**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: **Incorrect**

Your company wants to create a custom Power BI theme in Fabric for their marketing team. They prioritize a clean and modern look with a focus on data clarity and accessibility.

Which elements should you prioritize when designing the custom theme?

A. Vibrant color palette with contrasting shades for emphasis.

B. Extensive use of chart and graph types for detailed data exploration.

C. Highlighting key metrics with bold fonts and large sizes.

D. Implementing custom icons and logos for brand consistency.

* **C**

**(Correct)**

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

**(Incorrect)**

**Explanation**

The correct answer is C. Highlighting key metrics with bold fonts and large sizes. Here's why:

Data clarity and accessibility: The theme should prioritize clear and concise presentation of data, especially for key metrics, using features like bold fonts, large sizes, and high contrast.

Clean and modern look: Vibrant colors and extensive chart types can clutter the visuals and distract from the data. Custom icons and logos can be incorporated subtly for branding without compromising clarity.

Explanation of wrong answers:

A. Vibrant colors can be overwhelming and difficult for color-blind users.

B. While exploring data is important, the theme should prioritize clear communication of key insights over an abundance of chart types.

D. Custom icons and logos should be used sparingly to maintain a clean and focused visual experience.

Additional Tips:

Consider using accessibility tools and guidelines to ensure your custom theme is inclusive for users with different abilities.

Test the theme with real users and gather feedback to refine it for optimal usability.

Share the custom theme within the organization to ensure consistency and promote its adoption.

Bottom of Form

Top of Form

Question 2: **Incorrect**

You are tasked with designing a data analytics environment in Microsoft Fabric for a large retail company with 100+ stores across multiple countries. The company requires near real-time analytics on store sales data, including daily sales trends, product popularity analysis, and anomaly detection. Additionally, they need to share insights with different departments, including marketing, merchandising, and operations.

Which settings in the Fabric admin portal would you recommend for optimal performance and scalability?

A. Enable Azure Synapse Analytics as the data lakehouse, configure Premium SKU for data warehouses, and set high concurrency for workspaces.

B. Implement Azure Data Lake Storage for raw data, configure Standard SKU for data warehouses, and enable workspace sharing with granular access control.

C. Use Azure SQL Database for the data warehouse, set low concurrency for workspaces, and limit access to analytics features for specific departments.

D. Utilize Azure Data Factory for data pipelines, configure Standard SKU for lakehouses, and enable data sharing across all departments with open access.

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

**(Correct)**

* **D**

**(Incorrect)**

* **A**

**Explanation**

The correct answer is B. Here's why:

Azure Data Lake Storage (ADLS) is a more cost-effective and scalable solution for storing large amounts of raw data compared to Azure SQL Database.

Standard SKU for data warehouses strikes a balance between performance and cost for near real-time analytics, while Premium SKU might be overkill for this scenario.

Workspace sharing with granular access control allows different departments to access specific data sets and insights relevant to their needs, ensuring data security and compliance.

Azure Data Factory (ADF) is a robust and flexible platform for building and orchestrating data pipelines, efficiently moving data between ADLS and the data warehouse.

Explanation of wrong answers:

A: While Azure Synapse Analytics is a powerful data lakehouse, it might be too expensive for this scenario. High concurrency for workspaces is unnecessary with controlled access.

C: Azure SQL Database is not optimized for large-scale analytics and near real-time processing. Low concurrency would hinder performance, and limiting access restricts collaboration.

D: Open data sharing across departments is a security risk. Standard SKU for lakehouses might not handle the data volume effectively.

Bottom of Form

Top of Form

Question 3: **Correct**

You are managing a large financial reporting model in Power BI Desktop with many measures and relationships. You need to bulk edit measures across the model to implement updates in data formatting and aggregation logic.

Which tool would be most effective for performing these bulk edits and managing the overall structure of the semantic model?

A. Power BI Desktop, using the measure editor and relationship management tools.

B. Tabular Editor 2, enabling bulk editing of measures, calculated columns, and relationships.

C. DAX Studio, focusing on individual DAX function authoring and testing.

D. Azure Analysis Services, managing model deployment and security but not offering editing functionalities.

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

**(Correct)**

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

**Explanation**

B. Tabular Editor 2

Explanation:

Option A requires manual adjustments in Power BI Desktop, a time-consuming and error-prone process for bulk edits across many measures.

Option B, Tabular Editor 2, excels in this scenario. It offers efficient bulk editing functionalities for measures, calculated columns, and relationships, allowing you to implement formatting and logic updates across the entire model simultaneously.

Option C focuses on individual DAX function authoring, not model-wide modifications.

Option D deals with deployment and security, not model editing capabilities.

Bottom of Form

Top of Form

Question 4: **Incorrect**

You're building a Power BI report for a retail chain to analyze top-performing products by region and year. The data model includes:

Sales (sale\_id, product\_id, region\_id, year, sales\_amount)

Products (product\_id, name)

Regions (region\_id, name)

You want to create a measure that identifies the top 5 products with the highest total sales for each region in each year. Which DAX expression would achieve this effectively?

A. Top 5 Products = VAR TopCount = 5; CALCULATE(SUM(Sales[sales\_amount]), FILTER(Sales, Sales[region\_id] = CURRENTREGION() && Sales[year] = YEAR(CURRENTDATE())))

B. Top 5 Products = VAR TopProducts = TOPN(5, Sales, [sales\_amount], DESC); RETURN SUMX(TopProducts, [sales\_amount])

C. Top 5 Products = VAR TopRank = CALCULATE(COUNTROWS(Sales), FILTER(Sales, Sales[region\_id] = CURRENTREGION() && Sales[year] = YEAR(CURRENTDATE()))) - ROW\_NUMBER() OVER (PARTITION BY Sales[region\_id], Sales[year] ORDER BY SUM(Sales[sales\_amount]) DESC); RETURN IF(TopRank <= 5, [sales\_amount], BLANK())

D. Top 5 Products = CALCULATE(SUM(Sales[sales\_amount]), TOPN(5, Sales, [sales\_amount], DESC))

**A**

**(Incorrect)**

* **C**

**(Correct)**

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

**Explanation**

The correct answer is C. Top 5 Products = VAR TopRank = CALCULATE(COUNTROWS(Sales), FILTER(Sales, Sales[region\_id] = CURRENTREGION() && Sales[year] = YEAR(CURRENTDATE()))) - ROW\_NUMBER() OVER (PARTITION BY Sales[region\_id], Sales[year] ORDER BY SUM(Sales[sales\_amount]) DESC); RETURN IF(TopRank <= 5, [sales\_amount], BLANK())

Here's why:

A filters the Sales table for the current region and year but doesn't identify the top 5 products within that context.

B uses TOPN but doesn't filter by region or year and only outputs the total sales amount for the top 5 products, not individual product details.

C utilizes variables and window functions effectively:

TopRank calculates the current product's rank within its region and year based on total sales.

ROW\_NUMBER() assigns a rank to each product within the filtered context.

IF statement returns the sales amount only for products within the top 5 rankings.

D again uses TOPN without filtering and only provides the total sales for the top 5 products, not individual details.

Bottom of Form

Top of Form

Question 5: **Incorrect**

You are developing a real-time COVID-19 case dashboard powered by Microsoft Fabric. The data includes rapidly incoming updates on cases, vaccinations, and hospitalizations from various sources, requiring quick refresh and responsiveness.

Which storage mode for the semantic model would be most appropriate to ensure efficient data consumption, minimize refresh times, and optimize performance for real-time analysis?

A. Import mode, pre-aggregating data in the semantic model for faster query execution.

B. DirectQuery mode, accessing data directly from Azure Synapse Analytics without model copies.

C. Incremental refresh mode, automatically updating the model with new data at predefined intervals.

D. Live connections mode, dynamically refreshing the model with near real-time data from Azure Data Lake Storage.

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

**(Incorrect)**

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

**(Correct)**

**Explanation**

D. Live connections mode, dynamically refreshing the model with near real-time data from Azure Data Lake Storage.

Explanation:

Option A might be efficient for static data sets, but wouldn't provide near real-time updates for COVID-19 data.

Option B offers flexibility but can have latency issues accessing large Synapse tables.

Option C works well for periodic updates, but doesn't fulfill the near real-time requirement.

Option D directly addresses the need for immediate data reflection. Live connections mode dynamically refreshes the semantic model with data changes in Azure Data Lake Storage as they occur, ensuring seamless near real-time analysis of the evolving COVID-19 situation.

Bottom of Form

Top of Form

Question 6: **Correct**

You've implemented a star schema for your hospital Power BI report. Several users frequently need to analyze diagnoses across different date ranges. To improve query performance for these specific queries, what additional design element should you consider?

A. Implement materialized views for frequently used date ranges.

B. Create a separate fact table for diagnoses with date-specific columns.

C. Denormalize dimension tables by including frequently used diagnosis details.

D. Implement partitioning strategies on the Admissions fact table based on date.

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

**(Correct)**

* **C**

**Explanation**

The correct answer is D. Implement partitioning strategies on the Admissions fact table based on date.

Here's why:

D addresses the specific need for improved query performance on date-based diagnoses analysis. Partitioning the Admissions fact table by date segments the data for faster retrieval when querying specific date ranges.

A: Materialized views can be helpful, but they require additional maintenance and might not be necessary if partitioning effectively addresses the performance needs.

B: Creating a separate fact table for diagnoses adds complexity and potentially redundancy, increasing storage and management overhead.

C: Denormalizing dimension tables could improve performance for some queries but might lead to data redundancy and inconsistency issues.

Explanation of wrong answers:

A: While materialized views can be helpful, they require additional maintenance and might not be necessary if partitioning effectively addresses the performance needs.

B: Creating a separate fact table for diagnoses adds complexity and potentially redundancy, increasing storage and management overhead.

C: Denormalizing dimension tables could improve performance for some queries but might lead to data redundancy and inconsistency issues.

Bottom of Form

Top of Form

Question 7: **Incorrect**

You're designing a data analytics environment in Microsoft Fabric for a pharmaceutical company with data residing in a mix of on-premises SQL Server databases and Azure cloud storage. They require secure access to both data sources for various analytics tasks, including data refresh for Power BI reports and ETL pipelines.

Which data gateway type would be the most suitable option for this scenario?

A. Cloud Gateway: Efficient for accessing Azure data sources but incompatible with on-premises data.

B. Hybrid Gateway: Enables access to both on-premises and cloud data but requires additional configuration.

C. Virtual Machine Gateway: Offers flexibility for custom configurations but comes with higher management overhead.

D. On-premises Gateway: Simple to set up for on-premises data but lacks cloud data access capabilities.

* **D**

**(Incorrect)**

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

**(Correct)**

**Explanation**

The correct answer is B. Hybrid Gateway. Here's why:

A Hybrid Gateway acts as a bridge between on-premises data sources and Fabric, allowing secure access to both SQL Server databases and Azure cloud storage.

It eliminates the need for separate gateways for on-premises and cloud data, simplifying your environment and management overhead.

Explanation of wrong answers:

A. Cloud Gateway is only suitable for accessing data within Azure and wouldn't reach on-premises SQL Server databases.

C. While Virtual Machine Gateways offer more flexibility, they require virtual machine setup and management, adding complexity.

D. On-premises Gateway wouldn't grant access to Azure cloud storage within the Fabric environment.

Bottom of Form

Top of Form

Question 8: **Incorrect**

Your company needs to ingest data from various web APIs, log files, and social media feeds into a Fabric data lakehouse. The data requires different transformations and cleansing steps based on its source. The entire process needs to be automated and easily maintainable.

Which of the following options is the BEST tool for orchestrating and managing this data ingestion pipeline?

A. Develop separate Azure Databricks notebooks for each data source, performing all transformations and ingesting data into the lakehouse.

B. Schedule multiple Azure Data Factory (ADF) pipelines with individual copy activities for each source, using Data Flows for simple transformations within each pipeline.

C. Design a single ADF pipeline with dedicated activities for each source, employing nested dataflows for specific source-related transformations.

D. Implement a Data Factory pipeline and utilize linked services to connect to each source, building a central dataflow encompassing all data sources and transformations.

* **B**

**(Incorrect)**

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

**(Correct)**

* **A**

**Explanation**

Answer: D

Explanation:

A: Separate notebooks for each source are difficult to maintain and lack centralized orchestration.

B: While multiple ADF pipelines are manageable, using separate copies and dataflows for each source might be too granular and cumbersome.

C: Nested dataflows within each pipeline offer some centralized control, but managing multiple dataflows within pipelines can still be complex.

D: A single ADF pipeline with linked services for each source and a central dataflow encompassing all data sources and transformations delivers the best balance of control, flexibility, and maintainability. This enables centralized orchestration, simplifies future modifications, and provides a clear overview of the entire data ingestion process.

Bottom of Form

Top of Form

Question 9: **Correct**

Your company uses a Fabric data lakehouse with various datasets stored in folders and Azure Data Lake Storage Gen2 (ADLS Gen2) accounts. Your team needs to frequently access specific files across different locations within the lakehouse for analysis.

What is the MOST efficient and secure way to organize and access these files?

A. Create individual Power BI reports pointing to each file and share them with the team.

B. Develop custom scripts to navigate and locate the desired files across different locations.

C. Organize the files into logical folders within the same ADLS Gen2 account and grant access to your team.

D. Create Fabric lakehouse shortcuts in a central location, pointing to the specific files across different folders and accounts.

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

**(Correct)**

**Explanation**

Answer: D

Explanation:

A: Building individual Power BI reports for each file is inefficient and creates clutter. It doesn't address centralized access and security management.

B: Custom scripts add complexity and might not be user-friendly for everyone. They also require ongoing maintenance.

C: Organizing files within the same ADLS Gen2 account might improve navigation, but it doesn't provide a unified point of access or control permissions for specific files across different locations.

D: Fabric lakehouse shortcuts offer the most efficient and secure approach. You can create central shortcuts pointing to any file, regardless of its location within the lakehouse (folders or even different ADLS Gen2 accounts). This simplifies access, improves discoverability, and allows granular permission management for your team.

Bottom of Form

Top of Form

Question 10: **Correct**

Your company needs to migrate data from a legacy on-premises SQL Server database to a new Fabric data lakehouse. The data consists of several large tables, including customer information, product details, and historical sales transactions.

Which method is the MOST efficient and scalable for copying this data to the lakehouse, considering cost and performance?

A. Use Bulk Copy Program (BCP) to directly transfer the data from SQL Server to the lakehouse landing zone, then utilize Azure Databricks notebooks for further processing.

B. Develop Azure Synapse Pipelines with dedicated copy activities for each table, configuring compression and partitioning for optimal storage and performance.

C. Leverage Azure Data Factory (ADF) pipelines with the Copy Data activity, employing staged copy for large tables and optimizing throughput for faster data transfer.

D. Implement Azure Data Share to seamlessly share the SQL Server database with the Fabric environment, allowing direct access to the data within the lakehouse.

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

**(Correct)**

* **B**

**Explanation**

Answer: C

Explanation:

A: While BCP is efficient for large data transfers, it requires additional steps for lakehouse processing within Databricks, adding complexity.

D: While Data Share simplifies access, it doesn't physically move the data to the lakehouse, impacting performance for analytics workloads.

C: Azure Data Factory offers a centralized and scalable solution for data movement. The Copy Data activity supports staged copy for large tables, optimizing storage and transfer speed. Additionally, ADF allows configuring compression and partitioning for efficient lakehouse storage and performance.

B: Synapse Pipelines can also handle data movement, but they might be less cost-effective for this scenario compared to ADF, which is specifically designed for data integration.

Bottom of Form

Top of Form

Question 11: **Incorrect**

Your company plans to ingest real-time data from social media feeds into a Fabric data lakehouse using Azure Event Hubs for data streaming. The data needs to be cleansed and aggregated before storage.

Which of the following options BEST suits this real-time data ingestion scenario?

A. Schedule an 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 aggregation before writing to the lakehouse.

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

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

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

**(Correct)**

* **C**

**(Incorrect)**

**Explanation**

Answer: B

Explanation:

A: Scheduled ADF pipelines are not suitable for real-time data ingestion. They're best for periodic batch processing.

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

C: Databricks streaming notebooks can handle real-time data, but managing and deploying notebooks separately from the data pipeline adds complexity.

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 aggregating all types of data.

Bottom of Form

Top of Form

Question 12: **Correct**

You are a data analyst for a large retail company that uses Microsoft Fabric. The company has recently launched a new product line and you have been tasked with analyzing the sales data for this product line. You decide to use descriptive analytics to summarize the main characteristics of this data.

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

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

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

**(Correct)**

**Explanation**

Time-series analysis

Explanation: Descriptive analytics is used to summarize and interpret data to extract meaningful insights. 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 sales data. The other options, while useful in other scenarios, are not the best fit for this specific task.

Bottom of Form

Top of Form

Question 13: **Correct**

Your company is migrating its data analytics workloads from on-premises servers to Microsoft Fabric. You need to ensure seamless integration with existing Power BI reports and dashboards.

Which settings in the Fabric admin portal would enable smooth Power BI integration and minimize downtime during migration?

A. Configure a dedicated Power BI Gateway within the Fabric workspace.

B. Enable Azure Active Directory (AAD) authentication for Fabric workspaces and Power BI Desktop.

C. Use the Fabric Data Catalog to discover and connect to existing data sources.

D. Deploy the Power BI service directly into the Fabric environment.

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

**(Correct)**

**Explanation**

The correct answer is B. Here's why:

AAD authentication provides a single sign-on experience for accessing both Fabric and Power BI resources, simplifying user management and reducing migration complexity.

Enabling AAD within Fabric workspaces ensures secure access to data and analytics assets for Power BI users.

Explanation of wrong answers:

A: While a dedicated Power BI Gateway can be used for on-premises data sources, it's not necessary for migrating existing reports within Fabric.

C: The Fabric Data Catalog is helpful for discovering data sources within the Fabric environment, but it doesn't directly address Power BI integration.

D: Deploying the Power BI service inside Fabric might not be necessary depending on your existing infrastructure and licensing. Additionally, it might lead to separate authentication and management overhead.

Bottom of Form

Top of Form

Question 14: **Incorrect**

Your company utilizes a Fabric data lakehouse and Azure Databricks notebooks for various analytics tasks. Some notebooks need to run automatically on a schedule to refresh datasets used in Power BI reports. These notebooks perform complex ETL logic but have dependencies on pre-processed data generated by another notebook.

How can you efficiently schedule and automate these dependent notebooks in the Databricks workspace?

A. Implement Azure Data Factory (ADF) with trigger schedules to call each notebook sequentially, ensuring the pre-processed data is available.

B. Leverage Databricks notebook dependencies within the workspace, scheduling the main notebook and letting it automatically trigger the necessary pre-processing notebook.

C. Utilize Azure Kubernetes Service (AKS) to containerize the notebooks and use Kubernetes cron jobs for scheduling independent executions.

D. Manually trigger the pre-processing notebook before running the main notebook, relying on a manual process for ensuring data availability.

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

**(Correct)**

* **A**

**(Incorrect)**

**Explanation**

Answer: B

Explanation:

A: Adding an ADF layer adds unnecessary complexity when Databricks already offers built-in scheduling capabilities.

B: Leveraging Databricks notebook dependencies is the most efficient solution. You can define the pre-processing notebook as a dependency for the main notebook within the workspace. This ensures the pre-processing runs automatically before the main notebook execution, guaranteeing data availability without manual intervention.

C: Using AKS adds unnecessary containerization overhead for these specific notebooks. Databricks' native scheduling features are well-suited for this scenario.

D: Manually triggering the pre-processing notebook is inefficient and error-prone. Databricks dependencies automate the process and ensure reliable data availability for the main notebook execution.

Bottom of Form

Top of Form

Question 15: **Incorrect**

Your company has a Fabric data warehouse containing sales data. You need to build a new dataset for campaign analysis by merging sales data with marketing campaign data from a separate Azure SQL Database.

Which of the following methods is the MOST secure and efficient way to achieve this data merging in Fabric?

A. Use Power Query in Power BI Desktop to connect to both data sources and perform the merge, then upload the result to the data warehouse.

B. Develop a custom Azure Data Factory pipeline that extracts data from both sources, joins them in Azure Synapse Analytics, and loads the merged data into the warehouse.

C. Create a linked server connection in Azure Synapse Analytics to the SQL Database and directly query both data sources within a Synapse SQL script to create the merged dataset.

D. Implement external tables in the Fabric data warehouse, referencing the SQL Database directly and allowing joins within queries for campaign analysis.

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

**(Correct)**

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

**(Incorrect)**

**Explanation**

Answer: B

Explanation:

A: Power Query works for small-scale scenarios, but it's not secure for production use and lacks scalability for large datasets.

B: Azure Data Factory provides a secure and scalable solution. The pipeline can extract data from both sources, orchestrate the join in Synapse Analytics, and load the merged data into the warehouse, ensuring data integrity and security.

C: Linked server connections can be convenient, but they bypass audit trails and security controls in Fabric. Direct querying within Synapse SQL exposes the SQL Database to potential unauthorized access.

D: External tables allow referencing external data sources, but they don't automatically perform joins. You'd still need to write additional queries within the warehouse to merge the data, increasing complexity and maintenance overhead.

Bottom of Form

Top of Form

Question 16: **Incorrect**

You are developing a customer churn prediction model in Power BI. You need to create complex calculated columns involving conditional logic and nested functions to calculate customer lifetime value and risk of churn.

Which tool would be most appropriate for efficiently authoring and testing these complex DAX functions and calculations?

A. Power BI Desktop, using the DAX formula bar and calculated column editor.

B. Tabular Editor 2, allowing code editing, syntax highlighting, and advanced testing features.

C. DAX Studio, providing real-time formula validation, performance profiling, and version control.

D. Azure Analysis Services, directly deploying the model with calculated columns defined in Power BI Desktop.

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

**(Correct)**

* **A**

**(Incorrect)**

**Explanation**

C. DAX Studio

Explanation:

Option A is sufficient for simple calculations, but lacks advanced features like real-time validation, profiling, and version control for complex expressions.

Option B offers editing and testing, but DAX Studio excels in real-time validation, performance analysis, and easier collaboration with version control.

Option D bypasses authoring tools, while directly deploying the model might hinder iterative development and testing of complex calculated columns.

Option C, DAX Studio, provides the ideal environment for authoring and testing complex DAX expressions. Real-time validation ensures syntactical accuracy, performance profiling identifies potential bottlenecks, and version control facilitates collaboration and experimentation with these crucial calculated columns for the churn prediction model.

Bottom of Form

Top of Form

Question 17: **Incorrect**

You're building a Power BI report for a university to analyze student enrollment and course registrations. The data model includes several tables:

Students (student\_id, name, major)

Courses (course\_id, title, department)

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

How should you establish the relationship between Students and Courses to accommodate students registering for multiple courses in different semesters?

A. Create a one-to-many relationship from Students to Courses.

B. Create a many-to-many relationship directly between Students and Courses.

C. Implement a bridge table named "Registrations" with student\_id, course\_id, and semester columns.

D. Denormalize the "Enrollments" table by adding a "major" column from the Students table.

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

**(Correct)**

* **D**

**A**

**(Incorrect)**

**Explanation**

The correct answer is C. Implement a bridge table named "Registrations" with student\_id, course\_id, and semester columns.

Here's why:

A: A one-to-many relationship wouldn't accurately represent students registering for multiple courses.

B: While a many-to-many relationship is technically possible, it's not the most efficient solution and can lead to performance issues.

C is the optimal solution because a "Registrations" bridge table explicitly captures the many-to-many relationship between Students and Courses, including the specific semester for each registration. This design ensures data integrity and allows for efficient querying and analysis.

D Denormalization might improve performance for some queries, but it introduces data redundancy and increases storage requirements, making it less maintainable in the long run.

Explanation of wrong answers:

A: A one-to-many relationship wouldn't represent students having multiple registrations.

B: While a many-to-many relationship is technically possible, it's not ideal due to potential performance issues.

D: Denormalization adds complexity and data redundancy, making it less ideal than a well-designed bridge table.

Bottom of Form

Top of Form

Question 18: **Incorrect**

Your company is building a data analytics environment in Microsoft Fabric for a global manufacturing organization. The environment needs to:

Ingest real-time production data from on-premises factory floors across multiple countries.

Process and analyze large datasets for insights into machine performance, production efficiency, and quality control.

Share reports and dashboards with authorized personnel at headquarters and regional offices.

Which data gateway type would be the most suitable for this scenario?

A. On-premises data gateway: Provides secure and reliable connection to on-premises data sources.

B. Cloud data gateway: Offers scalability and flexibility for cloud-based data sources.

C. Hybrid data gateway: Combines the functionality of on-premises and cloud gateways.

D. Personal gateway: Enables individual users to connect to data sources directly from their devices.

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

**(Correct)**

* **D**

**(Incorrect)**

**Explanation**

The correct answer is C. Hybrid data gateway. Here's why:

The scenario involves both on-premises (factory data) and potentially cloud-based data sources (e.g., ERP system).

The hybrid gateway provides a single point of entry for securely accessing and managing both types of data sources within the Fabric environment.

Real-time data ingestion and large data processing suggest the need for a robust and scalable solution like the hybrid gateway.

Explanation of wrong answers:

A. On-premises data gateway: While suitable for connecting to on-premises data, it wouldn't handle potential cloud data sources effectively.

B. Cloud data gateway: Useful for cloud-based data only, not ideal for on-premises factory floor data.

D. Personal gateway: Not designed for large-scale data processing or sharing reports, only individual user connections.

Bottom of Form

Top of Form

Question 19: **Correct**

You're designing a Fabric environment for a healthcare organization with multiple clinics and departments. They require unified data analytics across the organization while ensuring each department can customize reports with their branding and preferred visuals.

Which approach would best enable custom Power BI report themes while maintaining consistency across the organization?

A. Develop a single master theme in Fabric and enforce it across all reports.

B. Create individual themes for each department and allow full customization.

C. Design a base theme in Fabric with customizable elements for departments.

D. Implement Power BI Desktop theme packages for individual departments.

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

**(Correct)**

**Explanation**

The correct answer is C. Design a base theme in Fabric with customizable elements for departments. Here's why:

Balance consistency and customization: A base theme ensures core branding and visual elements are consistent, while allowing departments to adjust colors, fonts, and specific data visualizations.

Maintain Fabric integration: Fabric themes offer native integration with Fabric data sources and security, unlike individual theme packages.

Centralized control: A base theme in Fabric allows for easier updates and management compared to individual department themes.

Explanation of wrong answers:

A. Enforcing a single theme limits department customization and might not align with specific needs.

B. Individual themes can lead to inconsistent branding and visual language across the organization.

D. Power BI Desktop theme packages require manual installation and management, creating additional overhead.

Bottom of Form

Top of Form

Question 20: **Correct**

Your company's Fabric data lakehouse stores customer data from various sources, including website visits, purchase transactions, and social media interactions. You need to enrich this data with additional information like customer demographics and geographic location for improved analysis.

Which of the following approaches is MOST efficient and scalable for data enrichment in this scenario?

A. Develop custom Azure Functions to call external APIs for each missing data point and update the customer data directly in the lakehouse.

B. Implement Azure Data Factory (ADF) pipelines with copy activities to merge pre-populated demographic and location datasets with the customer data.

C. Create external tables referencing external data sources within the lakehouse, allowing seamless access to enriched data without physical duplication.

D. Use Spark notebooks within Azure Databricks to join the customer data with separate demographic and location datasets, storing the enriched data in a new table.

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

**(Correct)**

**Explanation**

Answer: C

Explanation:

A: Custom functions can be time-consuming to develop and manage, and calling external APIs for each customer might be inefficient.

B: While ADF pipelines are good for data movement, merging large datasets can be resource-intensive.

C: External tables provide the most efficient and scalable approach. They act as virtual views referencing data in external sources, eliminating physical duplication and allowing seamless integration with the existing customer data for analysis without impacting performance.

D: Databricks notebooks offer flexibility, but storing the enriched data in a new table adds another layer of complexity and management overhead. External tables offer a more lightweight and efficient solution.

Bottom of Form

Top of Form

Question 21: **Incorrect**

Your company wants to ingest data from various sources, including website logs, social media feeds, and internal applications, into a Fabric lakehouse for customer analytics. The data needs to be transformed, cleaned, and formatted before being stored.

Which approach is the MOST scalable and efficient for ingesting and processing this diverse data in Fabric?

A. Use Azure Data Factory to schedule individual data pipelines for each data source, with custom scripts and transformations in each pipeline.

B. Develop a single dataflow in Fabric that ingests all data sources, using built-in data connectors and transformations within the dataflow.

C. Write a custom notebook in Azure Synapse Analytics to process and format the data, then upload it manually to the lakehouse.

D. Set up continuous ingestion using Azure Event Hubs to send data to the lakehouse, and rely on downstream processing for cleaning and transformation.

* **B**

**(Correct)**

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

**(Incorrect)**

**Explanation**

Answer: B

Explanation:

A: While individual pipelines can work, managing and maintaining numerous pipelines for diverse data sources can be complex and inefficient.

B: A single dataflow in Fabric offers a centralized and scalable solution. Built-in connectors handle data ingestion from various sources, and data transformations can be applied within the dataflow itself, eliminating the need for custom scripts and simplifying maintenance.

C: Custom notebooks require development and maintenance overhead, and manual uploading to the lakehouse introduces another step and potential bottlenecks.

D: Continuous ingestion with Azure Event Hubs can be efficient for real-time data, but it bypasses necessary cleaning and transformation steps before storage.

Bottom of Form

Top of Form

Question 22: **Incorrect**

You're building a Power BI report for a retail chain analyzing sales trends across different regions and product categories. You want to identify the top 3 performing products in each region for the last quarter.

Which DAX calculation best achieves this using variables, iterators, and table filtering?

A.

**Code snippet**

1. TopProducts = VAR TopCount = 3,
2. FilteredData = FILTER(Sales, Sales[Date] >= TODAY() - 90),
3. TopPerRegion =
4. ADDCOLUMNS(
5. VAR CurrentRegion = CALCULATE(MIN(Sales[Region]))
6. RETURN
7. VAR TopProduct = CALCULATE(TOPCOUNT(FilteredData, Sales[ProductID], 3, DESC(Sales[SalesAmount])))
8. RETURN
9. { "Region": CurrentRegion, "TopProduct": TopProduct }
10. )
11. RETURN
12. TopPerRegion

B.

**Code snippet**

1. TopProducts = CALCULATE(
2. TOPN(3, Sales,
3. [Sales Amount], DESC,
4. [Region]))

C.

1. TopProducts = VAR TopCount = 3,
2. RegionList = DISTINCT(Sales[Region]),
3. TopPerRegion =
4. ADDCOLUMNS(RegionList,
5. { "TopProduct": CALCULATE(TOPCOUNT(Sales, Sales[ProductID], TopCount, DESC(Sales[SalesAmount]))) }
6. )
7. RETURN
8. TopPerRegion

D.

**Code snippet**

1. TopProducts = VAR TopCount = 3,
2. FilteredData = FILTER(Sales, Sales[Date] >= TODAY() - 90),
3. TopProductIDs = CALCULATE(TOPN(TopCount, FilteredData, Sales[ProductID], DESC(Sales[SalesAmount])),
4. RETURN
5. CALCULATETABLE(Sales, Sales[ProductID] IN TopProductIDs)

* **A**

**(Correct)**

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

**(Incorrect)**

* **D**

**Explanation**

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

Here's why:

A utilizes variables, iterators, and table filtering effectively:

TopCount variable defines the number of top products.

FilteredData filters sales data for the past quarter.

TopPerRegion iterates through each region, calculates the top 3 products with ADDCOLUMNS, and returns a table with region and top product information.

Nested VAR TopProduct uses another iterator within ADDCOLUMNS to calculate the top product for each region using TOPCOUNT with DESC order on sales amount.

B misses the regional breakdown. It simply identifies the top 3 products across the entire dataset, ignoring the regional requirement.

C iterates through regions but calculates the top product across the entire dataset within each iteration, not specific to the region.

D filters the sales data based on top product IDs calculated across the whole dataset, not considering regional context.

Explanation of wrong answers:

B: It doesn't consider region-specific analysis.

C: It calculates top products for all regions based on the entire dataset within each iteration.

D: It filters data based on top product IDs across the entire dataset, not considering regional context.

Bottom of Form

Top of Form

Question 23: **Incorrect**

Your company's data lakehouse stores sensor data from IoT devices in JSON files. The data is partitioned by device ID and timestamp, but queries analyzing data trends across all devices often experience slow performance.

How can you further optimize file partitioning for these cross-device trend analysis queries?

A. Implement additional partitioning on the data within each device ID partition based on a specific time range (e.g., hour, day).

B. Combine all device ID partitions into a single file to improve scan performance for large-scale queries.

C. Convert the JSON files to a columnar format like Parquet for better compression and data access.

D. Use Azure Synapse Analytics to create a materialized view of the data with pre-aggregated trends for faster analysis.

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

**(Correct)**

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

**(Incorrect)**

**Explanation**

Answer: A

Explanation:

A: Implementing additional partitioning based on a specific time range within each device ID partition allows for efficient filtering and scanning of data relevant to specific time periods for cross-device trend analysis.

B: Combining all partitions into a single file might improve scan performance for large-scale queries, but it would negate the benefits of device- and timestamp-based partitioning for smaller, targeted queries.

C: While converting to a columnar format like Parquet can improve overall data access, it doesn't address the specific need for optimized partitioning for cross-device trend analysis.

D: Using a materialized view in Synapse Analytics with pre-aggregated trends can be helpful for specific analysis scenarios, but it adds an additional layer of complexity and requires maintaining the materialized view alongside the raw data.

Bottom of Form

Top of Form

Question 24: **Correct**

Your company needs to build an Azure Data Factory (ADF) pipeline to ingest data from a CSV file, perform cleansing and transformation using Python code, and then load the processed data into a Fabric data warehouse. The Python code involves complex logic for data validation and anomaly detection.

Which of the following approaches should you use to integrate the Python code within the ADF pipeline?

A. Develop a custom Azure Function to run the Python code and call it as an external activity within the ADF pipeline.

B. Use an ADF Data Flow with a Python script activity, directly embedding the Python code within the Data Flow canvas.

C. Implement an Azure Databricks notebook containing the Python code and connect it to the ADF pipeline as a linked service.

D. Create a stored procedure in the data warehouse using Python code and call it from the ADF pipeline after loading the data.

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

**(Correct)**

**Explanation**

Answer: B. Use an ADF Data Flow with a Python script activity, directly embedding the Python code within the Data Flow canvas.

Explanation:

A: Custom Azure Functions are suitable for complex logic, but they add an extra layer of complexity and require separate deployment and management. ADF’s Data Flow with a Python script activity offers a more integrated and efficient solution.

C: Databricks notebooks provide flexibility but are overkill for this scenario. ADF’s Python script activity offers better integration within the pipeline.

D: Stored procedures are not suitable for Python code execution. They are primarily used for T-SQL within the data warehouse.

Bottom of Form

Top of Form

Question 25: **Correct**

You are a data analyst working for Contoso Pharmaceuticals, responsible for analyzing clinical trial data to predict patient response to a new cancer treatment. The initial analysis identified several key factors influencing treatment effectiveness. Now, you need to create a report for the research team that integrates both predictive and prescriptive analytics to optimize treatment decisions.

Which of the following approaches would BEST achieve this goal?

A. Develop a dashboard with interactive charts showcasing predicted treatment outcomes for different patient profiles.

B. Build a machine learning model that recommends the optimal treatment dosage based on patient characteristics and predicted response.

C. Create a static report summarizing the key factors influencing treatment response and highlighting potential risks.

D. Implement A/B testing to compare the new treatment with existing ones in a real-world setting.

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

**(Correct)**

**Explanation**

B. Build a machine learning model that recommends the optimal treatment dosage based on patient characteristics and predicted response.

Explanation:

Option A focuses solely on predictive analytics through visualizations, ignoring prescriptive recommendations.

Option C provides static summaries, lacking both prediction and prescription.

Option D involves real-world testing, valuable for long-term evaluation but not suitable for immediate actionable insights.

Option B directly integrates both aspects - predicting response through the model and prescribing optimized dosage based on that prediction, aligning perfectly with the scenario and exam objectives.

Bottom of Form

Top of Form

Question 26: **Incorrect**

You are working on a project where you need to diagnose why sales have been declining for a particular product over the past six months. You decide to use diagnostic analytics to find out the cause.

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. Anomaly detection

* **A**

**(Correct)**

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

**(Incorrect)**

**Explanation**

A. Decision tree analysis

Explanation: Diagnostic analytics is used to determine why something happened. In this scenario, decision tree analysis would be the most appropriate technique as it allows you to explore possible outcomes and the statistical likelihood of each outcome, which can help identify the factors contributing to the decline in 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 27: **Correct**

Your company needs to ingest data from various sources, including social media APIs, web logs, and customer surveys, into a new Fabric data lakehouse. The data needs to be transformed and structured before landing in the lake.

Which of the following approaches is MOST efficient and scalable for this scenario?

A. Develop custom Azure Functions code to handle data ingestion, transformation, and loading for each source individually.

B. Build a data pipeline in Azure Data Factory (ADF) with separate activities for each source, using Data Flows for transformations.

C. Create Azure Databricks notebooks for each data source, performing ingestion, transformation, and loading within the notebooks.

D. Use Azure Synapse Pipelines with dedicated copy activities for each source and script-based transformations in embedded notebooks.

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

**(Correct)**

**Explanation**

Answer: B

Explanation:

A: Developing custom Azure Functions for each source is inefficient and lacks scalability. Maintaining and monitoring many individual functions is cumbersome.

B: Azure Data Factory offers a centralized orchestration platform for data pipelines. Building a single pipeline with separate activities for each source leverages ADF's scalability and simplifies management. Data Flows provide a user-friendly interface for visual transformations within the pipeline.

C: While Databricks notebooks are powerful for complex transformations, managing individual notebooks for each source can be difficult, especially for large-scale data ingestion.

D: Azure Synapse Pipelines are suitable for simpler scenarios, and script-based transformations in notebooks might not be user-friendly for non-technical users. ADF's Data Flows offer a better balance between flexibility and ease of use.

Bottom of Form

Top of Form

Question 28: **Incorrect**

Your company has a Fabric data warehouse containing financial data from various ERP systems. Your team needs to create separate reports for different departments, each requiring access to specific subsets of data within the warehouse.

How can you effectively manage data access and security for these reports while preventing unnecessary exposure of sensitive financial information?

A. Grant each department read access to the entire data warehouse for their reports.

B. Create copies of the warehouse for each department with specific data subsets, but this leads to data redundancy and maintenance challenges.

C. Implement row-level security within the data warehouse, restricting access to specific rows based on departmental roles.

D. Use Fabric lakehouse shortcuts to point to filtered views of the data warehouse tailored for each department's report, without replicating data.

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

**(Correct)**

* **A**

**(Incorrect)**

**Explanation**

Answer: C

Explanation:

A: Granting full access to the entire warehouse is insecure and exposes sensitive data beyond what each department needs.

B: Data replication is inefficient and increases storage costs and maintenance overhead.

C: Row-level security within the data warehouse provides granular control, restricting access to specific data rows based on pre-defined rules and user roles. This ensures departments only see the relevant data for their reports.

D: While Fabric lakehouse shortcuts can simplify access, they don't address data security directly. Combining shortcuts with row-level security offers a comprehensive solution for access management and data protection.

Bottom of Form

Top of Form

Question 29: **Incorrect**

Your company needs to build an ADF pipeline to process data from a variety of sources, including social media APIs, web logs, and CRM systems. The pipeline involves different data transformations and validations for each source, with some requiring specific data partitioning before loading into a Fabric data lakehouse.

How can you best organize and manage the various data processing steps within the ADF pipeline?

A. Develop a single pipeline with separate activities for each data source, including Data Flows and copy activities for each transformation and loading step.

B. Create multiple pipelines, each dedicated to a specific data source, with its own set of Data Flows and copy activities for individual processing steps.

C. Utilize ADF's modularity with linked services and datasets, building reusable Data Flows for common transformations and calling them from different pipelines based on the source.

D. Implement Azure Data Lake Analytics (ADLA) jobs after copying data from each source, using them for all data transformations and partitioning before loading into the lakehouse.

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

**(Correct)**

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

**(Incorrect)**

**Explanation**

Answer: C

Explanation:

A: A single large pipeline with many activities can become complex and difficult to manage.

B: Multiple dedicated pipelines for each source can lead to redundancy and complicate maintenance.

C: Leveraging ADF's modularity with linked services and reusable Data Flows promotes code reuse and simplifies pipeline management. You can build common transformations as Data Flows and call them from different pipelines based on the source data, improving efficiency and maintainability.

D: Using ADLA jobs after data movement adds an additional layer of complexity and can be less efficient than directly integrating transformations within the ADF pipeline.

Bottom of Form

Top of Form

Question 30: **Correct**

Your company needs to copy real-time data from social media APIs into a Fabric data warehouse for near-instantaneous analytics. The data volume is expected to be high and fluctuate based on social media activity.

Which method is MOST suitable for handling this real-time data copy to the data warehouse?

A. Schedule regular Azure Data Factory pipelines with copy activities to pull data from the social media APIs at specific intervals.

B. Implement Azure Stream Analytics to continuously process the real-time data stream and write it directly into the data warehouse tables.

C. Utilize Azure Event Hubs for real-time data ingestion and trigger Azure Functions connected to the data warehouse for data transformation and loading.

D. Develop Azure Databricks notebooks with streaming capabilities to continuously ingest and process the social media data into the data warehouse.

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

**(Correct)**

**Explanation**

Answer: B

Explanation:

A: Scheduled ADF pipelines are not suitable for real-time data, as they introduce latency due to their interval-based execution.

B: Azure Stream Analytics offers a real-time data processing engine capable of continuously ingesting and transforming the data stream before loading it into the data warehouse tables, providing near-instantaneous analytics.

C: While Azure Functions and Event Hubs can handle real-time data, connecting them directly to the data warehouse might not be optimal for performance and scalability. Stream Analytics offers a more integrated solution for real-time data pipelines.

D: Databricks notebooks can handle real-time data, but managing and deploying them continuously might be complex compared to Stream Analytics, which is specifically designed for this purpose.

Bottom of Form

Top of Form

Question 31: **Correct**

You're building a Power BI report for a marketing agency analyzing campaign performance across different regions and channels. You want to create a calculation group that allows users to dynamically filter data based on selected regions and channels.

**Which DAX calculation group effectively achieves this using field parameters and dynamic strings?**

A.

**Code snippet**

1. Calculation Group =
2. {
3. "TotalClicks" = SUM(Marketing[Clicks]),
4. "FilteredClicks" = VAR RegionParam = SWITCH(TRUE(),
5. SELECTEDVALUE('Region'[Region]) = BLANK(), ALL('Region'[Region]),
6. SELECTEDVALUE('Region'[Region])),
7. ChannelParam = SWITCH(TRUE(),
8. SELECTEDVALUE('Channel'[Channel]) = BLANK(), ALL('Channel'[Channel]),
9. SELECTEDVALUE('Channel'[Channel])),
10. FilteredData = FILTER(Marketing, Marketing[Region] IN RegionParam, Marketing[Channel] IN ChannelParam)
11. RETURN SUM(FilteredData[Clicks])
12. }

B.

**Code snippet**

1. Calculation Group =
2. {
3. "TotalClicks" = SUM(Marketing[Clicks]),
4. "FilteredClicks" = CALCULATE(SUM(Marketing[Clicks]),
5. FILTER(Marketing, Marketing[Region] IN SELECTEDVALUE('Region'[Region]), Marketing[Channel] IN SELECTEDVALUE('Channel'[Channel])))
6. }

C.

**Code snippet**

1. Calculation Group =
2. {
3. "TotalClicks" = SUM(Marketing[Clicks]),
4. "FilteredClicks" = SUMX(Marketing, Marketing[Clicks], CONTAINS(CONCATENATE('Region'[Region],'-', 'Channel'[Channel]), CONCATENATE(SELECTEDVALUE('Region'[Region]),'-', SELECTEDVALUE('Channel'[Channel]))))
5. }

D.

**Code snippet**

1. Calculation Group =
2. {
3. "TotalClicks" = SUM(Marketing[Clicks]),
4. "FilteredClicks" = CALCULATE(SUM(Marketing[Clicks]), Marketing[Region] = SELECTEDVALUE('Region'[Region]), Marketing[Channel] = SELECTEDVALUE('Channel'[Channel]))
5. }

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

**(Correct)**

**Explanation**

The correct answer is **A. VAR RegionParam = SWITCH(TRUE()...**

Here's why:

* **A** uses field parameters for Region and Channel to capture user selections.
* It defines dynamic strings based on selected or all values using SWITCH and CONCATENATE.
* The FilteredData variable filters the data based on dynamic region and channel strings, enabling dynamic filtering based on user choices.
* **B** uses SELECTEDVALUE directly in CALCULATE, which only works if a single value is selected for each field, not allowing for multi-selection or filtering based on all values.
* **C** uses CONCATENATE and CONTAINS but doesn't handle multi-selection or filtering based on all values effectively.
* **D** uses SELECTEDVALUE directly in CALCULATE and doesn't allow for multi-selection or filtering based on all values.

**Explanation of wrong answers:**

* **B and D**: These options rely on SELECTEDVALUE directly in CALCULATE and don't handle multi-selection or filtering based on all values effectively.
* **C**: While CONCATENATE and CONTAINS are used, the logic doesn't accurately handle multi-selection or filtering based on all values.

Bottom of Form

Top of Form

Question 32: **Incorrect**

You are working for a hospital that wants to improve patient outcomes by predicting the risk of hospital-acquired infections (HAIs). You have access to patient medical records, lab results, and environmental data from the hospital wards.

How can you effectively integrate prescriptive and predictive analytics into a report to help the hospital reduce HAI risk?

A. Develop a predictive model to identify patients at high risk of HAIs based on their medical history and lab results. Create a report with recommendations for preventive measures tailored to each patient's risk profile.

B. Analyze environmental data to identify areas within the hospital with higher HAI rates. Generate a report with recommendations for improving hygiene protocols in those areas.

C. Combine the predictive model results with environmental data to identify specific risk factors for individual patients based on their location in the hospital. Create a report with targeted recommendations for reducing those risks.

D. Build a real-time dashboard that tracks HAI rates and environmental conditions across the hospital. Use the dashboard to identify trends and implement preventive measures proactively.

* **C**

**(Correct)**

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

**(Incorrect)**

**Explanation**

Answer: C

Explanation:

Option A only uses patient-level data for prediction, ignoring potential environmental factors.

Option B focuses on environmental data and recommendations without considering individual patient risk.

Option D provides real-time monitoring but lacks specific prescriptive recommendations for individual patients.

Option C combines patient-level predictions with environmental data to identify specific risk factors for each patient, enabling tailored recommendations for reducing their HAI risk. This effectively integrates prescriptive and predictive analytics into a report for targeted risk mitigation.

Bottom of Form

Top of Form

Question 33: **Incorrect**

Your company's Fabric data warehouse contains a large sales table with detailed product and customer information. Your team frequently needs to analyze sales trends by product category and region, but the queries are complex and time-consuming.

How can you improve query performance and simplify access to this data for your team?

A. Create a materialized view in the data warehouse pre-calculated with aggregated sales data by product category and region.

B. Develop custom Azure Functions to process the sales data and provide pre-aggregated results for different categories and regions.

C. Implement stored procedures in the data warehouse that encapsulate the complex sales trend queries for specific categories and regions.

D. Design a data pipeline that extracts and transforms the sales data into a separate data lakehouse table optimized for analytics.

* **A**

**(Correct)**

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

**(Incorrect)**

**Explanation**

Answer: A

Explanation:

A: A materialized view pre-calculates and stores frequently used aggregations, significantly improving query performance for common sales trend analysis. This avoids repetitive calculations and optimizes data access for your team.

B: While Azure Functions can offer flexibility, they add complexity and require ongoing maintenance. They're not ideal for frequent data access needs.

C: Stored procedures can simplify complex queries, but they don't improve performance as they still need to process the entire sales table on-demand.

D: While a data lakehouse might be helpful for some scenarios, it doesn't address the immediate performance and access issues within the existing data warehouse.

Bottom of Form

Top of Form

Question 34: **Incorrect**

You are tasked with analyzing customer churn data for RetailCo, a large online retailer. Your goal is to identify at-risk customers and recommend interventions to prevent churn.

Which combination of Azure services would be MOST effective in achieving this goal?

A. Azure Databricks + Azure Machine Learning

B. Azure Data Lake Storage + Power BI

C. Azure Synapse Analytics + Azure Functions

D. Azure SQL Database + Azure App Service

* **C**

**(Correct)**

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

**(Incorrect)**

**Explanation**

C. Azure Synapse Analytics + Azure Functions

Explanation:

Option A focuses on building and deploying machine learning models, lacking real-time analysis and intervention.

Option B excels in data storage and visualization but lacks real-time processing and automated actions.

Option D is suitable for simple database storage and web applications but lacks advanced analytics and real-time capabilities.

Option C combines the strengths of both services - Azure Synapse Analytics provides scalable data warehousing and real-time analytics, while Azure Functions enable automated triggers and interventions based on churn predictions, making it the most comprehensive solution for the scenario.

Bottom of Form

Top of Form

Question 35: **Correct**

Your company uses Azure Data Factory (ADF) to orchestrate data pipelines within a Fabric data lakehouse. One pipeline involves a dataflow with complex transformations requiring frequent adjustments based on new business requirements. The pipeline runs hourly to refresh reports in Power BI.

How can you ensure efficient scheduling and management of updates to this dataflow within the ADF pipeline?

A. Modify the ADF pipeline directly to include the updated dataflow code whenever changes are made.

B. Leverage Azure DevOps to manage the dataflow as a separate versioned asset, deploying updated versions to the ADF pipeline.

C. Deploy the dataflow to an Azure Databricks workspace and trigger its execution within the ADF pipeline using a web activity.

D. Schedule separate ADF pipelines for each version of the dataflow, manually choosing the active pipeline based on business needs.

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

**(Correct)**

**Explanation**

Answer: B

Explanation:

A: Directly modifying the ADF pipeline for dataflow updates is inefficient and error-prone. It lacks version control and makes managing changes difficult.

B: Utilizing Azure DevOps to manage the dataflow as a separate versioned asset is the most efficient and recommended approach. You can track changes, deploy different versions to the ADF pipeline with controlled rollouts, and revert to previous versions if needed. This ensures better management and smoother updates without impacting the running pipeline.

C: While Databricks offers flexibility, deploying the dataflow there creates unnecessary complexity for integration with the existing ADF pipeline. Azure DevOps provides seamless version control and deployment within the Fabric ecosystem.

D: Manually managing multiple ADF pipelines based on versions is cumbersome and inefficient. Azure DevOps offers a centralized and controlled approach to versioning and deployment.

Bottom of Form

Top of Form

Question 36: **Incorrect**

You're investigating website traffic data in a Fabric application to diagnose a sudden decrease in conversions. Users suspect changes to the checkout process are responsible.

Which sequence of data exploration steps would be most effective in identifying the root cause of the conversion decline?

A. Analyze overall conversion rates before and after the checkout change, then drill down into specific product categories and user types.

B. Compare website traffic patterns before and after the change, focusing on engagement metrics like bounce rates and time on page.

C. Calculate purchase completion rates for different checkout steps before and after the change, pinpointing where drop-offs occur.

D. Identify user segments with the highest drop-off rates after the change, then analyze their purchase history and website interactions.

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

**(Correct)**

* **D**

**(Correct)**

* **B**

**(Incorrect)**

**Explanation**

Answer: C & D

Explanation:

C and D offer a stepwise approach to diagnose the conversion decline:

C (purchase completion rates) identifies which specific checkout steps experience drop-offs after the change, narrowing down the potential problem area.

D (user segments with high drop-off rates) analyzes specific user groups most affected, revealing possible correlations with purchase history or website interactions.

A provides initial high-level insights but lacks diagnostic depth:

Comparing overall conversion rates is helpful, but it doesn't pinpoint the exact step in the checkout process causing the decline.

B focuses on broader website behavior, not checkout specifics:

Analysing traffic patterns is valuable for initial context, but it doesn't directly diagnose how the checkout change impacts conversions.

Bottom of Form

Top of Form

Question 37: **Incorrect**

You are a data analyst for Contoso, a retail chain, tasked with improving the efficiency of their delivery operations. You have access to historical delivery data, including order details, customer demographics, and weather patterns. Your goal is to create a report that helps Contoso optimize delivery routes and reduce costs.

Which of the following options best integrates prescriptive and predictive analytics into a visual or report to achieve this goal?

A. Create a map visualization that displays historical delivery routes overlaid with weather forecasts.

B. Develop a machine learning model to predict delivery times based on order size, distance, and weather conditions. Integrate the model predictions into the map visualization to highlight high-risk routes.

C. Analyze the data to identify factors that contribute to delivery delays, such as specific weather events or peak traffic times. Create a report with recommendations for alternative routes or delivery time windows based on these insights.

D. Build a dashboard that displays real-time delivery progress alongside weather updates and traffic conditions. Use the dashboard to dynamically adjust delivery routes as needed.

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

**(Correct)**

* **C**

**(Incorrect)**

* **D**

**Explanation**

Answer: B

Explanation:

Option A only displays historical data and weather forecasts, not integrating predictive analytics.

Option C provides recommendations based on historical data analysis, but lacks real-time predictions.

Option D focuses on real-time monitoring but doesn't directly integrate prescriptive analytics for route optimization.

Option B combines historical data analysis (machine learning model) with real-time data (weather forecasts) to predict delivery times and highlight high-risk routes on a map, effectively integrating prescriptive and predictive analytics into a visual report.

Bottom of Form

Top of Form

Question 38: **Incorrect**

Your company's Fabric data lakehouse stores customer data from various sources, including CRM systems and loyalty programs. You need to develop a data quality check process to ensure data consistency and identify potential errors before it's used for further analysis.

Which approach is MOST effective for implementing data quality checks in this scenario?

A. Develop custom scripts within Azure Data Factory (ADF) to validate and clean the customer data in the lakehouse.

B. Create stored procedures in Azure Synapse Analytics to perform data quality checks against predefined rules and flags suspicious records.

C. Utilize Azure Databricks notebooks for data exploration and custom data quality checks tailored to specific data sources.

D. Implement data validation rules directly in the data pipelines ingesting the customer data into the lakehouse.

* **C**

**(Incorrect)**

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

**(Correct)**

**Explanation**

Answer: D

Explanation:

D: Integrating data validation rules within the data pipelines offers the most efficient and scalable approach. This catches data quality issues at the source, preventing them from propagating further into the lakehouse and potentially impacting downstream analytics.

A: While custom ADF scripts can be effective, they add complexity and require ongoing maintenance.

B: Stored procedures in Synapse Analytics can perform checks on existing data, but they don't prevent errors from entering the lakehouse in the first place.

C: Databricks notebooks provide flexibility, but managing and deploying them for continuous data quality checks adds overhead compared to pipeline-level validation.

Bottom of Form

Top of Form

Question 39: **Incorrect**

Your company stores clickstream data in a Fabric data lakehouse for analyzing website traffic and user behavior. The data volume is significant and grows daily. Your analytics queries often focus on specific date ranges.

How can you implement file partitioning in the lakehouse to optimize query performance and cost?

A. Partition the clickstream data files by user ID, as user behavior analysis is the primary focus.

B. Partition the data by file size to ensure even distribution across storage containers.

C. Split the data into monthly partitions based on the timestamp within each clickstream record.

D. Create daily partitions for finer granularity, even if query focus is typically on broader date ranges

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

**(Correct)**

* **D**

**(Incorrect)**

* **B**

**Explanation**

Answer: C

Explanation:

A: While user ID might be relevant for some analysis, partitioning based on timestamp directly aligns with your current query focus on specific date ranges. This improves query efficiency by quickly filtering to relevant partitions without scanning the entire dataset.

B: File size partitioning doesn't offer any optimization for date-based queries and might not evenly distribute data across storage containers due to varying record sizes.

D: While daily partitions offer even finer granularity, they might add unnecessary overhead for queries spanning broader date ranges. Monthly partitions provide a good balance between granularity and performance optimization.

C: Monthly partitions based on timestamps represent the best choice for optimizing queries focused on specific date ranges within the large clickstream dataset. This reduces scanning overhead and improves query performance while maintaining enough granularity for potential future analysis needs.

Bottom of Form

Top of Form

Question 40: **Incorrect**

Your company needs to schedule an ADF pipeline to process customer orders data from an on-premises SQL Server database every night. The pipeline needs to perform data cleansing, transformation, and then load the data into a Fabric data lakehouse. However, the on-premises database is only available during specific business hours.

Which scheduling and execution option within ADF is MOST suitable for this scenario?

A. Schedule the ADF pipeline to run hourly throughout the night, with retry logic to handle any failures due to database unavailability.

B. Implement a trigger-based pipeline triggered by a custom on-premises application that detects database availability before starting the data processing.

C. Utilize ADF Managed Pipelines with a schedule set for the desired time window when the database is available, eliminating the need for retry logic.

D. Set up an Azure Data Factory Integration Runtime (IR) on-premises to directly access the database and schedule the pipeline execution from within the IR.

* **C**

**(Correct)**

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

**(Incorrect)**

**Explanation**

Answer: C

Explanation:

A: Hourly schedules with retry logic can be resource-intensive and unnecessary if the database is unavailable for most of the night.

B: Trigger-based approach adds complexity and requires maintaining a separate application to detect database availability.

C: ADF Managed Pipelines offer scheduled execution within a specific window. By setting the schedule for the business hours when the database is available, the pipeline will automatically execute only during those times, eliminating unnecessary attempts and optimizing resource utilization.

D: While on-premises IR offers flexibility, it requires additional setup and maintenance compared to the built-in scheduling capabilities of ADF Managed Pipelines.

Bottom of Form

Top of Form

Question 41: **Correct**

You're analyzing sales data in a Fabric application to understand what factors contribute to high return rates. Users need insights into specific product categories and regions.

Which combination of descriptive and diagnostic analytics techniques would be most effective for this scenario?

A. Create bar charts of return rates by product category and region, then use drill-down filters to identify specific products with high return rates.

B. Calculate churn rates and customer lifetime values by region, then use heatmaps to visualize correlations between demographics and return rates.

C. Develop linear regression models for each region, predicting return rates based on product features and customer attributes.

D. Build decision trees to classify customers with high return rates, identifying key purchase patterns and demographic characteristics.

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

**(Correct)**

* **A**

**(Correct)**

**Explanation**

Answer: A & D

Explanation:

A and D address both descriptive and diagnostic aspects:

A (bar charts and drill-down) provides initial descriptive insights into return rates by category and region, allowing users to identify potential areas for further analysis.

D (decision trees) offer diagnostic capabilities by classifying high-return customers, revealing key influencing factors like purchase patterns and demographics.

B focuses primarily on descriptive analysis:

Churn rates and customer lifetime values provide valuable overall insights, but they don't directly reveal the specific reasons behind high return rates.

C is primarily predictive, not diagnostic:

Linear regression models can predict future return rates based on various factors, but they don't explain why current return rates are high in specific contexts.

Bottom of Form

Top of Form

Question 42: **Incorrect**

Your company needs to move data from a relational database in Azure SQL Database to a Fabric data warehouse. The data requires complex transformations, aggregation, and filtering before landing in the warehouse. The transformations need to be easily testable and reusable across different pipelines.

Which of the following approaches is MOST suitable for achieving this data movement and transformation?

A. Schedule an Azure Data Factory (ADF) pipeline with a copy activity directly pointing to the database and the data warehouse table, applying transformation scripts within the copy activity.

B. Utilize Azure Databricks notebooks to read the data from the database, perform the required transformations, and write it to the data warehouse.

C. Develop Azure Functions triggered by a timer event, reading data from the database, transforming it, and then loading it into the data warehouse with the Fabric connector.

D. Design an ADF pipeline with a copy activity connected to the database. Use an external transformation activity like Azure Data Lake Analytics or Databricks to process the data before loading it into the warehouse.

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

**(Correct)**

* **A**

**(Incorrect)**

**Explanation**

Answer: D

Explanation:

A: Applying transformations within the copy activity is limited and not suitable for complex scenarios.

B: Databricks notebooks offer flexibility, but managing and deploying them adds complexity.

C: Timer-triggered functions are not ideal for scheduled data movement.

D: ADF's flexibility shines here. The copy activity can efficiently move the data, while connecting to an external transformation activity like Data Lake Analytics or Databricks allows for complex, reusable transformations with dedicated tools and testing capabilities. This combines efficient data movement with powerful transformation capabilities.

Bottom of Form

Top of Form

Question 43: **Incorrect**

You're building a Power BI report for a hospital to analyze patient admissions, diagnoses, and treatments. The data includes several tables:

Patients (patient\_id, name, dob, address)

Admissions (admission\_id, patient\_id, date\_admitted, date\_discharged)

Diagnoses (diagnosis\_id, admission\_id, diagnosis\_code, diagnosis\_description)

Procedures (procedure\_id, admission\_id, procedure\_code, procedure\_description)

Medications (medication\_id, admission\_id, medication\_name, dosage)

Which star schema design would be MOST efficient for querying patient admissions and associated details in this scenario?

A. Fact table: Admissions. Dimension tables: Patients, Diagnoses, Procedures, Medications.

B. Fact table: Diagnoses. Dimension tables: Patients, Admissions, Procedures, Medications.

C. Fact table: Admissions (including all diagnosis and procedure details). Dimension table: Patients.

D. Fact table: Patients. Dimension tables: Admissions, Diagnoses, Procedures, Medications.

* **A**

**(Correct)**

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

**(Incorrect)**

**Explanation**

The correct answer is A. Fact table: Admissions. Dimension tables: Patients, Diagnoses, Procedures, Medications.

Here's why:

A follows the classic star schema pattern with a central fact table containing transactional data (admissions) and surrounding dimension tables with descriptive attributes (patients, diagnoses, procedures, medications). This design optimizes query performance for analyzing patient admissions and linked details.

B uses Diagnoses as the fact table, which would lead to repetitive data for each diagnosis per admission, impacting query performance and storage efficiency.

C combines all details (diagnoses, procedures) into the Admissions fact table, making it overly complex and hindering efficient queries on specific dimensions.

D places Patients as the fact table, leading to data redundancy and inefficient queries for analyzing admissions-related information.

Explanation of wrong answers:

B: Placing Diagnoses as the fact table leads to data redundancy and inefficient queries for analyzing admissions-related information.

C: Combining all details into the Admissions fact table creates a complex structure and hinders efficient queries on specific dimensions.

D: Placing Patients as the fact table is not suitable for analyzing admissions-related details and leads to data redundancy.

Bottom of Form

Top of Form

Question 44: **Correct**

Your company needs to cleanse customer data stored in a Fabric data lakehouse before analysis. The data contains inconsistencies like missing address values, invalid postal codes, and duplicate records.

Which Azure service and approach would be most efficient for implementing a data cleansing process in this scenario?

A. Develop Azure Data Factory (ADF) pipelines with Data Flows for handling all cleansing tasks, including missing values imputation, normalization, and deduplication.

B. Utilize Azure Databricks notebooks with custom Python scripts to perform complex data validation, string manipulation, and outlier detection for cleansing.

C. Implement Data Quality Services (DQS) within SQL Server Integration Services (SSIS) to profile the data, identify inconsistencies, and suggest cleansing rules for execution.

D. Leverage Azure Machine Learning Studio to train a machine learning model for anomaly detection and data correction, applying it to the customer data for automated cleansing.

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

**(Correct)**

**Explanation**

Answer: A

Explanation:

A: Azure Data Factory with Data Flows offers a centralized and efficient solution for data cleansing within the Fabric ecosystem. Data Flows provide built-in capabilities for handling missing values, data normalization, and deduplication with visual tools, eliminating the need for complex scripting.

B: Databricks notebooks offer flexibility for custom algorithms, but managing and deploying multiple notebooks for specific cleansing tasks can be cumbersome. Data Flows provide a readily available and efficient solution for common cleansing scenarios.

C: DQS within SSIS is primarily designed for on-premises data and might not be the best fit for Fabric data lakehouse. ADF with Data Flows offers a more native and integrated approach.

D: While machine learning can be valuable for complex outlier detection, training and maintaining a model solely for customer data cleansing might be overcomplicated. Data Flows offer sufficient capabilities for this scenario.

Bottom of Form

Top of Form

Question 45: **Incorrect**

You are a Fabric Data Architect tasked with designing a data analytics environment for a large retail chain with over 1000 stores globally. They require near real-time analytics on sales, inventory, and customer behavior to optimize operations. The data sources include:

Transaction data from point-of-sale systems (high volume, frequent updates)

Inventory data from warehouse management systems (moderate volume, daily updates)

Customer loyalty program data (medium volume, weekly updates)

Social media sentiment data (low volume, streaming)

Which Fabric components and features are essential for this solution, considering both performance and cost efficiency?

Choices:

A. Data Lakehouse (Standard SKU) and Dataflows (Gen2)

B. Data Warehouse (Premium SKU) and Notebooks (Standard SKU)

C. Data Lakehouse (Premium SKU) and Semantic Models (Standard SKU)

D. Data Warehouse (Standard SKU) and Azure Synapse Analytics (Standard SKU)

* **A**

**(Correct)**

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

**(Incorrect)**

**Explanation**

A. Data Lakehouse (Standard SKU) and Dataflows (Gen2)

Explanation:

A Data Lakehouse (Standard SKU) provides a scalable and cost-effective storage solution for all data sources, including high-volume transaction data and streaming social media data.

Dataflows (Gen2) enable efficient data ingestion, transformation, and orchestration without requiring expensive pre-provisioning. They can handle both real-time and batch processing, making them ideal for near real-time analytics.

A Data Warehouse is more suited for structured data and historical analysis, not ideal for the variety and volume of data in this scenario.

Notebooks offer flexibility but are not as scalable or cost-effective for large-scale data processing.

Semantic Models are best for pre-aggregated data and visualization, not essential for the initial data pipeline and analysis.

Azure Synapse Analytics adds unnecessary complexity and cost when a simpler Fabric solution can suffice.

Incorrect Answer Explanations:

B: While a Data Warehouse can handle structured data efficiently, it's not ideal for the high-volume and real-time requirements of this scenario. Notebooks are useful for exploration, but not a core component of the data pipeline.

C: A Premium Data Lakehouse offers additional features but isn't necessary for this scenario. Semantic Models are beneficial for visualization but not essential for initial data processing.

D: Azure Synapse Analytics overlaps with Fabric components and adds unnecessary cost for this scenario.

Bottom of Form

Top of Form

Question 46: **Incorrect**

You are building a large-scale financial analysis model in Microsoft Fabric using historical transaction data stored in Azure Data Lake Storage Gen2. The model requires complex calculations and aggregation, but storage cost optimization is also a significant concern.

Which storage mode for the semantic model would best balance efficient data processing for complex calculations with cost-effective storage for large datasets?

A. Import mode, pre-aggregating and storing all data in the model for fast calculations.

B. DirectQuery mode, avoiding data redundancy in the model but potentially impacting query performance.

C. Incremental refresh mode, only updating relevant data portions, keeping storage and refresh costs in check.

D. Hybrid mode, combining DirectQuery for frequently accessed data with import mode for infrequently accessed calculations.

* **D**

**(Correct)**

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

**(Incorrect)**

**Explanation**

D. Hybrid mode, combining DirectQuery for frequently accessed data with import mode for infrequently accessed calculations.

Explanation:

Option A can be fast but incurs high storage costs for the entire dataset.

Option B reduces storage but might lead to slow queries for complex calculations on large data volumes.

Option C focuses on cost but doesn't address efficient processing for complex calculations.

Option D provides the best balance. By using DirectQuery for frequently accessed data (e.g., daily analysis) and importing infrequently used data for complex calculations (e.g., historical trends), you optimize performance for frequently used features while minimizing storage costs for less-accessed calculations.

Bottom of Form

Top of Form

Question 47: **Incorrect**

You've implemented the "Registrations" bridge table in your university data model. Some users need to analyze course registrations by department. However, the Courses table doesn't directly contain department information. How should you address this scenario?\*\*

A. Create a new calculated column in the "Registrations" table to join with the department from the Courses table.

B. Add a "department" column to the "Enrollments" table by merging data from the Courses table.

C. Modify the relationship between "Registrations" and Courses to include a join on the department from the Courses table.

D. Create a separate dimension table for departments and link it to the Courses table.

* **A**

**(Correct)**

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

**(Incorrect)**

**Explanation**

The correct answer is A. Create a new calculated column in the "Registrations" table to join with the department from the Courses table.

Here's why:

A is the most efficient and maintainable solution. By creating a calculated column in "Registrations" to join with the department from Courses, you leverage existing relationships without modifying the data structure or creating redundancies.

B Denormalizing "Enrollments" with department information would introduce redundancy and complicate data updates.

C Modifying the relationship wouldn't automatically bring department information into "Registrations" and might introduce performance issues.

D Creating a separate department dimension table might be overkill for this specific scenario if the only need is to access department information for course registrations.

Explanation of wrong answers:

B: Denormalization adds complexity and data redundancy, making it less ideal than a calculated column.

C: Modifying the relationship might not achieve the desired outcome and could impact performance.

D: Creating a separate department dimension table is unnecessary for this specific scenario.

Bottom of Form

Top of Form

Question 48: **Incorrect**

Your company's Fabric data warehouse contains sales data for different products. You need to add a new column to this table calculating the profit margin for each product based on its cost and selling price.

Which of the following options is the BEST way to implement this data enrichment within the data warehouse?

A. Update the existing sales table by directly adding a new column with the calculated profit margin formula embedded.

B. Create a new view based on the sales table, adding the calculated profit margin as a virtual column using the formula.

C. Implement a user-defined function (UDF) encapsulating the profit margin calculation, then use it to add the new column to the sales table.

D. Design a separate table to store the calculated profit margins for each product, linked to the sales table via foreign keys.

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

**(Correct)**

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

**(Incorrect)**

**Explanation**

Answer: B

Explanation:

A: Modifying the existing table directly is risky and might impact historical data or performance.

B: Views offer a non-intrusive approach. You can add the calculated profit margin as a virtual column using the formula within the view, without modifying the original sales table. This preserves the original data while providing easy access to the enriched information.

C: While UDFs are suitable for reusable calculations, adding a new column directly to the table might not be necessary. Views provide a simpler and less intrusive solution in this case.

D: Creating a separate table adds unnecessary complexity and storage overhead. A view based on the existing sales table with the calculated profit margin is sufficient.

Bottom of Form

Top of Form

Question 49: **Incorrect**

Your company imports product sales data from various sources into a Fabric data warehouse. Some data sources contain incompatible currency formats and units of measurement. You need to transform the data into a consistent format for accurate analysis.

Which data transformation techniques are most suitable for handling currency and unit conversions in this scenario?

A. Utilize currency exchange APIs within Azure Functions triggered by data arrival to convert individual values before loading into the data warehouse.

B. Implement case statements within stored procedures in the data warehouse to dynamically convert currency and units based on source and product information.

C. Apply a mapping transformation within an ADF dataflow to match currency and unit codes with conversion rates and factors stored in separate reference tables.

D. Develop Azure Databricks notebooks with custom Python libraries for complex unit conversions and calculations depending on product categories and data provenance.

* **C**

**(Correct)**

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

**(Incorrect)**

**Explanation**

Answer: C

Explanation:

C: Utilizing a mapping transformation within an ADF dataflow offers the most efficient and centralized approach. You can leverage built-in or custom mapping functions to match currency and unit codes with conversion rates and factors stored in separate reference tables, achieving consistent data format without complex scripting.

A: Triggering Azure Functions for individual value conversions can be resource-intensive and create unnecessary latency. Mapping transformations within ADF handle bulk data efficiently.

B: Case statements within stored procedures can be cumbersome to maintain for multiple currencies and units. ADF's mapping tools offer a more manageable and reusable solution.

D: Databricks notebooks offer flexibility for complex algorithms, but they might be overkill for simple currency and unit conversions. ADF's dataflows provide a readily available and optimized solution for this scenario.

Bottom of Form

Top of Form

Question 50: **Incorrect**

You're tasked with designing a large format dataset for a research institute analyzing global climate change data. The data includes various sources like satellite imagery, weather station measurements, and scientific publications.

Which of the following approaches would be MOST effective for data ingestion and transformation in this scenario?

A. Develop a custom ETL pipeline using Python and pandas.

B. Utilize Azure Data Factory with built-in connectors for various data sources.

C. Leverage Azure Databricks for scalable data processing and transformation.

D. Use Power BI Desktop for data import and transformation.

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

**(Correct)**

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

**(Incorrect)**

**Explanation**

The correct answer is C. Leverage Azure Databricks for scalable data processing and transformation.

Here's why:

C addresses the key challenges of large format datasets: scalability, parallel processing, and complex transformations. Azure Databricks offers a cloud-based platform for Spark, enabling efficient data processing and transformation at scale.

A while flexible, might not be scalable and efficient for large datasets.

B while convenient for smaller datasets, Azure Data Factory might not be optimal for complex transformations and high-volume data ingestion.

D Power BI Desktop is primarily for data visualization and analysis, not designed for large-scale data processing and transformation.

Explanation of wrong answers:

A: While Python and pandas offer flexibility, they might not be scalable for large datasets and require additional development effort.

B: Azure Data Factory is excellent for data orchestration and data movement, but it might not be ideal for complex data transformation tasks.

D: Power BI Desktop is a powerful tool for analysis and visualization, but it's not designed for large-scale data processing and transformation.

Bottom of Form