

* **D**
* **C**

**(Correct)**

* **A**
* **B**

**(Incorrect)**

**Explanation**

Answer: C

Explanation:

A: Using the same folder with different shortcuts can lead to confusion and potential unauthorized access if permissions aren't tightly controlled.

B: Azure Data Factory adds unnecessary complexity for simple access control within the data lake.

C: Leveraging Azure Active Directory groups and associating them with specific shortcut permissions offers a secure and manageable solution. This allows granular control over data access based on department needs and ensures users only see the relevant data through their assigned shortcuts.

D: While custom scripts offer flexibility, they require additional development and maintenance effort. Azure AD groups provide a built-in and scalable solution for this scenario.

Bottom of Form

Top of Form

Question 5: **Incorrect**

You are a data analyst at MContoso, a retail chain, tasked with analyzing customer purchase data to personalize marketing campaigns. You have access to customer profiles including demographics, purchase history, and website browsing behavior.

Which of the following Fabric tools would be MOST effective in enriching customer profiles for targeted marketing campaigns?

A. Azure Data Factory

B. Azure Machine Learning

C. Azure Databricks

D. Azure Data Lake Storage

* **D**
* **C**
* **A**

**(Correct)**

* **B**

**(Incorrect)**

**Explanation**

A. Azure Data Factory

Explanation:

Azure Machine Learning excels at building and deploying machine learning models, not directly enriching profiles.

Azure Databricks offers advanced data processing but lacks built-in features for profile enrichment.

Azure Data Lake Storage provides data storage but lacks functionalities for data transformation and orchestration.

Azure Data Factory specializes in data pipelines and transformations, enabling you to integrate various data sources (purchase history, website behavior) into enriched customer profiles, making it the ideal tool for this scenario.

Bottom of Form

Top of Form

Question 6: **Incorrect**

Your company needs to implement access control for specific data within a Fabric data warehouse based on user roles. The warehouse contains sensitive financial data for different departments.

Which of the following options is the MOST secure and efficient way to achieve this data access control?

A. Create separate views for each department, filtering out sensitive data not relevant to their roles.

B. Implement row-level security within the warehouse tables, restricting access to specific rows based on user roles.

C. Develop custom database triggers to intercept queries and filter data based on user permissions.

D. Grant different levels of access permissions to the warehouse tables directly for each department.

* **D**
* **C**
* **B**

**(Correct)**

* **A**

**(Incorrect)**

**Explanation**

Answer: B

Explanation:

A: While views can simplify data access, they don't address the security aspect of controlling access to sensitive data.

B: Row-level security offers the most granular and secure approach. By defining policies based on user roles and columns, you can restrict access to specific data rows, ensuring only authorized users see sensitive information relevant to their department.

C: Custom triggers add complexity and might not be scalable for managing access control across a large user base.

D: Granting direct table permissions is insecure and lacks granular control. Users might access confidential data beyond their authorized scope.

Bottom of Form

Top of Form

Question 7: **Incorrect**

Your company performs monthly data analysis reports requiring complex transformations on large datasets in a Fabric data lakehouse. You need to schedule multiple dataflows for this process, but some dataflows depend on the successful completion of others.

How can you schedule the dataflows efficiently and ensure proper execution order while minimizing unnecessary runs?

A. Schedule each dataflow individually with manual dependencies, specifying which dataflows must finish before the next one starts.

B. Create an ADF pipeline with chained copy activities to move data through different dataflows based on their dependencies.

C. Utilize ADF triggers on completion of one dataflow to trigger the next dataflow in the sequence, ensuring proper execution order.

D. Develop a single dataflow with conditional branches based on data availability, dynamically executing the necessary transformations depending on the previous steps' completion.

* **B**
* **C**

**(Correct)**

* **A**
* **D**

**(Incorrect)**

**Explanation**

Answer: C

Explanation:

A: Manual dependencies are tedious to manage and error-prone for multiple dataflows.

B: ADF copy activities are not designed for complex transformations within dataflows.

C: ADF triggers on dataflow completion offer an efficient and automated approach. You can configure each dataflow to trigger the next one upon successful execution, ensuring proper order and avoiding unnecessary runs if a previous dataflow fails.

D: While conditional branches within a single dataflow offer flexibility, they might be complex for multiple dependent dataflows. Utilizing triggers on completion provides a clearer and more reliable approach for managing execution order.

Bottom of Form

Top of Form

Question 8: **Correct**

You are building a semantic model for Azure Analysis Services that combines sales and customer data stored in different sources: Azure SQL Database and Azure Data Lake Storage. The sales data updates hourly, while customer data changes less frequently.

Which storage mode would be the most appropriate for this scenario, considering data refresh efficiency and query performance?

A. Import mode: Load all data into the model for fast queries, refreshing the entire model hourly for sales updates.

B. DirectQuery mode: Connect to both data sources directly for real-time updates, potentially impacting query performance.

C. Composite model: Combine import and DirectQuery modes, keeping frequently updated sales data in the model and connecting to the lake for static customer data.

D. Processing mode: Use a scheduled processing job to refresh the entire model with the latest data hourly.

* **D**
* **B**
* **A**
* **C**

**(Correct)**

**Explanation**

C. Composite model

Explanation:

Option A wastes resources importing static customer data and unnecessarily refreshes the entire model every hour.

Option B offers real-time updates, but might not be performant for large datasets or complex queries.

Option D overloads the processing schedule with hourly refreshes for both sales and customer data.

Option C optimizes data refresh efficiency and query performance. By keeping the fast-changing sales data in the model for quick access and connecting to the lake for static customer data, you achieve efficient hourly updates for sales data without impacting performance for queries involving customer data.

Bottom of Form

Top of Form

Question 9: **Incorrect**

You're building a Power BI report for a university to analyze student enrollment across different courses and semesters. The data model includes the following tables:

Students (student\_id, name, major)

Courses (course\_id, name, department)

Semesters (semester\_id, start\_date, end\_date)

Enrollments (enrollment\_id, student\_id, course\_id, semester\_id, grade)

Which type of relationship should you establish between the Enrollments table and the other three tables to accurately represent student enrollment in courses across semesters?

A. One-to-one relationships between Enrollments and each of Students, Courses, and Semesters.

B. Many-to-one relationships between Enrollments and Students, Courses, and Semesters.

C. Many-to-many relationship between Enrollments and Students and many-to-one with Courses and Semesters.

D. Many-to-many relationship between Enrollments and Students and Courses and many-to-one with Semesters.

* **C**

**(Incorrect)**

* **A**
* **B**
* **D**

**(Correct)**

**Explanation**

The correct answer is D. Many-to-many relationship between Enrollments and Students and Courses and many-to-one with Semesters.

Here's why:

A and B: These options represent one-to-one or many-to-one relationships, which wouldn't accurately capture the scenario where a student can enroll in multiple courses across different semesters.

C: This option has a correct relationship between Enrollments and Students but incorrectly defines the relationship with Courses as many-to-one. One student can enroll in multiple courses, hence a many-to-many relationship is needed.

D accurately reflects the scenario. A student can enroll in many courses across different semesters, and a course can have many students enrolled in it. The relationship with Semesters is correctly defined as many-to-one because each enrollment belongs to one semester.

Explanation of wrong answers:

A and B: These options don't account for the possibility of a student taking multiple courses across different semesters.

C: This option incorrectly defines the relationship between Enrollments and Courses as many-to-one.

Bottom of Form

Top of Form

Question 10: **Incorrect**

Your company needs to share labeled data from the Fabric lakehouse with a partner organization for a joint analytics project. However, the partner's security policies differ from yours.

Which approach would best enable secure and controlled data sharing while respecting both sets of security policies?

A. Grant the partner direct access to the labeled data within the Fabric environment.

B. Export the data to a secure cloud storage platform and share it with the partner, trusting their security measures.

C. Utilize Fabric's external data sharing feature with data masking and dynamic labeling based on partner policies.

D. Develop a custom application to manage data access and apply partner-specific security policies.

* **D**

**(Incorrect)**

* **B**
* **A**
* **C**

**(Correct)**

**Explanation**

The correct answer is C. Utilize Fabric's external data sharing feature with data masking and dynamic labeling based on partner policies. Here's why:

Secure sharing: Fabric's external data sharing provides controlled access with granular permissions.

Data masking: Sensitive data can be masked or redacted to comply with partner policies and minimize exposure.

Dynamic labeling: Fabric can automatically apply partner-specific labels to shared data, ensuring consistent security measures.

Explanation of wrong answers:

A. Direct access poses security risks and might violate partner policies.

B. Sharing unmasked data to a third-party platform creates unnecessary security vulnerabilities.

D. Developing a custom application is unnecessary and time-consuming for a secure data sharing scenario.

Bottom of Form

Top of Form

Question 11: **Incorrect**

Your company wants to share a large marketing data set from their Fabric workspace with an external partner agency for a specific campaign. They need to ensure secure access and data expiration after the campaign.

Which method would best facilitate this temporary data sharing?

A. Grant the partner agency direct access to the data within the Fabric workspace.

B. Export the data set to a cloud storage platform and share the link with the agency.

C. Utilize Fabric's external data sharing feature with expiration settings and access controls.

D. Develop a custom application to manage and control data sharing with the agency.

* **A**

**(Incorrect)**

* **B**
* **D**
* **C**

**(Correct)**

**Explanation**

The correct answer is C. Utilize Fabric's external data sharing feature with expiration settings and access controls. Here's why:

Secure and controlled: Fabric's built-in feature offers secure access with granular permissions and expiration settings, minimizing risk and complying with data regulations.

Simplified management: This option avoids the need for manual data transfer or custom application development, reducing overhead and complexity.

Expiration settings: Setting an expiration date ensures the data access automatically terminates after the campaign, enhancing data security and compliance.

Explanation of wrong answers:

A. Direct workspace access offers poor control and can lead to unauthorized data access beyond the campaign.

B. Sharing a cloud storage link might not provide adequate access control or security features.

D. Developing a custom application is unnecessary and time-consuming for a temporary data sharing scenario.

Bottom of Form

Top of Form

Question 12: **Correct**

You're tasked with building a semantic model for a retail company that tracks sales, products, customers, and locations. The data includes numerous tables with varying granularities. You want to optimize query performance and user experience for complex analytical queries across different dimensions.

Which star schema design would be MOST suitable for this scenario?

A. Single central fact table with all dimension columns denormalized.

B. Multiple fact tables, each linked to specific dimension tables.

C. Single central fact table with normalized dimension columns.

D. Snowflake schema with multiple fact and dimension tables connected through intermediate dimension tables.

* **C**

**(Correct)**

* **A**
* **B**
* **D**

**Explanation**

The correct answer is C. Single central fact table with normalized dimension columns.

Here's why:

A: Denormalizing all dimensions might improve performance for simple queries, but it leads to data redundancy and complexity for managing updates and joins in complex analytical scenarios.

B: Multiple fact tables can be suitable for specific use cases, but for complex queries across all dimensions, a single central fact table simplifies joins and improves overall query performance.

D: Snowflake schemas are more complex and suitable for specialized scenarios with multiple hierarchies or relationships between dimensions. In this case, a single central fact table with normalized dimensions offers better performance and manageability for general analytical queries.

C best fits the scenario because:

A single central fact table centralizes all sales data for efficient joins and aggregations.

Normalized dimension columns avoid data redundancy and improve data integrity.

Explanation of wrong answers:

A: Denormalization can lead to performance issues for complex queries and maintenance overhead.

B: Multiple fact tables can complicate joins and hinder performance for cross-dimensional queries.

D: Snowflake schemas are best suited for specific scenarios with complex dimensional hierarchies, not general analytical queries across multiple dimensions.

Bottom of Form

Top of Form

Question 13: **Incorrect**

You are working on a project where you need to forecast sales for the next quarter and recommend actions to increase them. You decide to use predictive and prescriptive analytics and integrate the findings into a visual.

Which of the following techniques would be the most appropriate to use in this scenario?

A. Decision tree analysis B. Sentiment analysis C. Cohort analysis D. Time-series analysis

* **B**

**(Incorrect)**

* **C**
* **D**

**(Correct)**

* **A**

**Explanation**

D. Time-series analysis

**Explanation:** Predictive analytics is used to make predictions about future outcomes, and prescriptive analytics is used to recommend actions to take for optimal outcomes. In this scenario, time-series analysis would be the most appropriate technique as it allows you to analyze data points collected over time, which is ideal for forecasting sales. The other options, while useful in other scenarios, are not the best fit for this specific task.

Bottom of Form

Top of Form

Question 14: **Incorrect**

You want to create a KPI card displaying the percentage of orders completed within a specified timeframe (e.g., 30 days) based on a user-defined parameter for the timeframe. You also want to compare this percentage to an average completion time calculated across all orders.

Which approach best utilizes calculation groups, dynamic strings, and field parameters?

A. Create a calculation group with:

\* PercentCompleted: CALCULATE(SUM(IF(Orders[CompletionDate] <= Orders[OrderDate] + [Timeframe Parameter], 1, 0)) / COUNTROWS(Orders), ALL())

\* AvgCompletionTime: AVERAGE(Orders[CompletionDate] - Orders[OrderDate])

\* ComparisonString: DAX CONCATENATE(FORMAT([PercentCompleted], "Percent"), " completed vs. ", FORMAT([AvgCompletionTime], "Days"))

B. Create a single dynamic measure with CALCULATE, IF, and FORMAT functions to determine both percentage and average time based on the parameter, then format and display the information directly in the KPI card.

C. Use a field parameter for Timeframe and create separate measures for PercentCompleted and AvgCompletionTime with CALCULATE and FILTER functions based on the parameter value. Display these measures individually in the KPI card with dynamic titles using DAX CONCATENATE.

D. Define a measure for PercentCompleted with CALCULATE and IF functions based on the parameter. Use conditional formatting in the KPI card to visually compare the percentage to a pre-defined average completion time threshold.

* **D**

**(Incorrect)**

* **C**
* **A**

**(Correct)**

* **B**

**Explanation**

The correct answer is A. Create a calculation group with...

Here's why:

A leverages calculation groups for modularity and clarity:

PercentCompleted calculates the percentage based on the parameter using CALCULATE and IF statements.

AvgCompletionTime calculates the average completion time across all orders.

ComparisonString dynamically formats and combines both values using DAX CONCATENATE.

B while possible, a single dynamic measure can be complex and less modular than separate calculations.

C separate measures with field parameters work, but displaying them and adding titles requires additional configuration in the KPI card.

Bottom of Form

Top of Form

Question 15: **Incorrect**

Your company has a Fabric environment with multiple workspaces for different departments (e.g., marketing, finance, operations). They need to share specific data sets and reports across these workspaces while maintaining data security and ownership.

Which combination of data sharing features would best achieve this objective?

A. Enable public workspace access with view-only permissions for shared data.

B. Implement data connectors to directly link data between workspaces.

C. Utilize workspace sharing with granular access control based on user roles.

D. Create shared data lakes and lakehouses across workspaces with controlled data access.

* **A**

**(Incorrect)**

* **D**

**(Correct)**

* **C**
* **B**

**Explanation**

The correct answer is D. Create shared data lakes and lakehouses across workspaces with controlled data access. Here's why:

Shared data lakes and lakehouses: Provide a central location for storing and accessing data across workspaces, facilitating collaboration and analysis.

Controlled data access: Enables granular permissions based on user roles and data sensitivity, ensuring security and ownership.

Flexibility: This approach allows for sharing raw data, processed data, and pre-built reports, catering to different needs.

Explanation of wrong answers:

A. Public workspace access compromises data security and doesn't offer granular control.

B. Data connectors might not be suitable for sharing complete data sets and lack access control features.

C. While workspace sharing allows controlled access within a single workspace, it doesn't address cross-workspace data sharing needs.

Bottom of Form

Top of Form

Question 16: **Incorrect**

Your company's data warehouse includes a product dimension table with a Type 1 SCD for product name. You need to handle situations where products are discontinued and no longer included in sales data.

Which approach is MOST efficient for maintaining historical information while avoiding misleading analysis?

A. Update the product name in the dimension table to "Discontinued" for inactive products.

B. Remove inactive products from the dimension table altogether.

C. Add a new flag column "Active" to the dimension table, marking inactive products.

D. Create a separate "historical product" dimension table for discontinued products.

* **D**

**(Incorrect)**

* **C**

**(Correct)**

* **A**
* **B**

**Explanation**

Answer: C

Explanation:

A: Updating the name to "Discontinued" masks historical context and can influence calculations.

B: Removing inactive products eliminates historical data for analysis.

C: Adding an "Active" flag preserves historical data while clearly identifying which products are currently relevant for analysis. This avoids misleading interpretations.

D: Creating a separate historical table adds complexity and might not be necessary for simple discontinued product handling.

Bottom of Form

Top of Form

Question 17: **Incorrect**

You want to calculate the average sales per customer for the past year, excluding any customer with less than 5 orders. You need to use windowing functions and information functions.

Which DAX calculation achieves this?

A.

**Code snippet**

1. AvgSalesPerCustomer = CALCULATE(AVERAGE(Sales[SalesAmount]),
2. DATESYTD(),
3. FILTER(Sales, Sales[CustomerID] = CALCULATE(COUNTROWS(Sales), ALLEXCEPT(Sales, Sales[CustomerID]))))

B.

**Code snippet**

1. AvgSalesPerCustomer = CALCULATE(
2. AVERAGE(Sales[SalesAmount]),
3. DATESYTD(),
4. Sales[OrderID] - 5
5. )

C.

**Code snippet**

1. AvgSalesPerCustomer = CALCULATE(
2. AVERAGE(Sales[SalesAmount]),
3. DATESYTD(),
4. FILTER(Sales, Sales[CustomerID] IN (CALCULATE(VAR OrderCount = COUNTROWS(Sales), ALLEXCEPT(Sales, Sales[CustomerID])), OrderCount >= 5)))

D.

**Code snippet**

1. AvgSalesPerCustomer = CALCULATE(
2. AVERAGE(Sales[SalesAmount]),
3. DATESYTD() - 12,
4. DATESYTD()

* **B**

**(Incorrect)**

* **C**

**(Correct)**

* **A**
* **D**

**Explanation**

The correct answer is C.

AvgSalesPerCustomer = CALCULATE(

    AVERAGE(Sales[SalesAmount]),

    DATESYTD(),

    FILTER(Sales, Sales[CustomerID] IN (CALCULATE(VAR OrderCount = COUNTROWS(Sales), ALLEXCEPT(Sales, Sales[CustomerID])), OrderCount >= 5)))

Explanation: This DAX calculation correctly calculates the average sales per customer for the past year, excluding any customer with less than 5 orders. It uses the CALCULATE function to modify the context in which the AVERAGE function is evaluated. The FILTER function is used to include only those customers in the calculation who have 5 or more orders. The COUNTROWS function counts the number of orders per customer, and ALLEXCEPT function is used to remove all context filters in the table except filters that have been applied to the specified columns (in this case, Sales[CustomerID]).

Options A, B, and D are incorrect because they do not correctly implement the requirement of excluding customers with less than 5 orders. Option A incorrectly equates the CustomerID to the count of rows in Sales, option B subtracts 5 from the OrderID which doesn’t make sense in this context, and option D incorrectly subtracts 12 from the DATESYTD function

Bottom of Form

Top of Form

Question 18: **Incorrect**

Your company is migrating existing data analytics workloads from on-premises servers to Microsoft Fabric. They prioritize cost optimization and require minimal downtime during migration.

Which Fabric-enabled workspace settings would best achieve these objectives?

A. Enable Azure Data Lake Storage for raw data and configure Standard SKU for data warehouses.

B. Utilize Azure Synapse Analytics as the data lakehouse and set high concurrency for workspaces.

C. Implement Azure Data Factory for data pipelines and configure Premium SKU for lakehouses.

D. Configure workspace sharing with granular access control and enable cost optimization features.

* **C**

**(Incorrect)**

* **A**
* **D**

**(Correct)**

* **B**

**Explanation**

The correct answer is D. Configure workspace sharing with granular access control and enable cost optimization features. Here's why:

Workspace sharing: Allows for efficient resource utilization by sharing resources across multiple workloads, reducing costs.

Granular access control: Ensures data security and compliance while minimizing resource access.

Cost optimization features: Fabric offers automatic scaling and cost management tools to optimize resource allocation and reduce unnecessary spending.

Explanation of wrong answers:

A. Standard SKU for data warehouses balances cost and performance, but might not be necessary for initial migration. High concurrency is not crucial for cost optimization.

B. Azure Synapse Analytics offers powerful features but can be expensive for large-scale migrations. High concurrency also increases cost.

C. Azure Data Factory is useful for data pipelines, but Premium SKU for lakehouses is unnecessary for initial migration and might be cost-prohibitive.

Bottom of Form

Top of Form

Question 19: **Incorrect**

Your company uses a Fabric data warehouse containing customer transaction data. Your team analyzes customer purchasing behavior across various product categories. Currently, reports require complex calculations of average order value and purchase frequency for each category.

How can you optimize query performance and code reusability for these reports using views, functions, or stored procedures?

A. Create individual views for each product category containing the calculated average order value and purchase frequency.

B. Develop a stored procedure that accepts the product category as input and returns the calculated values.

C. Implement user-defined functions (UDFs) within the data warehouse to encapsulate the calculations for average order value and purchase frequency.

D. Use materialized views for each product category, pre-computing and storing the calculated values for faster retrieval.

* **A**

**(Incorrect)**

* **D**
* **C**

**(Correct)**

* **B**

**Explanation**

Answer: C

Explanation:

A: Creating individual views for each category increases storage overhead and redundant calculations. It's not scalable for many categories.

B: Stored procedures are suitable for complex logic but might not be as efficient for repetitive calculations within reports.

C: User-defined functions encapsulate the calculation logic in a reusable way, allowing them to be called within any query for any product category. This promotes code reusability and optimizes performance by avoiding redundant calculations in each report.

D: While materialized views can improve performance for frequently accessed data, they require maintenance and additional storage cost. UDFs provide a more flexible and lightweight solution for this scenario.

Bottom of Form

Top of Form

Question 20: **Correct**

Your company plans to implement Type 2 SCDs for the "product category" dimension in the Fabric data warehouse. This dimension might have new categories added or existing categories renamed.

Which Type 2 SCD approach is MOST suitable for this scenario, and how should you handle changes in product category assignments?

A. Implement a historical record approach with a new record for each change and maintain historical category assignments.

B. Use an additive approach with a flag column to indicate active categories and maintain separate tables for current and historical categories.

C. Utilize a hybrid approach with both historical records and flags to track category changes and maintain historical assignments.

D. Implement a dimension attribute column to store the current category and update it as products change categories, leaving historical records untouched.

* **C**

**(Correct)**

* **D**
* **B**
* **A**

**Explanation**

Answer: C

Explanation:

C: A hybrid approach offers the best balance. Historical records capture category changes and maintain historical assignments for analysis, while flags indicate the current active category for efficient queries.

A: While historical records offer full history, they can increase table size and complexity. Flags provide a way to identify current categories without bloating the table.

B: Separate tables for current and historical categories can be cumbersome to manage and query. A hybrid approach within one table is more efficient.

D: Updating the dimension attribute column directly might lead to data loss for historical analysis. Maintaining historical records with flags is a more accurate and comprehensive approach for tracking category changes.

Bottom of Form

Top of Form

Question 21: **Incorrect**

You are a data analyst for Fabrikam, a manufacturing company, tasked with investigating production line efficiency. You have access to sensor data capturing machine performance metrics and operator activity logs.

Which Azure service(s) would be most effective in profiling operator activity and its impact on machine performance for exploratory analysis?

A. Azure Data Factory + Azure Machine Learning Studio

B. Azure Databricks + Azure Synapse Analytics

C. Azure Data Explorer + Power BI

D. Azure Application Insights + Azure Monitor

* **C**

**(Incorrect)**

* **D**
* **A**
* **B**

**(Correct)**

**Explanation**

B. Azure Databricks + Azure Synapse Analytics

Explanation:

Option A focuses on data pipeline orchestration and machine learning model building, not directly addressing profile analysis.

Option C excels in real-time log analysis and visualization but lacks advanced data wrangling and feature engineering capabilities.

Option D is suitable for monitoring application performance and infrastructure health but not ideal for in-depth profile analysis involving sensor data and operator logs.

Option B combines the strengths of both services - Azure Databricks provides powerful data manipulation and feature engineering for sensor and log data, while Azure Synapse Analytics offers scalable data warehousing and exploratory analysis tools, making it the most comprehensive solution for profiling operator activity and its impact on machine performance.

Bottom of Form

Top of Form

Question 22: **Correct**

You've implemented a star schema for your retail data model. The fact table includes foreign key relationships with each dimension table. You want to optimize query performance further for frequently used metrics like average sales per product and total sales per customer.

Which DAX Studio/Tabular Editor 2 feature would be MOST effective in achieving this optimization?

A. Define calculated columns for the desired metrics within the fact table.

B. Implement column partitions on the fact table based on relevant dimensions.

C. Create aggregations (measures) for the metrics within the Power BI Desktop interface.

D. Configure relationship filters on the dimension tables to restrict data based on user selections.

* **B**

**(Correct)**

* **A**
* **C**
* **D**

**Explanation**

The correct answer is B. Implement column partitions on the fact table based on relevant dimensions.

Here's why:

A: While calculated columns can be helpful, they add processing overhead and might not significantly improve performance for frequently used metrics.

C: Creating measures in Power BI Desktop is a standard practice, but it doesn't directly address query performance optimization within the semantic model.

D: Relationship filters can be useful for filtering data, but they don't directly improve query performance on large fact tables.

B is the most effective option because column partitions allow you to pre-aggregate data based on frequently used dimensions like product or customer. This reduces the amount of data processed during query execution, significantly improving performance for specific metrics and queries.

Explanation of wrong answers:

A: Calculated columns can add processing overhead and might not be the best solution for frequently used metrics.

C: Creating measures is standard practice but doesn't directly address performance optimization.

D: Relationship filters are helpful for data filtering but not directly related to performance on large fact tables.

Bottom of Form

Top of Form

Question 23: **Incorrect**

Your company needs to cleanse customer data in a Fabric data warehouse table containing duplicate records with inconsistent names and addresses. You want to identify and remove duplicates while ensuring consistency in name and address formats.

Which data cleansing techniques should you prioritize for this scenario?

A. Implement deduplication based on a unique identifier like customer ID and standardize names using string functions.

B. Utilize fuzzy matching algorithms to identify similar names and addresses, then merge them after manual review.

C. Leverage geocoding services to validate and standardize address formats within the data warehouse table.

D. Develop custom SQL stored procedures to perform complex data cleansing logic and address inconsistencies in both names and addresses.

* **A**

**(Correct)**

* **C**

**(Correct)**

* **B**

**(Incorrect)**

* **D**

**Explanation**

Answer: A & C

Explanation:

A: Deduplication based on a unique identifier and address standardization are essential for eliminating duplicate records and improving data quality. This ensures consistent representation of customer information for analysis.

B: Fuzzy matching can be helpful for identifying potential duplicates, but it requires manual review and might not be necessary if a unique identifier exists.

C: Geocoding services can significantly improve address consistency and accuracy, making location data more usable for analysis. This complements data standardization for address formats.

D: While custom SQL stored procedures offer flexibility, they can be complex to develop and maintain. Utilizing built-in deduplication and geocoding services provides a more efficient and scalable solution.

Bottom of Form

Top of Form

Question 24: **Incorrect**

You are a data analyst for a healthcare company that uses Microsoft Fabric. The company wants to predict patient readmission rates and prescribe actions to reduce them. You decide to use predictive and prescriptive analytics and present the findings in a report.

Which of the following techniques would be the most appropriate to use in this scenario?

A. Regression analysis B. Time-series analysis C. Cluster analysis D. Association rule learning

* **C**

**(Incorrect)**

* **A**

**(Correct)**

* **D**
* **B**

**Explanation**

A. Regression analysis

Explanation: Predictive analytics is used to make predictions about future outcomes, and prescriptive analytics is used to recommend actions to take for optimal outcomes. In this scenario, regression analysis would be the most appropriate technique as it allows you to predict a dependent variable based on the values of one or more independent variables, which is ideal for predicting patient readmission rates.

Bottom of Form

Top of Form

Question 25: **Incorrect**

Your company's existing data pipeline in ADF uses dataflows for basic transformations but needs to implement a machine learning model for advanced analysis. The model requires specific data pre-processing steps before being applied.

How can you integrate the machine learning model into your ADF pipeline while ensuring efficient data pre-processing and model execution?

A. Directly embed the machine learning code within the ADF dataflow using Python scripting.

B. Develop a separate Azure Machine Learning pipeline and trigger it from the ADF pipeline after pre-processing the data in a Dataflow.

C. Utilize Azure Synapse Analytics to integrate the machine learning model within the data warehouse, allowing pre-processing and execution within the same environment.

D. Deploy the machine learning model as an Azure Function and call it from the ADF pipeline after passing the pre-processed data.

* **C**

**(Incorrect)**

* **B**

**(Correct)**

* **A**
* **D**

**Explanation**

Answer: B

Explanation:

A: Embedding complex machine learning code within the dataflow is not recommended due to maintainability and scalability concerns.

B: A separate Azure Machine Learning pipeline provides a dedicated environment for model training and execution. Triggering it from the ADF pipeline after pre-processing ensures efficient data flow and separation of concerns.

C: While Synapse Analytics can integrate machine learning models, it might not be the best option for this scenario if the existing pipeline is already in ADF. A separate Azure Machine Learning pipeline offers greater flexibility and control.

D: Calling an Azure Function might add unnecessary complexity and latency. Utilizing a dedicated Azure Machine Learning pipeline is more suitable for this specific use case.

Bottom of Form

Top of Form

Question 26: **Incorrect**

Your company is building a large-scale data analytics environment in Fabric for a global manufacturing organization with high data volume and complex analysis requirements. They need to ensure performance, scalability, and data governance.

Which Fabric-enabled workspace settings would best address these needs?

A. Utilize Azure SQL Database for the data warehouse and set low concurrency for workspaces.

B. Implement Azure Data Lake Storage for raw data and enable workspace sharing with open access.

C. Configure Azure Synapse Analytics as the data lakehouse and set high concurrency for workspaces.

D. Implement Azure data governance solutions and enable automatic scaling for resources.

* **A**

**(Incorrect)**

* **B**
* **D**

**(Correct)**

* **C**

**Explanation**

The correct answer is D. Implement Azure data governance solutions and enable automatic scaling for resources. Here's why:

Azure data governance solutions: Provide data lineage tracking, access control, and auditing, ensuring data security and compliance.

Automatic scaling: Allocates resources based on workload demands, optimizing performance and costs for large datasets and complex analysis.

Explanation of wrong answers:

A. Azure SQL Database is not scalable for high data volume and complex analysis. Low concurrency hinders performance.

B. Open access in a shared workspace compromises data security and governance.

C. While Azure Synapse Analytics is powerful, high concurrency might not be necessary initially. Implementing data governance solutions is crucial.

Additional Tips:

Consider using Fabric's built-in monitoring and alerting tools to track performance and resource utilization.

Regularly review and update workspace settings to adapt to changing data volumes and analysis needs.

Leverage Fabric's data catalog and search features to improve data discoverability and collaboration.

Bottom of Form

Top of Form

Question 27: **Incorrect**

Your company has a Fabric data lakehouse containing website clickstream data and product information tables. You need to create a new table with derived insights, such as user sessions, popular product combinations, and abandoned cart rates.

What is the BEST approach to implement this data transformation and create the new table?

A. Develop Azure Functions triggered by website clickstream events to calculate insights and store them in the new table in real-time.

B. Use Azure Data Factory pipelines to periodically extract data from the clickstream and product tables, perform transformations in Data Flows, and create the new table in the data lakehouse.

C. Implement Azure Databricks notebooks to process the clickstream data and generate insights, then write the results to the new table in the lakehouse.

D. Create a stored procedure within the data warehouse that reads clickstream data, performs transformations, and outputs the results to a new table in the lakehouse.

* **C**

**(Incorrect)**

* **A**
* **D**
* **B**

**(Correct)**

**Explanation**

Answer: B

Explanation:

A: Azure Functions are suitable for real-time processing, but they might not be ideal for complex transformations and managing large datasets.

B: Azure Data Factory pipelines offer a robust and scalable solution for data extraction, transformation, and loading (ETL) processes. Data Flows provide a visual interface for building transformations, simplifying the process of creating and maintaining the new table with derived insights.

C: Databricks notebooks are powerful for complex transformations, but managing them separately from the data pipeline adds complexity.

D: Stored procedures are primarily designed for data manipulation within the data warehouse. They're not ideal for ETL processes involving external data sources and creating new tables in the lakehouse.

Bottom of Form

Top of Form

Question 28: **Incorrect**

Your company is migrating existing Power BI reports to Fabric and needs to maintain existing access controls for individual users.

Which approach best leverages Fabric's access control features to ensure smooth user transition?

A. Manually recreate user permissions for each report within Fabric workspaces.

B. Utilize Azure Active Directory (AAD) groups and map them to Fabric workspace roles.

C. Migrate user accounts directly from the existing platform to Fabric, including access settings.

D. Develop a custom Power BI extension for managing user access within Fabric.

* **A**

**(Incorrect)**

* **D**
* **C**
* **B**

**(Correct)**

**Explanation**

The correct answer is B. Utilize AAD groups and map them to Fabric workspace roles. Here's why:

Leverages existing AAD infrastructure: AAD is likely already used for company-wide authentication, making it efficient to map existing user groups to Fabric roles.

Simplifies user management: Mapping AAD groups streamlines user access updates and avoids manual recreation of permissions.

Reduces migration complexity: Mapping existing groups minimizes the need for user migration and data transfer.

Explanation of wrong answers:

A. Manually recreating permissions is time-consuming and error-prone.

C. Direct user account migration might not seamlessly integrate with Fabric's access control model.

D. Developing a custom extension is unnecessary for a common scenario like existing user access.

Additional Tips:

Thoroughly test your access control configuration before migrating users to ensure accurate permissions.

Consider using Fabric's self-service access request feature to empower authorized users to manage their own access to specific data sets.

Document your access control policies and procedures clearly for future reference and compliance adherence.

Bottom of Form

Top of Form

Question 29: **Correct**

You are troubleshooting website performance issues for Contoso's online store, stored in a Fabric lakehouse. The lakehouse contains webserver logs, user session data, and product information. You need to identify the most frequently accessed products with slow loading times.

Which SQL features would be MOST helpful in achieving this goal?

A. Window functions

B. Joins with subqueries

C. Regular expressions

D. Group by and aggregate functions

* **D**

**(Correct)**

* **A**
* **C**
* **B**

**Explanation**

D. Group by and aggregate functions

Explanation:

Window functions are useful for advanced calculations within partitions, not suitable for finding most frequent products.

Joins with subqueries could be used to combine data, but group by and aggregate functions offer a more straightforward approach.

Regular expressions are typically used for text parsing, not relevant for analyzing product access and loading times.

Group by and aggregate functions like COUNT, AVG, and MAX allow you to group products by access frequency and calculate average loading times within each group, effectively identifying the products with the most frequent slow loading issues.

Bottom of Form

Top of Form

Question 30: **Correct**

Your company ingests customer data from various sources into a Fabric data lakehouse. This data contains missing values, inconsistencies, and formatting errors. You need to implement a data cleansing process to prepare the data for analysis.

Which approach is MOST efficient and scalable for this scenario?

A. Develop custom Azure Functions for each data source to address specific cleansing requirements before loading the data into the lakehouse.

B. Implement Azure Data Factory (ADF) Data Flows with pre-built cleansing transformations to handle common issues like missing values and formatting.

C. Utilize Azure Databricks notebooks with custom Python scripts to perform complex data cleansing logic on the entire dataset within the lakehouse.

D. Design an Azure Data Lake Analytics (ADLA) U-SQL script for large-scale data cleansing using its built-in functions and custom logic.

* **B**

**(Correct)**

* **C**
* **A**
* **D**

**Explanation**

Answer: B

Explanation:

A: Custom Functions for each source add complexity and maintenance overhead. ADF Data Flows offer a centralized and scalable solution.

B: ADF Data Flows with pre-built cleansing transformations provide a readily available and efficient way to handle common data issues like missing values, formatting, and data types. This minimizes the need for custom coding and offers a visual interface for easy configuration.

C: Databricks notebooks offer flexibility for complex logic, but they can be resource-intensive and require separate deployment and management. ADF Data Flows provide a more efficient and scalable solution for common cleansing tasks.

D: ADLA U-SQL scripts can be complex for data cleansing and might not be as user-friendly as ADF Data Flows for visual configuration. Data Flows offer a more accessible option for this scenario.

Bottom of Form

Top of Form

Question 31: **Correct**

Your company needs to orchestrate data pipelines for various analytics workloads within a Fabric data lakehouse. Some pipelines require daily processing of large datasets, while others need near-real-time updates based on streaming data sources.

Which Azure service is MOST suitable for scheduling both types of data pipelines efficiently?

A. Azure Data Factory (ADF) with trigger schedules and linked services for different data sources.

B. Azure Data Lake Analytics (ADLA) with U-SQL scripts and scheduled triggers for batch processing.

C. Azure Databricks notebooks with schedule settings and custom scripts for both batch and streaming data.

D. Azure Event Grid with event handlers and linked services to trigger processing pipelines for streaming data.

* **A**

**(Correct)**

* **C**
* **D**
* **B**

**Explanation**

Answer: A

Explanation:

A: Azure Data Factory offers a versatile and centralized platform for scheduling and orchestrating both batch and streaming data pipelines. Its trigger schedules can handle daily recurring executions for large datasets, while linked services seamlessly integrate with various data sources and sinks. For streaming data, ADF leverages triggers like Azure Event Hubs, allowing near-real-time processing.

B: ADLA is primarily designed for large-scale batch analytics using U-SQL scripts. While offering scheduled triggers, it lacks the flexibility and versatility of ADF in handling diverse data workloads and streaming events.

C: Databricks notebooks provide flexibility for custom scripts and scheduling, but managing individual notebooks for multiple schedules can be cumbersome and less centralized than ADF.

D: Azure Event Grid focuses on event routing and triggering actions, not on orchestrating complex data pipelines with diverse sources, transformations, and destinations. It can be used within an ADF pipeline for streaming data triggers, but not as the primary scheduling solution.

Bottom of Form

Top of Form

Question 32: **Incorrect**

You're developing a Power BI report for a retail company to analyze sales performance across different regions and product categories. The data model includes a large fact table with millions of rows and several dimension tables. To improve query performance and user experience, you need to implement semantic models.

Which of the following scenarios BEST justifies using DAX Studio and Tabular Editor 2 for semantic model design and development?

A. You need to create calculated columns and measures to enrich the data model with new insights.

B. You want to define relationships between dimension tables and the fact table using Relationship Editor.

C. You need to configure data partitioning and aggregation strategies to optimize query performance.

D. You want to import data from various sources and create a data model from scratch.

* **A**

**(Incorrect)**

* **D**
* **B**
* **C**

**(Correct)**

**Explanation**

The correct answer is C. You need to configure data partitioning and aggregation strategies to optimize query performance.

Here's why:

A: While DAX Studio and Tabular Editor 2 can be used to create calculated columns and measures, they are not the only tools available for this task. Power BI Desktop also offers similar functionalities.

B: Defining relationships is primarily done through Power BI Desktop's Relationship Editor, not DAX Studio or Tabular Editor 2.

D: Importing data and creating a data model from scratch are core functionalities of Power BI Desktop, not DAX Studio or Tabular Editor 2.

C is the most relevant scenario because DAX Studio and Tabular Editor 2 provide advanced features for semantic model design, specifically data partitioning and aggregation strategies. These features are crucial for optimizing query performance in large data models like the one described in the scenario.

Bottom of Form

Top of Form

Question 33: **Incorrect**

Your company analyzes website traffic data stored in a Fabric data lakehouse. The data includes website visits, user sessions, and product interactions. You need to design a star schema for efficient analysis, including considerations for slowly changing dimensions (SCDs).

Which dimension attributes should be implemented as Type 1 and Type 2 SCDs, respectively?

A. User name (Type 1), Product name (Type 2)

B. User ID (Type 1), Website URL (Type 2)

C. Visit timestamp (Type 1), User session ID (Type 2)

D. Session duration (Type 1), User type (e.g., new, returning) (Type 2)

* **C**

**(Incorrect)**

* **B**
* **A**
* **D**

**(Correct)**

**Explanation**

Answer: D

Explanation:

A: User name can change (e.g., marriage), hence Type 2 SCD. Product name might change due to marketing or upgrades, also making it Type 2.

B: User ID is permanent, so Type 1 SCD. Website URL might change due to website redesign, making it Type 2.

C: Visit timestamp is a factual data point, not changing (Type 1). User session ID is temporary and unique, so Type 1.

D: Session duration is a calculated measure, not directly stored. User type can change based on behavior, making it Type 2.

Bottom of Form

Top of Form

Question 34: **Incorrect**

You are a data scientist at Fabrikam, a manufacturing company, tasked with improving production efficiency. You have access to a vast dataset containing sensor readings from machines, production logs, and employee profile data.

Which of the following approaches would be MOST effective in using profile data for exploratory analytics to identify potential efficiency gains?

A. Create a scatter plot comparing machine sensor readings with employee productivity metrics.

B. Build a machine learning model to predict machine failures based on historical sensor data.

C. Segment employee profiles based on skillset, experience, and machine operation data, then analyze production output by segment.

D. Implement anomaly detection algorithms on sensor readings to identify unusual machine behavior.

* **B**

**(Incorrect)**

* **C**

**(Correct)**

* **A**
* **D**

**Explanation**

C. Segment employee profiles based on skillset, experience, and machine operation data, then analyze production output by segment.

Explanation:

Option A focuses on direct correlation between machine data and employee productivity, neglecting the potential influence of other factors like skillset or experience.

Option B focuses solely on machine learning for anomaly detection, ignoring the potential value of understanding human factors impacting efficiency.

Option D addresses machine anomalies but doesn't explore how employee profiles influence production output.

Option C directly leverages profile data (skillset, experience) to segment employees and analyze their impact on production output, providing valuable insights into potential efficiency gains based on human factors.

Bottom of Form

Top of Form

Question 35: **Incorrect**

You are designing a semantic model for a Power BI dashboard that analyzes website traffic data stored in Azure Data Lake Storage. The data updates daily and includes large volumes of clickstream events. You need the model to provide fast responses for real-time ad campaign performance monitoring and historical trend analysis.

Which storage mode for the semantic model would be the most suitable in this scenario?

A. Import mode: Load the entire dataset into the semantic model for fast query performance.

B. DirectQuery mode: Connect the model directly to the data lake for real-time updates but potentially slower queries.

C. Composite model: Combine import and DirectQuery modes, keeping frequently used data in the model for speed and connecting to the lake for large datasets.

D. Direct Lake mode: Access and process the data directly within the lake using Power BI, leveraging the lake's scalability for large datasets.

* **B**

**(Incorrect)**

* **C**
* **A**
* **D**

**(Correct)**

**Explanation**

D. Direct Lake mode

Explanation:

Option A is inefficient for large datasets, leading to storage and performance issues.

Option B provides real-time updates but might not be performant for large datasets due to latency in DirectQuery connections.

Option C offers flexibility, but frequent data updates might create maintenance challenges.

Option D directly addresses all requirements. Direct Lake mode allows Power BI to access and process the data within the lake without data movement or model storage overhead. This provides real-time updates, efficient processing for large datasets, and fast query performance for both ad campaign monitoring and historical analysis.

Bottom of Form

Top of Form

Question 36: **Incorrect**

You're building a Power BI report for a marketing agency analyzing campaign performance across different regions and product categories. The data model includes dimensions for Region, Product Category, and Marketing Channel. You want to create a dynamic chart title based on user selections in slicers for region and product category.

Which combination of calculation groups, dynamic strings, and field parameters best achieves this?

A. Create a calculation group with calculated columns:

\* RegionTitle: DAX CONCATENATE("[", DAX VALUE(Region[Name]), "]")

\* CategoryTitle: DAX CONCATENATE("[", DAX VALUE(ProductCategory[Name]), "]")

\* ChartTitle: CONCATENATE(IF(ISBLANK(RegionTitle), "", RegionTitle), IF(ISBLANK(CategoryTitle), "", " - "), CategoryTitle)

B. Create a single calculated column with DAX SWITCH, CONCATENATE, and ISBLANK functions to dynamically generate the entire chart title based on slicer selections.

C. Create field parameters for Region and Product Category. Use SWITCH statements in the chart title text box to dynamically include selected parameter values.

D. Use dynamic measures with CALCULATE and FILTER functions to dynamically change the measure displayed based on slicer selections, then reference the dynamically chosen measure's name in the chart title.

* **B**

**(Incorrect)**

* **C**
* **A**

**(Correct)**

* **D**

**Explanation**

The correct answer is A. Create a calculation group with calculated columns...

Here's why:

A effectively utilizes calculation groups and dynamic strings:

Separate calculated columns for RegionTitle and CategoryTitle allow independent dynamic construction with CONCATENATE and DAX VALUE.

ChartTitle combines these titles with IF statements to check for blank values and handle empty selections without adding unnecessary brackets.

B while possible, using a single SWITCH statement can be complex and difficult to maintain compared to separate, modular calculations.

C field parameters can influence visuals through filters but are not directly used for constructing dynamic titles.

D dynamic measures are suitable for changing data calculations, not directly manipulating titles based on user selections.

Explanation of wrong answers:

B: A single SWITCH statement can be complex and less understandable than separate calculations.

C: Field parameters are primarily for filtering data, not directly constructing titles.

D: Dynamic measures change data calculations, not titles based on user selections.

Bottom of Form

Top of Form

Question 37: **Incorrect**

Your company needs to analyze real-time social media data from Twitter and Facebook feeds for marketing campaigns. The analysis requires near-instantaneous data ingestion, cleansing, and aggregation into Fabric data warehouse tables.

How should you schedule dataflows and notebooks for this scenario to achieve efficient real-time data analysis?

A. Use ADF triggers based on new tweets and Facebook posts to trigger dataflows for processing and loading into the data warehouse.

B. Schedule dataflows to run every minute to pull data from social media APIs and perform analysis within the dataflows themselves.

C. Develop Azure Functions triggered by social media events to process data and write it to staging tables in the data warehouse. Schedule separate dataflows to analyze these tables hourly.

D. Deploy Azure Databricks notebooks with continuous streams connected to social media APIs. Configure notebooks to continuously analyze and write results to the data warehouse tables.

* **A**

**(Incorrect)**

* **B**
* **D**

**(Correct)**

* **C**

**Explanation**

Answer: D

Explanation:

A: While ADF triggers can react to events, relying on them for every tweet and post might be resource-intensive and inefficient.

B: Scheduling dataflows every minute is unnecessary for real-time data. It adds overhead and consumes resources.

C: Functions triggered by events offer efficient data processing, but writing to staging tables and then scheduling separate dataflows adds unnecessary complexity and latency.

D: Azure Databricks notebooks with continuous streams provide the best solution for this scenario. They connect directly to social media APIs and continuously analyze data as it arrives, writing the results directly to the data warehouse tables in near-real-time, achieving the desired responsiveness for campaign analysis.

Bottom of Form

Top of Form

Question 38: **Incorrect**

You're working on a complex financial reporting application for a bank. The data model includes multiple fact and dimension tables with intricate relationships and calculations. You need to ensure data quality and consistency throughout the development process.

Which combination of DAX Studio and Tabular Editor 2 features would be MOST effective for achieving data quality and consistency in this scenario?

A. Using calculated columns and measures to enforce data validation rules.

B. Implementing data lineage tracking to trace the origin of data points.

C. Employing schema compare and merge functionalities to manage model versions.

D. Leveraging veritas table creation and management for data quality checks.

* **D**

**(Incorrect)**

* **C**

**(Correct)**

* **A**
* **B**

**Explanation**

The correct answer is C. Employing schema compare and merge functionalities to manage model versions.

Here's why:

A: While calculated columns and measures can be used for data validation, they are not the primary tools for managing data quality in complex data models.

B: Data lineage tracking is a valuable feature, but it doesn't directly address data quality and consistency issues.

D: Veritas tables are primarily used for data governance and security, not directly for data quality and consistency.

C is the most relevant option because schema compare and merge functionalities in DAX Studio and Tabular Editor 2 allow you to track changes and differences between model versions. This helps to identify and address potential data inconsistencies and maintain data quality throughout the development process.

Explanation of wrong answers:

A: While calculated columns and measures can be used for data validation, they are not the primary tools for managing data quality in complex data models.

B: Data lineage tracking is a valuable feature, but it doesn't directly address data quality and consistency issues.

D: Veritas tables are primarily used for data governance and security, not directly for data quality and consistency.

Bottom of Form

Top of Form

Question 39: **Incorrect**

Your company needs to copy real-time data from Azure Event Hubs into a Fabric data lakehouse for near-instantaneous analytics. The data needs to be cleansed and transformed before landing in the lakehouse.

Which of the following methods is the MOST appropriate for this scenario?

A. Schedule an Azure Data Factory (ADF) pipeline with Data Flows to run periodically and process streamed data from Event Hubs.

B. Develop Azure Functions triggered by Event Hubs messages to perform cleansing and transformation before writing to the lakehouse using the Fabric connector.

C. Utilize Azure Databricks notebooks connected to Event Hubs to process the data in real-time and store it in the lakehouse with desired transformations.

D. Implement Azure Stream Analytics with a tumbling window function to aggregate real-time data and periodically write it to the lakehouse in batches.

* **C**

**(Incorrect)**

* **A**
* **D**
* **B**

**(Correct)**

**Explanation**

Answer: B

Explanation:

A: Scheduled ADF pipelines are not suitable for real-time data as they run periodically.

B: Azure Functions triggered by Event Hubs messages offer a reactive and scalable approach for real-time data processing. They can perform cleansing and transformation before writing the data to the lakehouse using the Fabric connector, achieving near-instantaneous availability.

C: Databricks notebooks can handle real-time data, but managing and deploying notebooks separately adds complexity. Functions are more readily integrated with the data pipeline.

D: Azure Stream Analytics works well for real-time analytics and filtering, but it might not be ideal for writing entire datasets to the lakehouse. Also, tumbling window functions may not be optimal for all types of data transformations.

Bottom of Form

Top of Form

Question 40: **Incorrect**

Your company needs to continuously integrate data from a changing CSV file into a Fabric data warehouse. The data needs to be validated and filtered before loading into specific tables based on its content.

Which option is the MOST suitable for this continuous data integration scenario?

A. Schedule an ADF pipeline with a copy activity to run periodically and load the entire CSV file into the warehouse, followed by separate activities for validation and filtering.

B. Implement Azure Data Factory with a data flow connected to the CSV file and perform validation, filtering, and loading within the data flow for real-time data processing.

C. Develop an Azure Function triggered by file changes in the CSV location to perform validation, filtering, and loading the data into the appropriate tables in the warehouse.

D. Utilize Azure Databricks notebooks to continuously monitor the CSV file, perform transformations, and load the data into the relevant tables within the warehouse.

* **A**

**(Incorrect)**

* **D**
* **C**
* **B**

**(Correct)**

**Explanation**

Answer: B

Explanation:

A: Scheduled ADF pipelines are not ideal for continuous data integration as they run periodically.

C: While Azure Functions are good for reacting to file changes, their logic might not be suitable for complex data transformations and loading into specific tables based on content.

B: ADF's data flows offer a powerful and flexible solution for continuous data integration. You can connect the data flow to the CSV file, perform real-time validation, filtering, and loading into the appropriate tables based on the data content. This eliminates the need for separate activities and ensures timely data processing.

D: Databricks notebooks can handle continuous data integration, but their complexity might not be necessary for this scenario. ADF's data flow provides a simpler and more efficient approach.

Bottom of Form

Top of Form

Question 41: **Incorrect**

Your company's Fabric data lakehouse stores website clickstream data for analyzing user behavior. The data is stored in large parquet files, but query performance is slow for specific date ranges frequently accessed by marketing teams.

How can you implement file partitioning to improve query performance for these date-specific queries?

A. Create separate folders for each date within the data lakehouse and store the corresponding data files in each folder.

B. Partition the existing parquet files based on the date column and store them within the same folder.

C. Convert the data to a columnar format like Parquet and leverage the file format's built-in partitioning capabilities.

D. Use Azure Data Factory (ADF) to copy the data into new parquet files partitioned by date and store them in a separate location.

* **C**

**(Incorrect)**

* **A**
* **B**

**(Correct)**

* **D**

**Explanation**

Answer: B

Explanation:

A: While creating separate folders might improve organization, it doesn't directly address query performance.

B: Partitioning the existing parquet files based on the date column leverages the file format's efficiency for accessing specific date ranges. This avoids scanning the entire dataset, significantly improving query performance for date-specific queries.

C: Converting to a columnar format like Parquet is already done and doesn't address the specific issue of date-based partitioning.

D: Copying the data to new files with ADF adds an additional processing step and might not be necessary. Partitioning the existing files within the same location is more efficient.

Bottom of Form

Top of Form

Question 42: **Incorrect**

Your company's Fabric environment houses multiple data lakes containing various datasets. You frequently access specific folders within these lakes for analysis. To improve efficiency, you want to create shortcuts to these folders.

Which of the following methods is the MOST efficient and flexible way to create and manage shortcuts in this scenario?

A. Create symbolic links within the data lakes themselves, pointing to the desired folders.

B. Develop custom Azure Functions to map folder locations and access them through function calls.

C. Use Fabric's built-in shortcut functionality, creating shortcuts within the "Lake" view of the lakehouse explorer.

D. Configure Azure Data Factory datasets to point to specific folders within the data lakes and access them through pipeline activities.

* **D**

**(Incorrect)**

* **B**
* **C**

**(Correct)**

* **A**

**Explanation**

Answer: C

Explanation:

A: While symbolic links work, they only function within the specific data lake and are not easily managed across Fabric.

B: Custom Azure Functions add unnecessary complexity and maintenance overhead for simple shortcut creation.

C: Fabric's built-in shortcut functionality provides a dedicated and user-friendly way to create and manage shortcuts to any folder within any Fabric data lake. Shortcuts appear as folders in the Lake view, offering convenient access and visibility.

D: Azure Data Factory datasets are intended for data processing within pipelines, not for general file access.

Bottom of Form

Top of Form

Question 43: **Incorrect**

Your company wants to build a large format dataset for analyzing customer behavior across multiple online platforms. The data sources include website clickstream data, purchase transactions, and social media interactions. The dataset needs to be scalable, performant, and support efficient querying for various user segments and timeframes.

Which approach would be the most suitable for designing and building this large format dataset, considering Microsoft Fabric and its capabilities?

A. Implement a traditional data warehouse using Azure SQL Database.

B. Utilize Azure Synapse Analytics as the primary data store and leverage Power BI for data analysis.

C. Develop a data lake in Azure Data Lake Storage and utilize Azure Databricks for data processing and analysis.

D. Design a data fabric using Azure Data Factory for data ingestion and Azure HDInsight for distributed data processing and storage.

* **C**

**(Incorrect)**

* **B**

**(Correct)**

* **A**
* **D**

**Explanation**

The correct answer is B. Utilize Azure Synapse Analytics as the primary data store and leverage Power BI for data analysis.

Here's why:

A: While Azure SQL Database is a robust relational database, it's not optimized for large-scale data volumes and diverse data types commonly found in customer behavior analysis.

B: Azure Synapse Analytics is specifically designed for large-scale data warehousing and analytics, offering scalability, performance, and integration with Power BI for efficient data exploration and visualization.

C: While Azure Data Lake Storage and Azure Databricks are powerful options for data lake and distributed analytics, they require more development effort and expertise compared to Azure Synapse Analytics' pre-built capabilities for data warehousing and integration with Power BI.

D: Azure Data Factory is primarily an orchestration tool for data movement and doesn't offer built-in data warehousing or analytics functionalities. Azure HDInsight can be used for distributed processing but requires additional configuration and management for a complete data analysis solution.

Explanation of wrong answers:

A: Azure SQL Database is not ideal for large-scale, diverse data volumes in this scenario.

C: Azure Data Lake Storage and Azure Databricks require more development effort and expertise compared to Azure Synapse Analytics for this specific use case.

D: Azure Data Factory and Azure HDInsight are more focused on data movement and processing, not directly catering to data warehousing and analytics.

Bottom of Form

Top of Form

Question 44: **Correct**

Your company needs to build a data pipeline in Azure Data Factory (ADF) to cleanse and transform data from various sources before loading it into a Fabric data warehouse. The pipeline requires invoking complex stored procedures within the source databases and performing advanced data manipulations.

How can you effectively leverage stored procedures, notebooks, and dataflows within your ADF pipeline to achieve this objective?

A. Develop custom Azure Functions to call the stored procedures and perform transformations, then use Data Flows to integrate the results into the pipeline.

B. Directly embed the stored procedure calls within ADF dataflows using scripting or linked services, followed by data manipulation activities within the flow.

C. Utilize Azure Databricks notebooks connected to the source databases for complex transformations, then use an ADF copy activity to load the processed data into the data warehouse.

D. Implement stored procedures within the data warehouse to handle complex transformations, triggered by the ADF pipeline after data is loaded.

* **B**

**(Correct)**

* **C**
* **A**
* **D**

**Explanation**

Answer: B

Explanation:

A: Custom Functions are unnecessary as ADF can directly integrate with stored procedures.

B: Embedding stored procedure calls within ADF dataflows offers the most efficient and integrated approach. You can leverage scripting or linked services to connect to the source databases and execute the procedures, while dataflows handle subsequent data manipulation tasks within the pipeline.

C: Databricks notebooks add unnecessary complexity and require separate deployment and management. ADF's built-in capabilities are sufficient for this scenario.

D: Implementing stored procedures within the data warehouse introduces additional complexity and adds unnecessary processing after data is already loaded. ADF's data manipulation activities can handle transformations more efficiently within the pipeline.

Bottom of Form

Top of Form

Question 45: **Incorrect**

Your company needs to update specific data in a Fabric data lakehouse based on real-time changes from Azure Event Hubs. The updates involve modifying existing files within the lakehouse.

Which of the following approaches is MOST suitable for this real-time data update scenario?

A. Schedule an ADF pipeline with Data Flows to run periodically and check for updates in Event Hubs, modifying lakehouse files accordingly.

B. Develop Azure Functions triggered by Event Hubs messages, reading updates, and directly modifying the relevant files within the lakehouse using Fabric APIs.

C. Utilize Azure Databricks notebooks connected to Event Hubs to process updates and write the modified data to a separate staging location, triggering an ADF pipeline for lakehouse file updates.

D. Implement Azure Stream Analytics with a join function to combine updates from Event Hubs with existing lakehouse data, generating new files with the updated information.

* **A**

**(Incorrect)**

* **C**
* **B**

**(Correct)**

* **D**

**Explanation**

Answer: B

Explanation:

A: Scheduled ADF pipelines are not suitable for real-time data updates.

B: Azure Functions triggered by Event Hubs offer a reactive and efficient approach. They can directly modify the relevant lakehouse files using Fabric APIs, providing near-instantaneous updates without requiring additional staging or pipeline runs.

C: Databricks notebooks add unnecessary complexity and increase latency compared to direct file modification.

D: Azure Stream Analytics might not be the best fit for file manipulation. Functions can handle the updates more effectively and directly within the lakehouse.

Bottom of Form

Top of Form

Question 46: **Correct**

You are analyzing sales data stored in a Fabric lakehouse for AdventureWorks, an outdoor gear retailer. You need to identify regions with the highest potential for growth based on customer demographics and purchase history. The lakehouse contains multiple data sources, including customer tables, sales tables, and geographic data.

Which approach would be MOST efficient for writing a SQL query to achieve this goal?

A. Write a single SQL query joining all data sources directly.

B. Use the Fabric visual query editor to create a visual representation of the query.

C. Create separate SQL queries for each data source and combine the results in a final query.

D. Leverage external tools like Excel or Power BI to connect to the lakehouse and analyze the data

* **B**

**(Correct)**

* **D**
* **C**
* **A**

**Explanation**

B. Use the Fabric visual query editor to create a visual representation of the query.

Explanation:

Option A would be complex and prone to errors when joining multiple data sources directly in a single SQL query.

Option C involves writing and managing multiple queries, increasing development time and maintenance complexity.

Option D bypasses Fabric's native lakehouse querying capabilities and introduces external dependencies.

Option B leverages the Fabric visual query editor, specifically designed for lakehouse exploration. It allows visually building and testing queries through drag-and-drop functions, simplifying complex joins and data visualizations, thus achieving the goal efficiently.

Bottom of Form

Top of Form

Question 47: **Correct**

Your company manages a data pipeline in ADF that cleanses and transforms data from on-premises SQL Server databases before loading it into a Fabric data warehouse. The pipeline runs daily at midnight but sometimes encounters errors and requires manual intervention for restart.

How can you improve the pipeline's reliability and reduce dependence on manual intervention?

A. Implement retry policies within the ADF pipeline activities to automatically handle and recover from certain errors.

B. Schedule additional ADF pipelines with different triggers to run the same processing logic in case of failure.

C. Leverage Azure Monitor alerts to notify you of pipeline failures and manually restart the pipeline using the ADF interface.

D. Convert the ADF pipeline to an Azure Data Lake Analytics (ADLA) U-SQL script for simpler error handling and automatic retries.

* **A**

**(Correct)**

* **C**
* **D**
* **B**

**Explanation**

Answer: A

Explanation:

A: Implementing retry policies within the ADF pipeline activities is the most efficient solution. You can configure retries for specific errors, allowing the pipeline to automatically recover without manual intervention. This significantly improves reliability and reduces downtime.

B: Running duplicate pipelines is inefficient and adds unnecessary resource consumption. It doesn't address the issue of error handling within the primary pipeline.

C: While Azure Monitor alerts are helpful for notification, they require manual intervention to restart the pipeline, which defeats the purpose of improving automation and reducing dependency on manual efforts.

D: Converting the pipeline to ADLA U-SQL offers limited benefits in this scenario. Retry policies within ADF provide a more targeted and efficient solution for automatic error handling.

Bottom of Form

Top of Form

Question 48: **Incorrect**

Your company's Fabric data lakehouse stores sensor data from various manufacturing machines, with files partitioned by machine ID and timestamp. You need to analyze data trends across multiple machines for a specific time period.

Which partitioning scheme is most suitable for this scenario, considering both performance and flexibility for future analysis?

A. Maintain the existing partitioning by machine ID and timestamp without any changes.

B. Add an additional level of partitioning by date within each machine ID directory.

C. Implement hierarchical partitioning with the first level based on machine ID and the second level based on date.

D. Convert all data to a columnar format like Parquet, regardless of the partitioning scheme.

* **D**

**(Incorrect)**

* **A**
* **C**

**(Correct)**

* **B**

**Explanation**

Answer: C

Explanation:

A: While existing partitioning is helpful, it doesn't facilitate efficient analysis across multiple machines.

B: Adding a date level within each machine ID might improve performance for specific date-based queries, but it can lead to many small partitions and hinder analysis across multiple machines.

C: Hierarchical partitioning offers the best balance of performance and flexibility. The first level based on machine ID keeps data organized by source, while the second level based on date allows efficient querying across multiple machines for specific timeframes. This structure is also flexible for future analysis involving additional dimensions.

D: While columnar formats like Parquet can improve compression and reading performance, they don't address the partitioning scheme itself. Combining a suitable partitioning strategy with a columnar format optimizes both data access and storage efficiency.

Bottom of Form

Top of Form

Question 49: **Correct**

Your company needs to migrate historical data from an on-premises SQL Server database to a Fabric data warehouse. The data will be used for various reports and requires minimal transformation beyond basic mapping and filtering.

Which of the following methods is the MOST efficient and cost-effective for this data migration?

A. Use Azure Data Factory (ADF) with a copy activity directly pointing to the SQL Server database and the target data warehouse table.

B. Develop a custom Azure Function leveraging the ODBC driver to extract data from SQL Server and stage it in Azure Blob Storage before loading it into the warehouse.

C. Migrate the data using the Azure Import/Export service with a bulk copy operation for fast data transfer.

D. Deploy Azure Databricks and write Spark code to read the SQL Server data, transform it as needed, and write it to the data warehouse.

* **A**

**(Correct)**

* **B**
* **D**
* **C**

**Explanation**

Answer: A

Explanation:

A: Azure Data Factory's copy activity with a built-in connector for SQL Server offers a simple and efficient solution for bulk data migration. This eliminates the need for custom code and provides built-in data transfer optimization.

B: Custom Azure Functions add unnecessary complexity and maintenance overhead. The direct ADF copy activity is readily available and efficient.

C: While Azure Import/Export is suitable for large dataset migrations, it's not optimal for real-time or frequent data transfers. ADF offers more flexibility and control.

D: Databricks provides flexibility for complex transformations, but it's overkill for this scenario requiring minimal data manipulation. ADF's copy activity is more efficient and cost-effective in this case.

Bottom of Form

Top of Form

Question 50: **Incorrect**

In the ABC university data model, you need to analyze student performance across different courses and semesters. However, the Enrollments table doesn't directly store student location information. There's a separate Locations table with city and country details. How would you best establish relationships to enable analysis based on student location?

A. Create a bridge table linking Enrollments and Locations based on student\_id and a new location\_id column.

B. Add a location\_id column to the Enrollments table and directly link it to the Locations table using the city and country data.

C. Denormalize the Locations table by adding city and country columns to the Students table.

D. Use calculated columns in the Enrollments table to derive location information based on student\_id and external data sources.

* **D**

**(Incorrect)**

* **B**
* **A**

**(Correct)**

* **C**

**Explanation**

The correct answer is A. Create a bridge table linking Enrollments and Locations based on student\_id and a new location\_id column.

Here's why:

A is the most efficient and flexible solution. The bridge table avoids data redundancy and allows for future expansion of location data without modifying the existing tables.

B directly links Enrollments and Locations, but it adds unnecessary location data to the Enrollments table, potentially increasing storage and performance overhead.

C denormalizes the Locations table, which might lead to data redundancy and inconsistencies if student location information changes.

D introduces complexity with calculated columns and might not be as efficient as a dedicated bridge table for location analysis.

Explanation of wrong answers:

B: Adds unnecessary location data to the Enrollments table, potentially impacting performance and storage.

C: Denormalization can lead to data redundancy and inconsistencies.

D: Calculated columns introduce complexity and might not be as efficient as a dedicated bridge table.

Bottom of Form