Microsoft 70-452

PRO: Designing a Business Intelligence Infrastructure Using Microsoft SQL Server 2008

Version: 33.0

QUESTION NO: 1

You design a Business Intelligence (BI) solution by using SQL Server 2008.

You plan to create a SQL Server 2008 Reporting Services (SSRS) solution that contains five sales dashboard reports.

Users must be able to manipulate the reports' parameters to analyze data.

You need to ensure that the following requirements are met:

- Users can manipulate the parameters for data analysis in a single trip to the data source.
- Reports are automatically rendered as soon as they are accessed for the first time.

Which two tasks should you perform?

(Each correct answer presents part of the solution. Choose two.)

A. Filter data by using expressions.
B. Specify the default values for each parameter.
C. Create an available values list for each parameter.
D. Create report parameters by using query parameters to filter data at the data source.

Answer: A,B
Explanation:

QUESTION NO: 2

You design a SQL Server 2008 Reporting Services (SSRS) solution. You create a report by using Microsoft Visual Studio .NET 2008.

The report contains the following components:

- A dataset named Customer that lists all active customers and their details. The dataset accepts no parameters.
- A dataset named SalesHistory that lists all sales transactions for a specified time period and accepts year and month as parameters.

You need to ensure that a summary of sales transactions is displayed for each customer after the customer details.

Which component should you add to the report?

A. List
B. Table
C. Matrix
D. Subreport

Answer: D
Explanation:

http://msdn.microsoft.com/en-us/library/ms160348(SQL.100).aspx
How to: Add a Subreport and Parameters (Reporting Services)
Add subreports to a report when you want to create a main report that is a container for multiple related reports. A subreport is a reference to another report. To relate the reports through data values (for example, to have multiple reports show data for the same customer), you must design a parameterized report (for example, a report that shows the details for a specific customer) as the subreport. When you add a subreport to the main report, you can specify parameters to pass to the subreport.
You can also add subreports to dynamic rows or columns in a table or matrix. When the main report is processed, the subreport is processed for each row. In this case, consider whether you can achieve the desired effect by using data regions or nested data regions.

QUESTION NO: 3

You design a Business Intelligence (BI) solution by using SQL Server 2008.

The solution includes a SQL Server 2008 Analysis Services (SSAS) database. The database contains a data mining structure that uses a SQL Server 2008 table as a data source. A table named OrderDetails contains detailed information on product sales. The OrderDetails table includes a column named Markup.

You build a data mining model by using the Microsoft Decision Trees algorithm. You classify Markup as discretized content.

The algorithm produces a large number of branches for Markup and results in low confidence ratings on predictable columns.

You need to verify whether the Markup values include inaccurate data.

What should you do?

A. Modify the content type of Markup as Continuous.

_____

B. Create a data mining dimension in the SSAS database from OrderDetails.

C. Create a data profile by using SQL Server 2008 Integration Services (SSIS).

D. Create a cube in SSAS. Use OrderDetails as a measure group. Recreate the data mining structure and mining model from the cube data.

Answer: C

Explanation:

Discretized The column has continuous values that are grouped into buckets. Each bucket is considered to have a specific order and to contain discrete values. Possible values for discretization method are automatic, equal areas, or clusters. Automatic means that SSAS determines which method to use. Equal areas results in the input data being divided into partitions of equal size. This method works best with data with regularly distributedvalues. Clusters means that SSAS samples the data to produce a result that accounts for "clumps" of data values. Because of this sampling, Clusters can be used only with numeric input columns. You can use the date, double, long, or text data type with the Discretized content type.

Microsoft Decision Trees Algorithm

Microsoft Decision Trees is probably the most commonly used algorithm, in part because of its flexibility— decision trees work with both discrete and continuous attributes—and also because of the richness of its included viewers. It's quite easy to understand the output via these viewers. This algorithm is used to both view and to predict. It is also used (usually in conjunction with the Microsoft Clustering algorithm) to find deviant values. The Microsoft Decision Trees algorithm processes input data by splitting it into recursive (related) subsets. In the default viewer, the output is shown as a recursive tree structure. If you are using discrete data, the algorithm identifies the particular inputs that are most closely correlated with particular predictable values, producing a result that shows which columns are most strongly predictive of a selected attribute. If you are using continuous data, the algorithm uses standard linear regression to determine where the splits in the decision tree occur. Clicking a node displays detailed information in the Mining Legend window. You can configure the view using the various drop-down lists at the top of the viewer, such as Tree,

Default Expansion, and so on. Finally, if you've enabled drillthrough on your model, you can display the drillthrough information—either columns from the model or (new to SQL Server 2008) columns from the mining structure, whether or not they are included in this model.

Data Profiling

The control flow Data Profiling task relates to business problems that are particularly prominent in BI projects: how to deal with huge quantities of data and what to do when this data originates from disparate sources.

Understanding source data quality in BI projects—when scoping, early in prototyping, and during package development—is critical when estimating the work involved in building the ETL processes to populate the OLAP cubes and data mining structures. It's common to underestimate the amount

of work involved in cleaning the source data before it is loaded into the SSAS destination structures. The Data Profiling task helps you to understand the scope of the source-data cleanup involved in your projects. Specifically, this cleanup involves deciding which methods to use to clean up your data. Methods can include the use of advanced package transformations (such as fuzzy logic) or more staging areas (relational tables) so that fewer in-memory transformations are necessary during the transformation processes. Other considerations include total number of tasks in a single package, or overall package size.

QUESTION NO: 4

You design a Business Intelligence (BI) solution by using SQL Server 2008.

The solution contains a SQL Server 2008 Analysis Services (SSAS) database. A measure group in the database contains log entries of manufacturing events. These events include accidents, machine failures, production capacity metrics, and other activities.

You need to implement a data mining model that meets the following requirements:

- Predict the frequency of different event types.
- Identify short-term and long-term patterns.

Which algorithm should the data mining model use?

A. the Microsoft Time Series algorithm
B. the Microsoft Decision Trees algorithm
C. the Microsoft Linear Regression algorithm
D. the Microsoft Logistic Regression algorithm

Answer: A
Explanation:
Microsoft Time Series Algorithm
Microsoft Time Series is used to impact a common business problem, accurate forecasting. This algorithm is often used to predict future values, such as rates of sale for a particular product. Most often the inputs are continuous values. To use this algorithm, your source data must contain at one column marked as Key Time. Any predictable columns must be of type Continuous. You can select one or more inputs as predictable columns when using this algorithm.

Time series source data can also contain an optional Key Sequence column.
Function The ARTxp algorithm has proved to be very good at short-term prediction. The ARIMA algorithm is much better at longer-term prediction. By default, the Microsoft Time Series algorithm

blends the results of the two algorithms to produce the best prediction for both the short and long term.

Microsoft Decision Trees Algorithm

Microsoft Decision Trees is probably the most commonly used algorithm, in part because of its flexibility— decision trees work with both discrete and continuous attributes—and also because of the richness of its included viewers. It's quite easy to understand the output via these viewers. This algorithm is used to both view and to predict. It is also used (usually in conjunction with the Microsoft Clustering algorithm) to find deviant values. The Microsoft Decision Trees algorithm processes input data by splitting it into recursive (related) subsets. In the default viewer, the output is shown as a recursive tree structure. If you are using discrete data, the algorithm identifies the particular inputs that are most closely correlated with particular predictable values, producing a result that shows which columns are most strongly predictive of a selected attribute. If you are using continuous data, the algorithm uses standard linear regression to determine where the splits in the decision tree occur. Clicking a node displays detailed information in the Mining Legend window. You can configure the view using the various drop-down lists at the top of the viewer, such as Tree, Default Expansion, and so on. Finally, if you'veenabled drillthrough on your model, you can display the drillthrough information—either columns from the model or (new to SQL Server 2008) columns from the mining structure, whether or not they are included in this model.

Microsoft Linear Regression Algorithm

Microsoft Linear Regression is a variation of the Microsoft Decision Trees algorithm, and works like classic linear regression—it fits the best possible straight line through a series of points (the sources being at least two columns of continuous data). This algorithm calculates all possible relationships between the attribute values and produces more complete results than other (non–data mining) methods of applying linear regression. In addition to a key column, you can use only columns of the continuous numeric data type. Another way to understand this is that it disables splits. You use this algorithm to be able to visualize the relationship between two continuous attributes. For example, in a retail scenario, you might want to create physical placement locations in a retail store and rate of sale for items. The algorithm result is similar to that produced by any other linear regression method in that it produces a trend line. Unlike most other methods of calculating linear regression, the Microsoft Linear Regression algorithm in SSAS calculates all possible relationships between all input dataset values to produce its results. This differs from other methods of calculating linear regression, which generally use progressive splitting techniques between the source inputs

QUESTION NO: 5

You design a Business Intelligence (BI) solution by using SQL Server 2008.

_____

The solution includes a SQL Server 2008 Analysis Services (SSAS) database. A cube in the database contains a large dimension named Customers. The database uses a data source that is located on a remote server.

Each day, an application adds millions of fact rows and thousands of new customers. Currently, a full process of the cube takes several hours.

You need to ensure that queries return the most recent customer data with the minimum amount of latency.

Which cube storage model should you use?

A. hybrid online analytical processing (HOLAP)
B. relational online analytical processing (ROLAP)
C. multidimensional online analytical processing (MOLAP)
D. automatic multidimensional online analytical processing (automatic MOLAP)

Answer: A
Explanation:
Relational OLAP

Relational OLAP (ROLAP) stores the cube structure in a multidimensional database. The leaf-level measures are left in the relational data mart that serves as the source of the cube. The preprocessed aggregates are also stored in a relational database table. When a decision maker requests the value of a measure for a certain set of dimension members, the ROLAP system first checks to determine whether the dimension members specify an aggregate or a leaf-level value. If an aggregate is specified, the value is selected from the relational table. If a leaf-level value is specified, the value is selected from the data mart.

Also, because the ROLAP architecture retrieves leaf-level values directly from the data mart, the leaf-level values returned by the ROLAP system are always as up-to-date as the data mart itself. In other words, the ROLAP system does not add latency to leaf-level data. The disadvantage of a ROLAP system is that the retrieval of the aggregate and leaf-level values is slower than the other OLAP architectures.

Multidimensional OLAP

Multidimensional OLAP (MOLAP) also stores the cube structure in a multidimensional database. However, both the preprocessed aggregate values and a copy of the leaf-level values are placed in the multidimensional database as well. Because of this, all data requests are answered from the multidimensional database, making MOLAP systems extremely responsive. Additional time is required when loading a MOLAP system because all the leaflevel data is copied into the multidimensional database. Because of this, times occur when the leaf-level data returned by the MOLAP system is not in sync with the leaf-level data in the data mart itself. A MOLAP system, therefore, does add latency to the leaf-level data. The MOLAP architecture also requires more disk

space to store the copy of the leaf-level values in the multidimensional database. However, because MOLAP is extremely efficient at storing values, the additional space required is usually not significant.

Hybrid OLAP

Hybrid OLAP (HOLAP) combines ROLAP and MOLAP storage. This is why we end up with the word "hybrid" in the name. HOLAP tries to take advantage of the strengths of each of the other two architectures while minimizing their weaknesses. HOLAP stores the cube structure and the preprocessed aggregates in a multidimensional database. This provides the fast retrieval of aggregates present in MOLAP structures. HOLAP leaves the leaf-level data in the relational data mart that serves as the source of the cube. This leads to longer retrieval times when accessing the leaf-level values. However, HOLAP does not need to take time to copy the leaf-level data from the data mart. As soon as the data is updated in the data mart, it is available to the decision maker. Therefore, HOLAP does not add latency to the leaf-level data. In essence, HOLAP sacrifices retrieval speed on leaf-level data to prevent adding latency to leaf-level data and to speed the data load.

QUESTION NO: 6

You design a Business Intelligence (BI) solution by using SQL Server 2008.

The solution includes a SQL Server 2008 Analysis Services (SSAS) database. The database contains a cube named Financials. The cube contains objects as shown in the exhibit.

A calculated member named Gross Margin references both Sales Details and Product Costs.

You need to ensure that the solution meets the following requirements:

- Managers must be able to view only their cost center's percentage of the company's gross margin.
- The impact on query performance is minimal.

What should you do?

_____

A. Add dimension-level security and enable the Visual Totals option.
B. Add cell-level security that has read permissions on the Gross Margin measure
C. Add cell-level security that has read contingent permissions on the Gross Margin measure.
D. Change the permissions on the Managers dimension level from Read to Read/Write.

Answer: A
Explanation:
http://msdn.microsoft.com/en-us/library/ms174927.aspx
User Access Security Architecture
Microsoft SQL Server Analysis Services relies on Microsoft Windows to authenticate users. By default, only authenticated users who have rights within Analysis Services can establish a connection to Analysis Services. After a user connects to Analysis Services, the permissions that user has within Analysis Services are determined by the rights that are assigned to the Analysis Services roles to which that user belongs, either directly or through membership in a Windows role.

Dimension-Level Security
A database role can specify whether its members have permission to view or update dimension members in specified database dimensions. Moreover, within each dimension to which a database role has been granted rights, the role can be granted permission to view or update specific dimension members only instead of all dimension members. If a database role is not granted permissions to view or update a particular dimension and some or all the dimension's members, members of the database role have no permission to view the dimension or any of its members. Note Dimension permissions that are granted to a database role apply to the cube dimensions based on the database dimension, unless different permissions are explicitly granted within the cube that uses the database dimension.

Cube-Level Security
A database role can specify whether its members have read or read/write permission to one or more cubes in a database. If a database role is not granted permissions to read or read/write at least one cube, members of the database role have no permission to view any cubes in the database, despite any rights those members may have through the role to view dimension members.

Cell-Level Security
A database role can specify whether its members have read, read contingent, or read/write permissions on some or all cells within a cube. If a database role is not granted permissions on cells within a cube, members of the database role have no permission to view any cube data. If a database role is denied permission to view certain dimensions based on dimension security, cell level security cannot expand the rights of the database role members to include cell members from that dimension. On the other hand, if a database role is granted permission to view members of a dimension, cell-level security can be used to limit the cell members from the dimension that

the database role members can view.

QUESTION NO: 7

You design a Business Intelligence (BI) solution by using SQL Server 2008.

The solution includes a SQL Server 2008 Reporting Services (SSRS) infrastructure in a scale-out deployment. All reports use a SQL Server 2008 relational database as the data source. You implement row-level security.

You need to ensure that all reports display only the expected data based on the user who is viewing the report.

What should you do?

A. Store the credential of a user in the data source.
B. Configure the infrastructure to support Kerberos authentication.
C. Configure the infrastructure to support anonymous authentication by using a custom authentication extension.
D. Ensure that all report queries add a filter that uses the User.UserID value as a hiddenparameter.

Answer: B
Explanation:

QUESTION NO: 8

You design a Business Intelligence (BI) solution by using SQL Server 2008.

You need to load data into your online transaction processing (OLTP) database once a week by using data from a flat file. The file contains all the details about new employees who joined your company last week. The data must be loaded into the tables shown in the exhibit. (Click the Exhibit button.) Employee.EmployeeID is an identity.