20467D

Designing Business Intelligence Solutions with Microsoft SQL Server 2014

Module 1

Planning a BI Solution

Contents:

Lesson 1
Elements of a BI Solution

Contents:

Question and Answers

Question: Think about a BI system that you are familiar with. Which of the components that this lesson describes are included in the system, and how were they implemented? Think about the data warehouse schema (for example, CIF or dimensional), the ETL process, data cleansing, master data management, data models, and reporting and analysis.

Answer: Answers will vary because students' experiences will differ.

Lesson 2
The Microsoft BI Platform

Contents:

Question and Answers

Question: Which of the following does SharePoint 2013 support?

( ) Sharing Excel workbooks in a SharePoint document library through a web browser.

( ) Graphical exploration of data models through Power View in a web browser.

( ) Delivery and management of reports through SharePoint document libraries.

( ) Share of Excel workbooks that contain a PowerPivot tabular data models.

( ) All of the above.

> **Answer:** SharePoint 2013, when integrated with SQL Server 2014 Reporting Services can provide all of the functionality described in the question.

Lesson 3
Planning a BI Project

Contents:

Question and Answers

Question: Is the following statement true or false? A BI project is primarily a technology-driven project, so they key stakeholders are likely to be mainly technical and IT people.

( ) True

( ) False

Answer:

( ) True

(√) False

Module Review and Takeaways

Review Question(s)

Question: What challenges have you experienced when gathering, interpreting, and prioritizing business requirements in previous projects? How did you overcome them? What tips can you share with your fellow students?

Answer: Answers will vary depending on the experience of the students.

Lab Review Questions and Answers

Lab: Planning a BI Solution

Question and Answers

Lab Review

Question: If you had been able to conduct discussions with real business stakeholders instead of reading paper-based interview transcripts, how would you have elicited clearer, more specific business requirements?

> **Answer:** There is no definitively correct answer, but a good approach is to have users think about the information they want to gain from the solution, and how they would like to consume that information. In some cases, the requirements can then be expressed as specifications for reports or data models, which can be refined to a considerable degree of detail. Some stakeholders may even present mock-ups or diagrams representing their envisioning of the solution.

Question: In a real project, how would you have determined the feasibility of the requirements you captured in the requirements matrix?

> **Answer:** In most cases, the feasibility is determined by the availability of source data from which the required information can be obtained. If source data is available, the next consideration is the completeness and reliability of that data, and the amount of effort and time required to consolidate it into a conformed dimensional schema and create the necessary reports and analytical models.
>
> Other considerations include the availability of the expertise required to create the envisioned solution, and the frequency with which the data must be refreshed from source systems.
>
> The key point is that, although many business requirements are *possible* to address, the specific constraints for an individual organization, project, or existing infrastructure may make them *unfeasible*.

Module 2

Planning SQL Server Business Intelligence Infrastructure

Contents:

Lesson 1
Considerations for BI Infrastructure

Contents:

Question and Answers

Question: What type of architecture do the data warehouses in your workplace use? In the light of what you have learnt in this lesson, do you think that this is the best architecture for the databases? Can you suggest any ways in which you could improve it?

Answer: Answers will vary.

Resources

Planning for High Availability

Additional Reading: For more information about using AlwaysOn Availability Groups with the SSIS catalog, see *SSIS with AlwaysOn* at http://blogs.msdn.com/b/mattm/archive/2012/09/19/ssis-with-alwayson.aspx.

Lesson 2
Planning Data Warehouse Hardware

Contents:

Question and Answers

Question: When working out the anticipated throughput for a data warehouse, you should start by assessing the CPU requirements and then plan the other components around this.

( ) True

( ) False

>   **Answer:**
>
>   (√) True
>
>   ( ) False

Resources

Microsoft Data Warehouse Fast Track Reference Architectures

🌐   **Additional Reading:** For more information about Data Warehouse Fast Track reference architectures, download the *Fast Track Data Warehouse Reference Guide for SQL Server 2012* white paper at http://msdn.microsoft.com/en-us/library/hh918452.aspx. For more information about specific vendor offerings in the Microsoft Data Warehouse Fast Track program, see the *Data Warehouse Fast Track* page at http://www.microsoft.com/en-us/server-cloud/data-warehouse-fast-track.aspx.

Demonstration: Calculating Maximum Consumption Rate

Demonstration Steps

Create tables for benchmark queries

1. Ensure that the 20467D-MIA-DC and 20467D-MIA-SQL virtual machines are both running, and then log on to 20467D-MIA-SQL as **ADVENTUREWORKS\Student** with the password **Pa$$w0rd**.

2. On the taskbar, click the **SQL Server Management Studio** icon.

3. In the **Connect to Server** dialog box, in the **Server type** list, click **Database Engine**, in the **Server name** field, type **localhost**, , in the **Authentication** list, click **Windows Authentication**, and then click **Connect**.

4. On the **File** menu, point to **Open**, and then click **File**. Browse to the D:\Demofiles\Mod02 folder, select **Create BenchmarkDB.sql**, and then click **Open**.

5. Click **Execute**, and wait for query execution to complete. This query creates a database containing two tables, one with a clustered index and one without. Both tables contain a substantial number of rows.

Execute a query to retrieve I/O statistics

1. On the **File** menu, point to **Open**, and then click **File**. In the D:\Demofiles\Mod02 folder, select **Measure MCR.sql**, and then click **Open**.

2. Click **Execute**, and wait for query execution to complete. The queries retrieve an aggregated value from each table, and are performed twice. This ensures that on the second execution (for which statistics are shown), the data is in cache so the I/O statistics do not include disk reads. Note that the MAXDOP=1 clause ensures that only a single core is used to process the query.

Calculate MCR from the I/O statistics

1. In the results pane, click the **Messages** tab. This shows the statistics for the queries.

2. Add the **logical reads** value for the two queries together, and then divide the result by two to find the average.

3. Add the **CPU time** value for the two queries together, and then divide the result by two to find the average. Divide the result by 1000 to convert it to seconds.

4. Calculate MCR by using the following formula:

    (average logical reads / average CPU time) * 8 / 1024

Module Review and Takeaways

Question: In a growing number of organizations, virtualization has become a core platform for infrastructure. Hyper-V in Windows Server 2012 R2 together with enterprise operations and management software such as Microsoft System Center 2012 has enabled IT departments to benefit from simpler provisioning, management, mobility, and recoverability of services.

What components of a BI infrastructure would you consider virtualizing, and why?

> **Answer:** Many database professionals are resistant to virtualization, particularly with regards to data warehouses, because of the additional layer of abstraction it adds between the database server and the physical hardware—in particular, the disk subsystem. However, advances in virtualization, such as support for virtual HBAs in Windows Server 2012 R2, mean that a virtualized environment can provide near-equivalent performance to a physical server, with advantages in terms of portability and recoverability (because the entire data warehouse server can be easily copied to new physical hardware or backed up).

> For Reporting Services and Analysis Services, a virtualized infrastructure could be extremely beneficial due to the ability to easily move entire virtual servers across physical hosts for hardware maintenance operations or to recover from a hardware failure, and the ability to dynamically scale up virtual hardware resources to match demand. For example, suppose an organization experiences a sharp spike in demand for report processing at the end of the financial year, and a corresponding drop in ad-hoc analysis as users shift their attention to generating year-end reports. The allocation of physical memory and processor resources to the virtual server hosting Reporting Services could be increased, if necessary by reducing the resources allocated to the virtual Analysis Services server. Then, after the year-end reporting activity is over, the resource allocations could be re-adjusted to support the normal workload balance.

> The SQLCAT team at Microsoft has conducted research into the performance of SQL Server database engine and Analysis Services on virtualized infrastructure, and you can review their findings by downloading the white paper *High Performance SQL Server Workloads on Hyper-V* from http://sqlcat.com/sqlcat/b/whitepapers/archive/2010/05/27/high-performance-sql-server-workloads-on-hyper-v.aspx

Lab Review Questions and Answers

Lab: Planning SQL Server Business Intelligence Infrastructure

Question and Answers

Lab Review

Question: Review **BI Topology.vsdx** in the D:\Labfiles\Lab02\Solution folder. How do the various suggested topologies (on multiple pages named A to F) compare to the topology you designed in the lab?

> **Answer:** Answers will vary, and there is no definitively correct answer. However, considering the requirements in the lab, topology B or C is probably the most appropriate based on the relatively light workload requirements. These topologies provide a balance between minimizing server hardware and licensing costs and separating data warehousing and reporting workloads. They also provide scope for using a cluster to provide high availability for the data warehouse server, and can be scaled out at a future point if demand on the BI solution increases.

Question: Review **DW Hardware Spec.xlsx** in the D:\Labfiles\Lab02\Solution folder. How does the hardware specification in this workbook compare to the specification you created in the lab?

> **Answer:** Again, answers will vary. Key points about the suggested solution are:
>
> - The amount of suggested memory is 64 GB per processor, and comes to more than 20 percent of the data volume.
>
> - The suggested storage solution is extensible (more disks can be added to the SAN) and offloads I/O processing overheads to the SAN. Moreover, it enables the solution to balance disk I/O for the storage arrays symmetrically with the processors.

Module 3
Designing a Data Warehouse

Contents:

Lesson 1
Data Warehouse Design Overview

Contents:

Question and Answers

What is the correct sequence of actions for the data warehouse design process that this lesson described? Put the following steps in order by numbering each to indicate the correct order.

	Steps
	Determine analytical and reporting requirements
	Identify the business processes that generate the required data
	Examine the source data for those business processes
	Conform dimensions across business processes
	Prioritize business processes and define a dimensional model for each
	Document and refine the models to determine the database logical schema
	Design the physical data structures for the database

Answer:

	Steps
1	Determine analytical and reporting requirements
2	Identify the business processes that generate the required data
3	Examine the source data for those business processes
4	Conform dimensions across business processes
5	Prioritize business processes and define a dimensional model for each
6	Document and refine the models to determine the database logical schema
7	Design the physical data structures for the database

Resources

The Data Warehouse Design Process

Additional Reading: For a detailed exploration of how to apply the Kimball dimensional modeling methodology to a SQL Server-based data warehouse design, read *The Microsoft Data Warehouse Toolkit* (Wiley, 2011).

Dimensional Modeling

Additional Reading: For more information about using a bus matrix as part of a data warehouse design project, read *The Microsoft Data Warehouse Toolkit* (Wiley, 2011).

Lesson 2
Designing Dimension Tables

Contents:

Question and Answers

Question: You need to include in your dimension design the ability to maintain previous versions of attribute values along with the current version of those attribute values. You want to implement this in a way that does not cause you to change any of the table schemas in the database. How could you best achieve this?

( ) Implement a type 1 slowly changing dimension.

( ) Implement a type 2 slowly changing dimension.

( ) Implement a type 3 slowly changing dimension.

( ) Implement a parent-child hierarchy.

> **Answer:** Implement a type 2 slowly changing dimension is the correct answer. A type 1 slowly changing dimension will not retain the previous versions of attribute values, A type 3 slowly changing dimension will retain the previous versions of attribute values, but requires changes to the table schema to do so. A parent-child hierarchy is a way of implementing hierarchical relationships and does not help to maintain previous and current versions of attribute values.

Lesson 3
Designing Fact Tables

Contents:

Question and Answers

Question: Although it might not always be possible to sum certain types of measures, you can often aggregate them in different ways

( ) True

( ) False

> **Answer:**
>
> (√) True
>
> ( ) False

Lesson 4
Designing a Data Warehouse Physical Implementation

Contents:

Question and Answers

Question: Which of the technologies that this lessons describes would you be likely to implement in your work environment and what benefits would you expect to gain by doing so?

> **Answer:** Answers will vary, and there is no single correct answer. It is important to understand the options that are available for implementing the physical data warehouse database and how the choices that you make will affect factors including performance, resource utilization, and manageability.

Demonstration: Partitioning a Fact Table

Demonstration Steps

Create a Partitioned Table

1. Ensure that the 20467D-MIA-DC and 20467D-MIA-SQL virtual machines are both running, and then log on to 20467D-MIA-SQL as **ADVENTUREWORKS\Student** with the password **Pa$$w0rd**.

2. Start SQL Server Management Studio and connect to the **localhost** instance of the database engine by using Windows authentication.

3. Open **Partitions.sql** from the D:\Demofiles\Mod03 folder.

4. Select the code under the comment **Create a database**, and then click **Execute**. This creates a database for the demonstration.

5. Select the code under the comment **Create filegroups**, and then click **Execute**. This creates four filegroups in the demo database.

6. Select the code under the comment **Create partition function and scheme**, and then click **Execute**. This creates a partition function that defines four ranges of values (less than 20000101, 20000101 to 20010100, 20010101 to 20020100, and 20020101 and higher), and a partition scheme that maps these ranges to the FG0000, FG2000, FG2001, and FG2002 filegroups.

7. Select the code under the comment **Create a partitioned table**, and then click **Execute**. This creates a partitioned table on the partition scheme you created previously.

8. Select the code under the comment **Insert data into the partitioned table**, and then click **Execute**. This inserts four records into the table.

View Partition Metadata

1. Select the code under the comment **Query the table**, and then click **Execute**. This retrieves rows from the table and uses the **$PARTITION** function to show which partition the **datekey** value in each row is assigned to. This function is useful for determining which partition of a partition function a specific value belongs in.

2. Select the code under the comment **View filegroups, partitions, and rows**, and then click **Execute**. This code uses system tables to show the partitioned storage and the number of rows in each partition. Note that there are two empty partitions; one at the beginning of the table, and one at the end.

Split a Partition

1. Select the code under the comment **Add a new filegroup and make it the next used**, and then click **Execute**. This creates a new filegroup named FG2003 and adds it to the partition scheme as the next used partition.

2.   Select the code under the comment **Split the empty partition at the end**, and then click **Execute**. This creates a new partition for values of 20030101 and higher and assigns it to the next used filegroup (FG2003), leaving an empty partition for values between 20020101 and 20030100.

3.   Select the code under the comment **Insert new data**, and then click **Execute**. This inserts two new rows into the partitioned table.

4.   Select the code under the comment **View partition metadata**, and then click **Execute**. This shows that the two rows inserted in the previous step are in partition 4, and that partition 5 (on FG2003) is empty.

Merge Partitions

1.   Select the code under the comment **Merge the 2000 and 2001 partitions**, and then click **Execute**. This merges the partition that contains the value 20010101 into the previous partition.

2.   Select the code under the comment **View partition metadata**, and then click **Execute**. This shows that partition 2 (on FG2000) now contains four rows, and that the partition that was previously on FG2001 has been removed.

3.   Close SQL Server Management Studio.

Demonstration: Creating Indexes

Demonstration Steps

Create Indexes on Dimension Tables

1.   Ensure that the 20467D-MIA-DC and 20467D-MIA-SQL virtual machines are both running, and then log on to 20467D-MIA-SQL as **ADVENTUREWORKS\Student** with the password **Pa$$w0rd**.

2.   Start SQL Server Management Studio and connect to the **localhost** instance of the database engine by using Windows authentication.

3.   Open **Indexes.sql** from the D:\Demofiles\Mod03 folder.

4.   Select the code under the comment **Create the data warehouse**, and then click **Execute**. This creates a database for the demonstration.

5.   Select the code under the comment **Create the DimDate dimension table**, and then click **Execute**. This creates a time dimension table named **DimDate**.

6.   Select the code under the comment **Populate DimDate with values from 2 years ago until the end of this month**, and then click **Execute**. This adds rows to the **DimDate** table.

7.   Select the code under the comment **Create indexes on the DimDate table**, and then click **Execute**. This creates a clustered index on the surrogate key column, and non-clustered indexes on commonly queried attribute columns.

8.   Select the code under the comment **Create the DimCustomer table**, and then click **Execute**. This creates a dimension table named **DimCustomer** and inserts some customer data.

9.   Select the code under the comment **Create indexes on the DimCustomer table**, and then click **Execute**. This creates a clustered index on the surrogate key column, and non-clustered indexes on commonly queried attribute columns.

10. Select the code under the comment **Create the DimProduct table**, and then click **Execute**. This creates a dimension table named **DimProduct** and inserts some product data.

11. Select the code under the comment **Create indexes on the DimProduct table**, and then click **Execute**. This creates a clustered index on the surrogate key column, and non-clustered indexes on a commonly queried attribute column.

View Index Usage and Execution Statistics

1. Select the code under the comment **Create a fact table**, and then click **Execute**. This creates a fact table named **FactOrder** that contains more than 7.5 million rows from the existing data in the dimension tables.

2. On the toolbar, click the **Include Actual Execution Plan** button.

3. Select the code under the comment **View index usage and execution statistics**, and then click **Execute**. This enables statistics messages and queries the tables in the data warehouse to view orders for the previous six months.

4. After query execution completes, in the results pane, click the **Messages** tab. Note the logical reads from each table—the number from the **FactOrder** table should be considerably higher than the dimension tables; and note the CPU time and elapsed time for the query.

5. Click the **Execution plan** tab, which shows a visualization of the steps the query optimizer used to execute the query. Scroll to the right and to the bottom, and note that a table scan was used to read data from the **FactOrder** table. Then hold the mouse pointer over each of the **Index Scan** icons for the dimension tables to see which indexes were used.

6. Execute the selected code again and compare the results when the data is cached.

Create Indexes on a Fact Table

1. Select the code under the comment **Create traditional indexes on the fact table**, and then click **Execute**. This creates a clustered index on the date dimension key, and non-clustered indexes on the other dimension keys (the operation can take a long time).

2. Select the code under the comment **Empty the cache**, and then click **Execute**. This clears any cached data.

3. Select the code under the comment **Test the traditional indexes**, and then click **Execute**. This executes the same query as earlier.

4. Click the **Messages** tab and compare the number of logical reads for the **FactOrders** table and the CPU and elapsed time values with the previous execution. They should all be lower.

5. Click the **Execution plan** tab and note that the clustered index on the date key in the fact table was used.

6. Execute the selected code again and compare the results when the data is cached.

Create a Columnstore Index

1. Select the code under the comment **Create a copy of the fact table with no indexes**, and then click **Execute**. This creates an un-indexed copy of the **FactOrder** table named **FactOrderCS**.

2. Select the code under the comment **Create a columnstore index on the copied table**, and then click **Execute**. This creates a columnstore index on all columns in the **FactOrderCS** table.

3. Select the code under the comment **Empty the cache again**, and then click **Execute**. This clears any cached data.

4. Select the code under the comment **Test the columnstore index**, and then click **Execute**. This executes the same query as earlier.

5. Click the **Messages** tab and compare the number of logical reads for the **FactOrdersCS** table and the CPU and elapsed time values with the previous execution. They should all be lower.

6. Click the **Execution plan** tab and note that the columnstore index on the fact table was used.

7. Execute the selected code again and compare the results when the data is cached.

8. Close SQL Server Management Studio.

Demonstration: Implementing Data Compression

Demonstration Steps

Create Uncompressed Tables and Indexes

1. Ensure that the 20467D-MIA-DC and 20467D-MIA-SQL virtual machines are both running, and then log on to 20467D-MIA-SQL as **ADVENTUREWORKS\Student** with the password **Pa$$w0rd**.

2. Use File Explorer to view the contents of the D:\Demofiles\Mod03 folder, and set the folder window to Details view and resize it if necessary so that you can see the **Size** column.

3. Start SQL Server Management Studio and connect to the **localhost** instance of the database engine by using Windows authentication.

4. Open **Compression.sql** from the D:\Demofiles\Mod03 folder.

5. Select the code under the comment **Create the data warehouse** (from line2 to line 113 in the script), and then click **Execute**. This creates a database with uncompressed tables.

6. While the script is still executing, view the contents of the D:\Demofiles\Mod03 folder and note the increasing size of **DemoDW.mdf**. This is the data file for the database.

    Note: The log file (DemoDW.ldf) will also be growing, but you can ignore this.

7. When execution is complete (after approximately 3 minutes), view the final size of **DemoDW.mdf** and return to SQL Server Management Studio.

Estimate Compression Savings

1. Select the code under the comment **Estimate size saving** (line 119 in the script), and then click **Execute**. This uses the **sp_estimate_data_compression_savings** system stored procedure to compress a sample of the **FactOrder** table (which consists of a clustered index and two non-clustered indexes).

2. View the results returned by the stored procedure, noting the current size and estimated compressed size of each index.

Create Compressed Tables and Indexes

1. Select the code under the comment **Create a compressed version of the database** (from line 125 to line 250 in the script), and then click **Execute**. This creates a database with compressed tables and indexes.

2. While the script is still executing, view the contents of the D:\Demofiles\Mod03 folder and note the increasing size of **CompressedDemoDW.mdf**. This is the data file for the database.

    Note: The log file (CompressedDemoDW.ldf) will also be growing, but you can ignore this.

3. When execution is complete (after approximately 3 minutes), compare the final size of **CompressedDemoDW.mdf** with **DemoDW.mdf** (the file for the compressed database should be smaller) and return to SQL Server Management Studio.

Compare Query Performance

1. Select the code under the comment **Compare query performance** (from line 255 to line 277 in the script), and then click **Execute**. This executes an identical query in the compressed and uncompressed databases and displays execution statistics.

2. When execution is complete, click the **Messages** tab and compare the statistics for the two queries. The execution time statistics (the second and third set of figures labeled "SQL Server Execution Times") should be similar, and the second query (in the compressed database) should have used considerably fewer logical reads for each table than the first.

3. Close SQL Server Management Studio.

Module Review and Takeaways

Question: When designing a data warehouse, is it better or worse to have a strong background in transactional database design?

Answer: Answers will vary, and the point of the question is to get students to consider the differences and similarities between transactional database design and data warehouse design. In general, a good understanding of fundamental database concepts, such as tables, keys, relationships, and indexes is critical when designing a data warehouse; and knowledge of transactional database schema techniques and Transact-SQL is useful when exploring source systems. However, there is a danger that a transactional database developer can easily fall into the trap of instinctively normalizing data, so that a data warehouse that would best suit a star schema ends up with snowflake dimensions and queries require multiple joins to retrieve de-duplicated data.

Ideally, a data warehouse developer should have some knowledge of general database design principles, combined with an ability to understand business processes and priorities. To design a successful data warehouse, you must apply dimensional modeling techniques to the design of the data warehouse, while being pragmatic about meeting the analytical and reporting priorities of the business.

Lab Review Questions and Answers

Lab A: Designing a Data Warehouse Logical Schema

Question and Answers

Lab Review

Question: Use Excel to open Matrix.xlsx in the D:\Labfiles\Lab03A\Solution folder and compare it to the matrix your group created during the lab. What are the significant differences between your solution and the suggested solution, and how would you justify your choices in the lab?

> **Answer:** Answers will vary, and the point of the question is to provoke discussion. The solution document includes some notes justifying the design, but students may have interpreted or prioritized requirements differently or identified a different set of dimensions. An important point to bring out is that for the process to succeed, you need a constant iteration of communication with the business stakeholders (which is not possible in a classroom environment).

Question: Use Visio to open Initial Sun Diagram.vsdx in the D:\Labfiles\Lab03A\Solution folder. How do the dimensional models in this document compare to your solution?

> **Answer:** Again, answers will vary.

Question: Use Visio to open DW Schema.vsdx in the D:\Labfiles\Lab03A\Solution folder. How does the database schema design in this document compare to your solution?

> **Answer:** Again, answers will vary.

Lab B: Designing a Data Warehouse Physical Implementation

Question and Answers

Question: Like in the previous lab, there is no definitive correct solution, but a sample solution has been provided. To view the sample data warehouse implementation, run Setup Solution.cmd in the D:\Labfiles\Lab03B\Solution folder as Administrator. Then, after the script finishes running, use SQL Server Management Studio to connect to the MIA-SQLDW instance of the database engine and examine the AWDataWarehouse database. You can use the AWDataWarehouse.docx document in the D:\Labfiles\Lab03B\Solution folder as a guide to the key features of the solution.

After spending some time reviewing the solution, what are the key aspects of the implementation that differ from your design in the lab, and how else might you have designed the solution?

> **Answer:** Answers will vary. The point of the question is to provoke discussion about the design considerations that needed to be addressed in the lab scenario.

Module 4

Designing an ETL Solution

Contents:

Lesson 1
Overview of ETL

Contents:

Question and Answers

What are the key elements for each of the three ETL architectures listed below? Categorize each element into the appropriate ETL category. Indicate your answer by writing the category number to the right of each item.

Items	
1	Source
2	Source
3	Source
4	Data Warehouse
5	Staging database
6	Landing zone
7	Data warehouse
8	Staging database
9	Data warehouse

Category 1		Category 2		Category 3	
Single-stage ETL		Two-stage ETL		Three-stage ETL	

Answer:

Category 1		Category 2		Category 3
Single-stage ETL		Two-stage ETL		Three-stage ETL
Source Data Warehouse		Source Staging database Data warehouse		Source Landing zone Staging database Data warehouse

Lesson 2
Planning Data Extraction

Contents:

Lesson 2
Planning Data Extraction

Question and Answers

Question: The only way that you can profile data from source systems is to use the SQL Server Integration Services data profiling task.

( ) True

( ) False

Answer:

( ) True

(√) False

Lesson 3
Planning Data Transformation

Contents:

Question and Answers

Question: What are the key factors you should consider when planning where in the data flow to perform data transformations?

> **Answer:** There is no single correct answer, but a plan should balance the need to minimize the impact on source systems, with the need to move the data efficiently across the network and into the data warehouse.

Demonstration: Comparing Transact-SQL and Data Flow Transformations

Demonstration Steps

1. Ensure that the 20467D-MIA-DC and 20467D-MIA-SQL virtual machines are both running, and then log on to 20467D-MIA-SQL as **ADVENTUREWORKS\Student** with the password **Pa$$w0rd**.

2. Start SQL Server Management Studio and connect to the **localhost** instance of the database engine by using Windows authentication.

3. Click **File**, point to **Open**, and then click **File**, navigate to the D:\Demofiles\Mod04 folder, click **Setup.sql**, click **Open**, wait for the file to open, and then click **Execute**.

4. In **Object Explorer**, expand **Databases**, expand **DemoExtract**, expand **Tables**, and then review the tables in the **DemoExtract** database. Note that the **Customers** table and the **MembershipLevels** table are both in the **src** schema, and the **DimCustomer** table is in the **stg** schema.

5. Right-click the **DimCustomer** table, and then click **Select Top 1000 Rows**. Note that the table is currently empty.

6. On the Taskbar, click **Visual Studio 2013**, click **File**, point to **Open**, and then click **File**, browse to the D:\Demofiles\Mod04 folder, click **DemoExtract.sln**, and then click **Open**.

7. In Solution Explorer, click **Transact-SQL.dtsx**, on the design surface, double-click **Truncate Staging Table**, in the Execute SQL Task Editor window, review the settings, and then click **Cancel**.

8. On the design surface, double-click **Extract Customers**, and then double-click **Customers**.

9. In the **OLE DB Source Editor** window, on the **Connection Manager** page, review the settings. Note that the connection manager uses a Transact-SQL statement to retrieve the data from the Customers and MembershipLevels tables.

10. Review the Transact-SQL statement and note the following points:

     a. The CAST function converts the data type of **CustomerID** to **nvarchar**.

     b. The **Firstname** and **Lastname** columns are concatenated into a single column called **CustomerName**.

     c. The ISNULL function replaces NULL values in the **MembershipLevelName** column with the value **Unknown**.

     d. The LEFT OUTER JOIN retrieves all customers form the **Customers** table, including those for which there is no matching row in the **MembershipLevels** table.

11. In the **OLE DB Source Editor** window, click **Cancel**.

12. On the design surface, double-click **DimCustomers**, in the **OLE DB Destination Editor** window, click **Mappings** and then review the column mapping settings.

13. In the **OLE DB Destination Editor** window, click **Cancel**.

14. In Solution Explorer, double-click **SSIS Transforms.dtsx**, on the design surface, double-click **Extract Customers**.

15. On the Data Flow design surface, double-click **Customers**, and then review the Transact-SQL statement, noting that it only retrieves rows form the Customers table and that it does not include the CAST, CONVERT, or ISNULL functions.

16. In the **OLE DB Source Editor** window, click **Cancel**.

17. On the Data Flow design surface, double-click **Convert Data Types**, and then in the **Data Conversion Transformation Editor** window, review the settings, noting that the transformation converts the **CustomerID**, **FirstName**, and **LastName** columns to the Unicode String datatype.

18. In the **Data Conversion Transformation Editor** window, click **Cancel**.

19. On the Data Flow design surface, double-click **Concatenate Names**, and then in the **Derived Column Transformation Editor** window, review the settings. Note that the transformation concatenates two of the output columns from the **Convert Data Types** transformation.

20. In the **Derived Column Transformation Editor** window, click **Cancel**.

21. On the Data Flow design surface, double-click **Lookup Membership Level**, and then in the **Lookup Transformation Editor** window, click **Connection**. Review the Transact-SQL statement, and note that it retrieves the rows from the **MembershipLevels** table.

22. Click **Columns**, and then review the settings. Note that the transformation matches rows in the **MembershipLevel** column in the data flow to the **MembershipLevelID** column from the **MemberShipLevels** table, and returns the **MembershipLevelName** for each row as the output.

23. In the **Lookup Transformation Editor** window, click **Cancel**.

24. On the Data Flow design surface, double-click **Handle NULLs**, and then in the **Derived Column Transformation Editor** window, review the expression that replaces NULL values in the **LookedUpMembershipLevel** output from the previous step with the value **Unknown**.

25. In the **Derived Column Transformation Editor** window, click **Cancel**.

26. On the design surface, double-click **DimCustomers**, in the **OLE DB Destination Editor** window, click **Mappings** and then review the column mapping settings.

27. In the **OLE DB Destination Editor** window, click **Cancel**.

28. In Solution Explorer, double-click **Transact-SQL.dtsx**, and then on the **DEBUG** menu, click **Start Debugging**. Wait until the package executes successfully, and then on the **DEBUG** menu, click **Stop Debugging**.

29. On the Taskbar, click **SQL Server Management Studio**, in Object Explorer, right-click **stg.DImCustomer**, and then click **Select Top 1000 Rows**. Review the results of the query, noting that there are three rows, customer names have been concatenated, and that there are no NULL values in the **MembershipLevel** column.

30. On the Taskbar, click **Visual Studio 2013**, in Solution Explorer, double-click **SSIS Transforms.dtsx**, and then on the **DEBUG** menu, click **Start Debugging**. Wait until the package executes successfully, and then on the **DEBUG** menu, click **Stop Debugging**.

31. On the Taskbar, click **SQL Server Management Studio**, in Object Explorer, right-click **stg.DImCustomer**, and then click **Select Top 1000 Rows**. Review the results of the query, noting that they are exactly the same as when you used the **Transact-SQL.dtsx** package.

32. Close **SQL Server Management Studio**, and then close **Visual Studio 2013**. Do not save any changes.

Lesson 4
Planning Data Loads

Contents:

Question and Answers

Question: Which trace flag can you use to enable tables that contain data and have a clustered index to support minimal logging for inserts?

(   ) Trace flag 610

(   ) Trace flag 634

(   ) Trace flag 260

(   ) Trace flag 1224

(   ) Trace flag 1448

> **Answer:** Trace flag 610 enables tables that contain data and have a clustered index to support minimal logging for inserts.

Demonstration: Loading a Partitioned Fact Table

Demonstration Steps

Split a Partition

1. Ensure that the 20467D-MIA-DC and 20467D-MIA-SQL virtual machines are both running, and then log on to 20467D-MIA-SQL as **ADVENTUREWORKS\Student** with the password **Pa$$w0rd**.

2. Start SQL Server Management Studio and connect to the **localhost** instance of the database engine by using Windows authentication.

3. Click **File**, point to **Open**, and then click **File**, navigate to the **D:\Demofiles\Mod04** folder, click **Partitions.sql**, and then click **Open**.

4. Select the code under the comment **Create a database** (from line 1 to line 45), and then click **Execute**. This creates a database with a partitioned fact table with a columnstore index.

5. Select the code under the comment **View partition metadata**, and then click **Execute**. This shows the partitions in the table with their starting and ending range values, and the number of rows they contain. Note that the partitions are shown once for each index (or for the heap if no clustered index exists). Note that the final partition (4) is for key values of 20020101 or higher and currently contains no rows.

6. Select the code under the comment **Add a new filegroup and make it the next used**, and then click **Execute**. This creates a filegroup, and configures the partition scheme to use it for the next partition to be created.

7. Select the code under the comment **Split the empty partition at the end**, and then click **Execute**. This splits the partition function to create a new partition for keys with the value 20030101 or higher.

8. Select the code under the comment **View partition metadata again**, and then click **Execute**. This time the query is filtered to avoid including the same partition multiple times. Note that the table now has two empty partitions (4 and 5).

Create a Load Table

1. Select the code under the comment **Create a load table**, and then click **Execute**. This creates a table on the same filegroup as partition 4, with the same schema as the partitioned table.

2. Select the code under the comment **Bulk load new data**, and then click **Execute**. This inserts the data to be loaded into the load table (in a real solution, this would typically be bulk loaded from staging tables).

3. Select the code under the comment **Add constraints and indexes to the load table**, and then click **Execute**. This adds a check constraint to the table that matches the partition function criteria, and a columnstore index that matches the index on the partitioned table.

Switch a Partition

1. Select the code under the comment **Switch the partition**, and then click **Execute**. This switches the load table with the partition on which the value 20020101 belongs. Note that the required partition number is returned by the $PARTITION function.

2. Select the code under the comment **Clean up and view partition metadata**, and then click **Execute**. This drops the load table and returns the metadata for the partitions. Note that partition 4 now contains two rows—these are the rows that were inserted into the load table.

3. Select the code under the comment **Drop the DemoDW database**, and then click **Execute**. Close SQL Server Management Studio, and do not save any changes to Partitions.sql.

Module Review and Takeaways

Question: In what scenarios would you consider using Transact-SQL for transformations, and in what scenarios are SSIS data flow transformations appropriate?

> **Answer:** Answers will vary depending on the level of experience and comfort students have with Transact-SQL and SSIS. Both approaches are appropriate in many scenarios, but generally you should avoid using intensive Transact-SQL queries when extracting data from source systems that must support other concurrent workloads. Additionally, SSIS data flow transformations offer greater flexibility for detecting and redirecting invalid rows, and incorporating complex custom logic into data flows.

Lab Review Questions and Answers

Lab: Designing an ETL Solution

Question and Answers

Lab Review

Question: Compare the source-to-target documentation in the D:\Labfiles\Lab04\Solution folder with your own documentation. What significant differences are there in the suggested solutions compared to your own, and how would you justify your own solutions?

> **Answer:** Answers will vary. The point of the exercise is to have students think about how they can create and document a detailed data flow plan in the form of high-level diagrams and source-to-target mappings. The formats used in the lab are not based on a particular standard, and students may have devised their own documentation styles that provide useful information.

Question: How might your design of the SSIS package that loads the **FactResellerSales** table have differed if the table was partitioned on **OrderDateKey** instead of **ShipDateKey**?

> **Answer:** If data loads were still performed on the last night of the month, and only orders that had been shipped were included, the SSIS package might need to load orders into multiple partitions. For example, the load could include orders that were placed the previous month but had not been shipped when the previous load was performed, as well as orders placed and shipped during the current month. To accomplish this, a load table could be created for each affected partition and the switch technique used to switch out the already-loaded partition from last month before dropping indexes, inserting new rows, recreating indexes, and switching it back in again.
>
> If the data load process was changed to include orders that have not yet shipped, the load process would need to support updates to already-loaded fact rows (to set the shipped date for orders that have been loaded before they shipped). Once again, a possible solution to this problem would be to switch out the already-loaded partitions, drop the indexes, update the data, recreate the indexes, and switch the partitions back in.

Module 5

Designing Analytical Data Models

Contents:

Lesson 1
Introduction to Analytical Data Models

Contents:

Question and Answers

Question: Which one of the following statements about SQL Server 2014 Analysis Services data models is *not* correct?

( ) Both multidimensional data models and tabular data models support MDX queries.

( ) Both multidimensional data models and tabular data models support data mining.

( ) Both multidimensional data models and tabular data models support DAX queries.

( ) Multidimensional data models use disk-based storage.

( ) Tabular data models can use in-memory storage.

> **Answer:** One of the important differences between multidimensional data models and tabular data models is that only multidimensional data models support data mining.

Resources

Data Model Features and Functionality

Additional Reading: For more information about the different features supported by multidimensional and tabular models, see "Comparing Tabular and Multidimensional Solutions (SSAS)" at https://msdn.microsoft.com/en-us/library/hh212940(v=sql.120).aspx

Lesson 2
Designing an Analytical Data Model

Contents:

Question and Answers

Question: When importing data into a tabular or multidimensional data model form a source database table, it is best practice to import all of the columns from the source table.

( ) True

( ) False

Answer:

( ) True

(√) False

Lesson 3
Designing Dimensions

Contents:

Question and Answers

Match each description to the type of hierarchy that it describes. Indicate your answer by writing the category number to the right of each item.

Items	
1	A hierarchy in which all branches of the hierarchy do not descend to the same level, for example a geography hierarchy that includes the attributes Country, State, and City.
2	A hierarchy in which the entities have a self-referencing relationship, for example an employee hierarchy that shows management chains.
3	A hierarchy in which all branches of the hierarchy descend to the same level, for example a calendar hierarchy that includes year, quarter, and month.

Category 1		Category 2		Category 3
Ragged hierarchy		Parent-child hierarchy		Balanced hierarchy

Answer:

Category 1		Category 2		Category 3
Ragged hierarchy		Parent-child hierarchy		Balanced hierarchy
A hierarchy in which all branches of the hierarchy do not descend to the same level, for example a geography hierarchy that includes the attributes Country, State, and City.		A hierarchy in which the entities have a self-referencing relationship, for example an employee hierarchy that shows management chains.		A hierarchy in which all branches of the hierarchy descend to the same level, for example a calendar hierarchy that includes year, quarter, and month.

Lesson 4
Enhancing Data Models

Contents:

Question and Answers

Question: You want to implement a data model that will support data mining and offer the best possible performance. Which of these answers best meets these requirements?

( ) A tabular data model in direct query mode

( ) A multidimensional data model in HOLAP mode

( ) A tabular data model that uses in-memory storage

( ) A multidimensional data model in MOLAP mode

( ) A multidimensional data model in ROLAP mode

> **Answer:** Tabular data models do not support data mining, so only a multidimensional data model would be suitable. Of the three storage modes for multidimensional data models, MOLAP generally offers the best performance. Note that in real-world implementations, you should always test performance because other factors, such as the hardware used on the data warehouse server and the OLAP server, can have an influence.

Module Review and Takeaways

Question: Now that you are familiar with both models, what are the key factors that you would influence your decision about which to use in a particular business scenario?

Answer: There are a variety of factors to consider, including:

- Specific requirements that can be met using only one model or the other, or that are significantly easier to implement in one model than the other. For example, ragged hierarchies are supported only in multidimensional models, whereas Power View can consume only tabular models.

- The level of familiarity of the data model developers with the model types.

- The volume of data and available hardware. Specifically, if the volume of data is extremely large and memory is limited, a multidimensional model is probably a better choice because of the flexibility it provides in terms of storing aggregations on disk. For smaller amounts of data that can be comfortably handled in memory, a tabular model may perform better.

Lab Review Questions and Answers

Lab A: Designing Analytical Data Models

Question and Answers

Lab Review

Question: Which model did you find easiest to create and work with?

> **Answer:** Answers will vary, depending on student experience. For students with little experience of OLAP design, the multidimensional model can be confusing, and the tabular model is more intuitive for database professionals accustomed to working with relational databases. However, the wizards in the multidimensional model do a great deal of the work for you—in particular, creating the measures and automatically assigning friendly names by adding spaces.

Lab B: Designing Dimensions and Hierarchies

Question and Answers

Question: How do the two models compare when designing dimensions and hierarchies?

> **Answer:** The multidimensional model offers more capabilities (such as ragged hierarchies and account intelligence), and makes it easier to create role-playing dimensions and parent-child hierarchies. However, configuring attribute properties and relationships in a multidimensional can become complex, and hierarchies are generally easier to create and manage in a tabular model.

Lab C: Enhancing Data Models

Question and Answers

Question: How did the experience of creating cube enhancements vary between the two models?

> **Answer:** Answers will vary. Many basic tasks are easier, and therefore quicker to perform, in tabular data models, but for more complex calculated measures and KPIs that include trends, multidimensional data models might be a better fit.

Module 6

Planning a BI Delivery Solution

Contents:

Lesson 1
Considerations for Delivering BI

Contents:

Question and Answers

Question: Which of the reporting tools that this lesson described have you used? Which tools do you think best met the needs of your end users?

> **Answer:** Answers will vary as the question refers to personal experience.

Lesson 2
Common Reporting Scenarios

Contents:

Question and Answers

Match the report types with the appropriate descriptions. Indicate your answer by writing the category number to the right of each item.

Items	
1	Based on a combination data from multiple sources
2	Designed to communicate specific information for a specific audience
3	An unstructured and unconstrained activity
4	Not typically a regular part of business operations
5	Often distributed in document formats
6	Performed by business users who are familiar with the business measures and dimensions
7	Scoped at regular intervals
8	Involves examining data to uncover relationships or patterns

Category 1		Category 2		Category 3
Data mashups		Formal Reports		Data Exploration and Analysis

Answer:

Category 1		Category 2		Category 3
Data mashups		Formal Reports		Data Exploration and Analysis
Based on a combination data from multiple sources Not typically a regular part of business operations		Designed to communicate specific information for a specific audience Often distributed in document formats Scoped at regular intervals		An unstructured and unconstrained activity Performed by business users who are familiar with the business measures and dimensions Involves examining data to uncover relationships or patterns

Lesson 3
Choosing a Reporting Tool

Contents:

Question and Answers

Question: A business user asks your advice about which tool is best for her to use for performing in-depth data exploration and which for creating formal reports. Which of the following tools might best meet her requirements?

(   ) Power View and Report Designer

(   ) PerformancePoint and Report Designer

(   ) PerformancePoint and Report Builder

(   ) Excel and Report Builder

> **Answer:** Of the four answers provided, Excel and Report Builder best meet the stated requirements. In a real-world scenario, you would need to find out more about what the user needs to do before making your recommendation.

Module Review and Takeaways

Question: How does the inclusion of a requirement for self-service BI influence the choice of data tools?

Answer: Answers will vary. In reality, most BI solutions are either fully provided by IT based on specific business requirements and requests for reports, or they contain a mixture of IT-provided and self-service reports. There are very few enterprise BI solutions that are completely self-service. That being the case, it is most common for formal reports, dashboards, and scorecards to be created by BI professionals. Business users might supplement the IT-provided formal reports with their own reports, but this doesn't really influence the choice of tool because both Reporting Services and Excel enable business users to author and publish reports. However, if Reporting Services is used, some consideration must be given to enabling self-service reporting in the form of training on the use of Report Builder and the creation of reusable data sources, datasets, and report parts that will make it easier for less technical users to build their own reports.

For interactive analysis scenarios, Excel provides a flexible environment that is familiar to most users. Business users can perform self-service reporting and analysis against IT-provided data models, or power users can create their own PowerPivot data models in Excel. The main considerations for supporting this type of user-driven data modeling is the installation and enabling of the PowerPivot add-in for Excel, and the SharePoint features that must be enabled to support sharing of user-created PowerPivot workbooks.

Lab Review Questions and Answers

Lab: Planning a BI Delivery Solution

Question and Answers

Lab Review

Question: Review **Reporting Requirements** in the D:\Labfiles\Lab06\Solution folder. How do the recommendations in this document compare to the decisions you made in the lab?

>   **Answer:** Answers will vary. Use the notes in the document to discuss each of the requirements and the different options available for fulfilling them.

Module 7

Designing a Reporting Services Solution

Contents:

Lesson 1
Planning a Reporting Services Solution

Contents:

Question and Answers

Question: Your proposed Reporting Services plan includes a requirement for linked reports. Which one of the answers below best describes the appropriate Reporting Services architecture based on this requirement?

(   ) You should use Reporting Services in SharePoint-Integrated mode.

(   ) You should use Reporting Services in Native mode.

(   ) You can use Reporting Services in SharePoint-Integrated mode or in Native mode.

> **Answer:** You should use Reporting Services in Native mode because only Native mode supports linked reports.

Lesson 2
Designing Reports

Contents:

Question and Answers

Question: Think about the reports that are produced in your work environment. How might you improve them based in the information provided in this lesson?

> **Answer:** Answers will vary because the question relates to students' personal experiences.

Lesson 3
Planning Report Consistency

Contents:

Question and Answers

Question: You create a template to standardize the reports that developers in your organization create. The developers use various different data sources. Some of the developers use Report Designer to create reports, and others use Report Builder. How should you enable the developers to use your template?

( ) Install the template locally on the computers belonging to the developers who use Report Builder, and create a linked report for the developers who use Report Designer.

( ) Install the template locally on the computers belonging to the developers who use Report Builder. Save the template to a share for the developers who use Report Designer, and set permissions on the report to prevent it being over-written.

( ) Install the template locally on the computers belonging to all of the developers.

( ) Install the template locally on the computers belonging to the developers who use Report Designer, and create a linked report for the developers who use Report Builder.

( ) Install the template locally on the computers belonging to the developers who use Report Designer. Save the template to a share for the developers who use Report Builder, and set permissions on the report to prevent it being over-written.

> **Answer:** Install the template locally on the computers belonging to the developers who use Report Designer. Save the template to a share for the developers who use Report Builder, and set permissions on the report to prevent it being over-written. Report Designer supports locally stored templates, but Report Builder does not. Linked reports are not appropriate in this scenario.

Module Review and Takeaways

Question: What are likely to be the key challenges in providing a self-service reporting solution that includes an Analysis Services data model as a data source, and how might you overcome them?

Answer: Many users are familiar with basic Transact-SQL syntax, and will be capable of creating datasets that retrieve data from a data warehouse. However, MDX syntax is more complex, and it is likely that business users with no formal training or background in multidimensional analysis will not be able to construct MDX queries. Report Builder provides a graphical query editor for MDX queries that can be useful for basic queries, but the MDX it generates is not always the most efficient code and it can be difficult to edit the source syntax. You can overcome these difficulties to an extent by providing shared datasets that encapsulate common analytical queries, which business users can then select when designing their reports.

However, although an Analysis Services data model is a suitable source for formal reports, in a self-service environment, users are likely to want to explore the data interactively. In many cases, using Report Builder to create reports from data models is not as effective a business solution as enabling users to connect to the data models from Excel, or if the data model is tabular from Power View, to analyze the data interactively and share their results in the form of an Excel workbook, Power View report, or PowerPoint presentation.

Lab Review Questions and Answers

Lab: Designing a Reporting Services Solution

Question and Answers

Lab Review

Question: What were the key organizational principles you applied when designing the report server folder structure, and what revisions did you consider when you started to create and publish report items?

> **Answer:** Answers will vary. The key factors in deciding a folder structure are described in the *Planning Report Navigation* topic in Lesson 1: *Planning a Reporting Services Solution*. It is not unusual for the folder hierarchy design to be revised as initial report items are created and published.

Module 8

Designing a Microsoft Excel-Based Reporting Solution

Contents:

Lesson 1
Using Excel for Data Analysis and Reporting

Contents:

Question and Answers

Question: You want to import data into a Microsoft Excel 2013 workbook from a data source for which there is no existing data source provider. How could you achieve this with the minimum effort?

(   ) It is not possible to import data from sources for which there is no data source provider.

(   ) Import the data into an Analysis Services cube and then connect to the cube from the Excel workbook.

(   ) Export the data to a common format such as a .csv file or an .xml file and then import the data from this file into Excel.

> **Answer:** In most cases, the most solution would be to export the data to a common format such as a .csv file or an .xml file and then import the data from this file into Excel. Creating a cube and importing the data into it, and then connecting to the cube from the workbook might work, but would require a greater effort.

Lesson 2
PowerPivot for Excel

Contents:

Question and Answers

Question: Your organization has decided to deploy PowerPivot for Excel. Some users use Excel 2013, and other users use Excel 2010. Which of one the following will you need to do to enable all users to use PowerPivot for Excel?

(   ) Download and install the appropriate PowerPivot for Excel add-in for users of Excel 2010. Ensure that the built-in PowerPivot for Excel Add-In is enabled for all users of Excel 2013.

(   ) Download and install the appropriate PowerPivot for Excel Add-In for users of Excel 2010 and Excel 2013.

(   ) Ensure that the built-in PowerPivot for Excel Add-In is enabled for all users of Excel 2010 and Excel 2013.

(   ) Upgrade all instances of Excel 2010 to Excel 2013, and then ensure that the built-in PowerPivot for Excel Add-In is enabled for all users.

> **Answer:** You should download and install the appropriate PowerPivot for Excel add-in for users of Excel 2010 and ensure that the built-in PowerPivot for Excel Add-In is enabled for all users of Excel 2013. Excel 2013 has the PowerPivot for Excel Add-In built-in, but it may be disabled. Excel 2010 does not have the Add-In built-in and you need to download and install it. There is no need to upgrade Excel 2010 to Excel 2013 because Excel 2010 supports PowerPivot for Excel if you install the Add-In.

Lesson 3
Power View for Excel

Contents:

Question and Answers

Question: Users in your organization want to start using Power View in Excel. Some users use Excel 2013, and other users use Excel 2010. Which of one the following will you need to do to enable all users to use PowerPivot for Excel?

( ) Upgrade all instances of Excel 2010 to Excel 2013, and then ensure that the built-in Power View COM Add-In is enabled for all users.

( ) Download and install the appropriate Power View Add-In for users of Excel 2010 and Excel 2013.

( ) Upgrade all instances of Excel 2010 to Excel 2013, and then ensure that the built-in Power View Excel Add-In is enabled for all users

( ) Download and install the Power View Add-In for users of Excel 2010, and ensure that the built-in COM Power View Add-In is enabled for users of Excel 2013.

> **Answer:** Power View in Excel is not available for versions of Excel earlier than Excel 2013, so it is not possible for users of Excel 2010 to use Power View. Consequently, you should upgrade all instances of Excel 2010 to Excel 2013, and then ensure that the built-in Power View COM Add-In is enabled for all users. Additionally, Power View is built-in to Excel as a 2013 COM Add-In, so you do not need to download and install it, although you may need to enable it.

Module Review and Takeaways

Question: What challenges do think organizations will face when empowering users to analyze data in Excel?

Answer: Answers will vary depending on the level of Excel expertise that already exists in the organization. Users may require training, and clear documentation about the data sources they can connect to and the tables and cubes they can retrieve data from. Additionally, suitable permissions must be applied to data sources.

When using PowerPivot, a significant challenge is likely to be the need to enable users to distribute and share their data models. PowerPivot workbooks can become very large, so sending them as email attachments is usually not a suitable solution. PowerPivot works best when implemented in a SharePoint environment that enables users to publish PowerPivot and consume workbooks in a SharePoint site.

Lab Review Questions and Answers

Lab: Designing a Microsoft Excel-Based Reporting Solution

Question and Answers

Lab Review

Question: How might you support the Sales VP's requirement to visually analyze the marketing data models in Excel 2010?

> **Answer:** If the Sales VP is restricted to Excel 2010, the marketing specialists could create the PowerPivot models using PowerPivot for Excel 2010, and the sales VP could create PivotCharts from the data instead of using Power View. This provides a less flexible and more complex chart development option than Power View, but it would achieve the objective.
>
> Alternatively, the marketing specialists could create PowerPivot workbooks using either Excel 2010 or Excel 2013 and publish them to a SharePoint Server site that has Reporting Services installed in SharePoint Integrated mode. The Sales VP could then use Power View in the SharePoint site to create visualizations from the workbooks.

Module 9

Planning a SharePoint Server BI Solution

Contents:

Lesson 1
Introduction to SharePoint Server as a BI Platform

Contents:

Question and Answers

Question: The business requirements for SharePoint-based BI solution include the following:

High-performance and minimal wait times.

Minimal downtime in the event of failure of a service or other component.

Strict separation of data between the different departments of the organization for reasons of security.

Which of the following high-level solutions is most appropriate?

( ) A scale-out solution with a single SharePoint site.

( ) A scale-out solution with multiple subsites.

( ) A high-availability solution with a single SharePoint site.

( ) A high-availability solution with multiple subsites

( ) A single server solution with multiple subsites.

> **Answer:** To meet the business requirements for performance and high availability, a high availability topology is the most appropriate solution. To meet the business requirements for security, you should implement multiple subsites.

Lesson 2
Planning Security for a SharePoint Server BI Solution

Contents:

Question and Answers

Configuring Kerberos. Put the following steps in order by numbering each to indicate the correct order.

	Steps
	List the computer name, FQDN, and port numbers of all servers involved in the business intelligence solution.
	For each server, note the BI service applications that are running on that server, and the associated service account for each service application.
	Determine which SPNs to use, including mandatory SPNs and custom SPNs.
	Use SetSPN to register the services in Active Directory.
	Define a path of authentication from client machined to back-end data sources to determine which accounts you need to delegate.
	Use the Active Directory Users and Computers tool to delegate the necessary accounts.

Answer:

	Steps
1	List the computer name, FQDN, and port numbers of all servers involved in the business intelligence solution.
2	For each server, note the BI service applications that are running on that server, and the associated service account for each service application.
3	Determine which SPNs to use, including mandatory SPNs and custom SPNs.
4	Use SetSPN to register the services in Active Directory.
5	Define a path of authentication from client machined to back-end data sources to determine which accounts you need to delegate.
6	Use the Active Directory Users and Computers tool to delegate the necessary accounts.

Demonstration: A Kerberos Planning Example

Demonstration Steps

Identify services and service accounts

1. Ensure that the 20467D-MIA-DC and 20467D-MIA-SQL virtual machines are both running, and then log on to 20467D-MIA-SQL as **ADVENTUREWORKS\Student** with the password **Pa$$w0rd**.

2. In the D:\Demofiles\Mod09 folder, double-click **Kerberos.pptx** to open it in PowerPoint, and then in the ribbon, on the **SLIDE SHOW** tab, click **From Beginning**. If a message notifying you that the graphics card is not configured properly is displayed, click **OK**.

3. View the first slide and note that in this example scenario, users will access data in a data warehouse by viewing SQL Server Reporting Services reports in a SharePoint site. However, the identity of the individual users who access the data must be audited for compliance reasons.

4. Click the slide to advance to the next slide and note that the FQDN of the host for each service in the scenario has been recorded, along with the type of service, the port number, and service account that the service runs under.

Define service principal names

1. Click the slide to advance to the next slide, which shows an empty command prompt. Now that the details for each service have been identified, you can use the SetSPN command to create service principal names for the services.

2. Click the slide. Two SetSPN command are shown for the HTTP service (the SharePoint site). Note that a command is executed for the NETBIOS name of the host, and another is executed for the FQDN of the host. The port is not specified because in this example, the site is accessed through the default port (80). If the site was configured to use a non-default port, the same commands would be executed and two more commands with the port appended to the host names would also be required.

3. Click the slide. Two SetSPN commands for the SSRS service (Reporting Services) are shown.

4. Click the slide again to show two SetSPN commands for the MSSQLSvc (SQL Server) service hosting the SharePoint content database.

5. Click the slide again to show two SetSPN commands for the MSSQLSvc (SQL Server) service hosting the data warehouse.

Identify delegation requirements

1. Click the slide to advance to the next slide.

2. Note that the SPSvc account used by the SharePoint Server site must delegate authentication to the following service accounts:

    o   The SSRSSvc account for the SSRS service. This enables the SharePoint site to pass user credentials to Reporting Services for authentication.

    o   The SQLSPSSvc account for the MSSQLSvc service. This enables the SharePoint site to pass user credentials to the SharePoint content database.

3. Note that the SSRSSvc account used by Reporting Services must delegate authentication to the SQLDWSvc account used by the MSSQLSvc service for the database server that hosts the data warehouse.
    To enable the required delegation, the service accounts must be configured in Active Directory users and Computers.

Configure delegation

1. Click the slide to advance to the next slide, which shows the **Delegation** tab of the properties page for the SSRSSvc Windows account in Active Directory Users and Computers (for the purposes of this example, assume that the SPSSvc account has already been configured).

2. Note that the account has been trusted for delegation to specific services.

3. Click the **Add** button on the slide to display the **Add Services** dialog box, which is used to specify the services that the account can delegate to.

4. Click the **Users or Computers** button to display the **Select Users or Computers** dialog box, in which the service accounts for the services to which authentication will be delegated have been selected – in this case, the SQLDWSvc account for the MSSQLSvc service on the data warehouse server.

5. Click **OK** to close the **Select Users or Computers** dialog box, and note that the services associated with the selected account are displayed and can be selected.

6.   Click **OK**, and note that when the list of services is expanded, it includes all instances of the services –
     including NETBIOS and FQDN host names for each service that was selected.

7.   Press **Esc** to exit the slide show, and then close PowerPoint without saving any changes.

Lesson 3
Planning Reporting Services in SharePoint Mode

Contents:

Question and Answers

Question: Which one of the following is the correct way to set up Reporting Services in SharePoint Integrated mode in a multi-tier topology?

(   ) Install Reporting Services in SharePoint Integrated mode on the application servers and the Reporting Services add-in for SharePoint on the WFE servers.

(   ) Install Reporting Services in SharePoint Integrated mode on the WFE servers and the Reporting Services add-in for SharePoint on the application servers.

(   ) Install both Reporting Services in SharePoint Integrated mode and the Reporting Services add-in for SharePoint on the application servers.

(   ) Install both Reporting Services in SharePoint Integrated mode and the Reporting Services add-in for SharePoint on the WFE servers.

> **Answer:** The correct answer is to install Reporting Services in SharePoint Integrated mode on the application servers and the Reporting Services add-in for SharePoint on the WFE servers. In a multi-tier topology, you should install the components in separate tiers. Installing Reporting Services in SharePoint Integrated mode on the WFE servers and the Reporting Services add-in for SharePoint on the application servers would result in an incorrect configuration. In a single tier topology, you would install both components on the same server.

Lesson 4
Planning PowerPivot for SharePoint

Contents:

Question and Answers

Question: What would be the most appropriate times to allow data refreshes to occur in your own work environment? What procedures and events would you have to consider when configuring data refresh?

Answer: Answers will vary depending upon students' own experiences.

Lesson 5
Planning for PerformancePoint Services

Contents:

Question and Answers

Question: You need to create a simple graphical representation that enables you to view sales performance against the defined business targets. Which one of the following PerformancePoint elements best meets your requirements?

( ) A dashboard

( ) A scorecard

( ) A KPI

( ) A report

> **Answer:** For the simple requirements specified, a KPI would be most suitable. KPIs display performance against targets in an easy to understand graphical fashion.

Module Review and Takeaways

Question: Now that you are familiar with the capabilities that SharePoint Server brings to a BI project, what considerations would there be for implementing SharePoint as part of a BI project in your organization?

Answer: There are a variety of factors to consider, including:

- The cost of licensing.

- The complexity of planning, installing, and configuring a SharePoint Server farm.

- The in-house skills to support SharePoint Server.

- The edition of SharePoint Server being used.

Lab Review Questions and Answers

Lab A: Implementing a SharePoint Server BI Solution

Question and Answers

Lab Review

Question: What is the benefit of creating a subsite in SharePoint Server for storing business intelligence content? Is it a mandatory process?

> **Answer:** It is not a mandatory requirement to create a separate subsite to store business intelligence content. However, creating a subsite makes sense for the following reasons:
>
> - It allows separation of BI content from non-BI content.
>
> - It makes it easier to secure BI content.
>
> - It enables BI applications to be self-contained in on their own sites.

Students may offer other examples; it is important to explore the answer to promote a discussion.

Lab B: Implementing PerformancePoint Services

Question and Answers

Question: You could implement a similar dashboard in the site by using Reporting Services or Excel Services instead of PerformancePoint Services? What are the benefits and disadvantages of using PerformancePoint Services?

> **Answer:** Answers will vary depending on students' familiarity with PerformancePoint Services. The key benefits of PerformancePoint Services is that it can be used with multiple data sources and includes support for rich, expression-based filtering and highly customized KPIs, scorecards, and reports. It also enables you to create interactive dashboards that integrate well into a SharePoint Server web site interface.
>
> However, if the KPIs and reports you want to display are already defined in SQL Server Analysis Services and SQL Server Reporting Services, it may be easier to create Reporting Services or Excel Services dashboards – especially in organizations where SQL Server skills are available but PerformancePoint Services skills are lacking.

Module 10

Monitoring and Optimizing a BI Solution

Contents:

Lesson 1
Overview of BI Monitoring

Contents:

Question and Answers

Question: When creating a performance baseline, which of the following answers best describes the workloads you should monitor?

( ) Data warehouse load operations, data model processing operations, and peak-time query activity.

( ) Data warehouse load operations, data model processing operations, normal everyday query activity, and peak-time query activity.

( ) Data warehouse load operations, normal everyday query activity, and peak-time query activity.

( ) Data warehouse load operations, data model processing operations, and normal everyday query activity.

( ) Normal everyday query activity and peak-time query activity.

> **Answer:** To create a baseline, you should monitor data warehouse load operations, data model processing operations, normal everyday query activity, and peak-time query activity.

Lesson 2
Monitoring and Optimizing the Data Warehouse

Contents:

Question and Answers

What is the recommended order for reviewing performance metrics when troubleshooting data warehouse performance? Put the following steps in order by numbering each to indicate the correct order.

	Steps
	Processor
	Memory
	Storage (disks)

Answer:

	Steps
1	Processor
2	Memory
3	Storage (disks)

Lesson 3
Monitoring and Optimizing Analysis Services

Contents:

Question and Answers

Question: Which of one the following answers best describes the recommended way of using the Flight Recorder tool?

( ) Leave the Flight Recorder tool running at all times so that you capture all available data.

( ) Run the Flight Recorder periodically during periods of heavy usage of Analysis Services.

( ) Enable Flight Recorder only when you have a specific problem that you need to troubleshoot, and then disable it again afterwards.

( ) Run the Flight Recorder periodically during periods of light usage of Analysis Services.

> **Answer:** You should enable Flight Recorder only when you have a specific problem that you need to troubleshoot, and then disable it again afterwards. Flight Recorder can have a negative impact on performance because it competes with Analysis Services for resources, so leaving it running all the time would degrade performance, as would running it during periods of heavy usage. Running Flight Recorder during period of light usage would have less of an impact on performance, but the data it collects would be of very limited use.

Resources

Considerations for Analysis Services Performance

Additional Reading: For more information about optimizing Analysis Services, download the e-book "Microsoft SQL Server Analysis Services Multidimensional Performance and Operations Guide" from the Microsoft Download Center.

Monitoring Tools for Analysis Services

Reference Links: For more information about the DMVs that you can use to monitor Analysis Services, view the article 'Use Dynamic Management Views (DMVs) to Monitor Analysis Services' at https://msdn.microsoft.com/en-us/library/hh230820(v=sql.120).aspx

Troubleshooting Analysis Server Queries

Additional Reading: For more information about diagnosing query performance problems, see "Analysis Services MOLAP Performance Guide for SQL Server 2012 and 2014" at https://msdn.microsoft.com/en-us/library/dn749781.aspx.

Lesson 4
Monitoring and Optimizing Reporting Services

Contents:

Question and Answers

Question: Which of the following RSReportServer.config file settings can you use to specify the percentage at which Reporting Services considers the level of memory pressure to be medium?

( ) WorkingSetMinimum

( ) MemoryThreshold

( ) MemorySafetyMargin

( ) WorkingSetMaximum

> **Answer:** To specify the percentage at which Reporting Services considers the level of memory pressure to be medium, you should use the MemorySafetyMargin option. When you define this, if memory usage reaches the value that you specified, Reporting Services will begin to refuse some requests for memory reduce memory allocations.

Resources

Considerations for Reporting Services Performance

Reference Links: For more information about modifying the RSReportServer.config file, see the article 'Modify a Reporting Services Configuration File (RSReportServer.config) at https://msdn.microsoft.com/en-us/library/bb630448(v=sql.120).aspx

Module Review and Takeaways

Question: In this module, you have considered the components of SQL Server that must be monitored and optimized in a BI solution. What other elements of the solution should you monitor and troubleshoot in the event of performance problems?

Answer: Answers will vary, but they might include:

- Windows event logs to detect service configuration problems or hardware failures.

- SharePoint logs and the SharePoint Server Health Analyzer in SharePoint Central Administration.

- The PowerPivot Management Dashboard in SharePoint Central Administration.

Lab Review Questions and Answers

Lab: Monitoring and Optimizing a BI Solution

Question and Answers

Lab Review

Question: How might the classifier function you would create to prioritize ETL workloads in a real solution differ from the one used in the lab?

> **Answer:** Most ETL operations are scheduled, so the classifier function would probably filter based on the current time as well as the application name. Additionally, ETL operations are likely to run in the context of a dedicated service account, so the system user name could also be used in the filter.

Module 11

Operating a BI Solution

Contents:

Lesson 1
Overview of BI Operations

Contents:

Question and Answers

There are several considerations for scheduling operational tasks, including data latency tolerance, data flow dependencies, and the available operational window. Match each the following descriptions to the consideration that it best describes. Indicate your answer by writing the category number to the right of each item.

Items	
1	The period during which all tasks, such as data loads, index rebuilds, and cube processing, must complete.
2	How out of date the data that information workers use is allowed to be.
3	The way that operational tasks relate to each other, which determines the order in which the tasks need to run.

Category 1		Category 2		Category 3
Operational window		Data latency tolerance		Data flow dependencies

Answer:

Category 1		Category 2		Category 3
Operational window		Data latency tolerance		Data flow dependencies
The period during which all tasks, such as data loads, index rebuilds, and cube processing, must complete.		How out of date the data that information workers use is allowed to be.		The way that operational tasks relate to each other, which determines the order in which the tasks need to run.

Resources

Automating Operational Tasks

Additional Reading: To learn more about automating database maintenance tasks with the SQL Server Agent, you can attend Course 20462: *Administering Microsoft SQL Server 2014 Databases*.

Lesson 2
ETL Operations

Contents:

Question and Answers

Question: You plan to create an SSIS package that you want to execute in both a test environment and in a production environment on two different SQL Server 2014 instances. You intend to use the project deployment model. What is the best way to achieve this?

( ) Create a package and deploy it. Create one environment for the test scenario and one environment for the production scenario, create a single variable for the server name, and then assign the variable the correct server name in each environment.

( ) Create a package and deploy it. Create a single environment for both the test scenario and the production scenario to use, create a single variable for the server name, and then assign the variable the correct server name.

( ) Create a package and deploy it. Create one environment for the test scenario and one environment for the production scenario, create two variables, one for each of the server names, and then assign each variable the correct server name for each environment.

( ) Create a package and deploy it. Create a single environment for both the test scenario and the production scenario to use, create two variables, one for each of the server names, and then assign each variable the correct server name for each environment.

> **Answer:** As you learnt in the topic Environments and Variables in this lesson, the correct way to achieve the desired result is to create one environment for the test scenario and one environment for the production scenario, create a single variable for the server name, and then assign the variable the correct server name in each environment.

Resources

Backing Up the SSIS Catalog

Additional Reading: For step by step instructions on how to restore the SSIS catalog on a new server, see "Backup, Restore, and Move the SSIS Catalog" in SQL Server Books Online.

Lesson 3
Data Warehouse Operations

Contents:

Question and Answers

Question: You are responsible for planning the backup strategy for a SQL Server 2014 medium-sized data warehouse that is a data source for BI applications. The database uses the simple recovery model, has a single filegroup for data, and it is not partitioned. The operational window for backing up the database is relatively short. Which backup strategy is the most appropriate for this scenario?

(   ) Back up the database periodically by using the FULL backup option only.

(   ) Use a partial backup strategy.

(   ) Back up the database periodically by using the FULL backup option, and backup the database log to capture the changes between each FULL backup.

(   ) Back up the database periodically by using the FULL backup option, and use the DIFFERENTIAL backup option to capture the changes between each FULL backup.

> **Answer:** The most appropriate strategy is to use a FULL plus DIFFERENTIAL strategy. Using FULL backups only would probably take too long given that there is a limited operational window in which to perform the backups. A partial backup strategy requires multiple filegroups, and it is not possible to back up the log when a database uses the simple recovery model.

Resources

🌐   **Additional Reading:** For more information about monitoring database server health, attend Course 20462: *Administering Microsoft SQL Server 2014 Databases*

Managing Partitions in the Data Warehouse

🌐   **Additional Reading:** For information about how to automate a sliding window archival process, see "How to Implement an Automatic Sliding Window in a Partitioned Table on SQL Server 2005" at http://msdn.microsoft.com/en-us/library/aa964122(SQL.90).aspx. Although this paper describes partitioning in SQL Server 2005, the same principles apply in SQL Server 2014.

Demonstration: Using a Partial Backup Strategy

Demonstration Steps

Implement a partial backup strategy

1. Ensure that the 20467D-MIA-DC and 20467D-MIA-SQL virtual machines are both running, and then log on to 20467D-MIA-SQL as **ADVENTUREWORKS\Student** with the password **Pa$$w0rd**.

2. In the D:\Demofiles\Mod11 folder, right-click **Setup.cmd** and click **Run as administrator**. When prompted, to allow the program to make changes, click **Yes**.

3. Start SQL Server Management Studio and connect to the **localhost** instance of the database engine by using Windows authentication.

4. Open **Partial Backup.sql** from the D:\Demofiles\Mod11 folder.

5. Select the code under the comment **View filegroups and partitions**, and then click **Execute**. This retrieves information about the filegroups and partitions in the **DemoDW** database.

6. Select the code under the comment **Make inactive filegroups read-only**, and then click **Execute**. This marks the filegroups containing partitions with inactive data as read-only.

7. Select the code under the comment **Backup inactive filegroups**, and then click **Execute**. This backs up the read-only filegroups.

8. Select the code under the comment **Backup read/write filegroups**, and then click **Execute**. This backs up the read/write filegroups.

9. Select the code under the comment **Perform a data load** (from line 45 to line 74), and then click **Execute**. This performs ETL tasks to load a new row into a dimension table and a new partition in a fact table.

10. Select the code under the comment **Make loaded filegroup read-only and back it up** and then click **Execute**. This marks the newly loaded filegroup as read-only and backs it up.

11. Select the code under the comment **Perform a differential backup of read/write filegroups** and then click **Execute**. This creates a differential backup of the read/write filegroups.

Perform a piecemeal restore

1. Select the code under the comment **Simulate a disaster** and then click **Execute**. This drops the DemoDW database.

2. Select the code under the comment **Restore the initial full backup with the partial option** and then click **Execute**. This restores the original full backup and specifies that the database will be recovered from partial backups.

3. Select the code under the comment **Restore the read/write filegroups and recover** and then click **Execute**. This restores the differential backup of the read/write filegroups.

4. Select the code under the comment **Access read/write data** and then click **Execute**. This queries the dimension table, which is on the PRIMARY filegroup and has been recovered.

5. Select the code under the comment **Restore the read-only filegroups** (from line 118 to line 141) and then click **Execute**. This restores the read-only filegroups, which were backed up individually.

6. Select the code under the comment **Access read-only data** and then click **Execute**. This queries the fact table to verify that it has been restored.

7. Close SQL Server Management Studio and do not save any changes.

Lesson 4
Analysis Services Operations

Contents:

Question and Answers

Question: You have used the Process Data option to process a dimension table in a tabular data model. What should you do now to ensure that users can query the table? You want to enable users to have access the table as quickly as possible.

( ) Run a Process Full operation.

( ) Run a Process Defrag operation.

( ) Run a Process Recalc operation.

( ) Run a Process Clear operation.

( ) It is not necessary to perform any further operations.

> **Answer:** You should run a Process Recalc operation after performing a Process Data operation. Process Data does not perform recalculations, so it may not be possible to query data models if you omit the Process Recalc step. Process Full would perform the recalculations, but it also performs additional tasks and so would probably not be as efficient as the Process Recalc operation. It is not necessary to perform a Process Defrag operation when you use the Process Data option on a table, because Process Data builds new dictionaries when it processes the table. Process Clear deletes data, so users would not be able to query the table and obtain any meaningful results.

Resources

Processing Data Models

Additional Reading: For more information about choosing a processing mode for multidimensional data models, see "Analysis Services Processing Best Practices" at http://msdn.microsoft.com/en-US/library/cc966525.

Lesson 5
Reporting Services Operations

Contents:

Question and Answers

Question: You have moved the server that hosts a native mode instance of SQL Server 2014 Reporting Services, and the server now resides in a different time zone. You have set the time zone on the server, and restarted the Reporting Services service. What else must you do to ensure that Reporting Services registers the new time zone so that scheduled tasks run correctly (choose one answer)?

( ) Restart SQL Server Agent.

( ) Manually update the scheduled tasks.

( ) Manually update the time zone in Reporting Services

( ) No further action is necessary.

> **Answer:** A Reporting Server picks up the time zone from the host computer when you restart the Reporting Services service, and you cannot update the time zone in Reporting Services. There is no need to restart SQL Server Agent, or manually update the scheduled tasks to force Reporting Services to register the new time zone.

Module Review and Takeaways

Question: As a BI specialist, your involvement in a BI solution may end when you deploy the solution into production. How can you ensure that the IT personnel who will support the solution are able to manage and troubleshoot the necessary operational tasks?

> **Answer:** Answers will vary, but a key requirement is to document all operational tasks (automated and manual) and all information about schedules, user accounts and permissions, and the planned disaster recovery processes.

Lab Review Questions and Answers

Lab: Operating a BI Solution

Question and Answers

Lab Review

Question: How might the operations solution you created in the lab have differed if the measure groups in the cube were partitioned on the same basis as the fact tables in the relational database?

> **Answer:** If the measure groups in the cube were partitioned on the same basis as the fact tables that they are based on, you would need to use an Analysis Services command to create a new partition in each measure group based on the same partitioning criteria as the relational tables.
>
> The complexity of determining the required new relational table partitions and measure group partitions might lead you to decide to implement the required partitioning and processing tasks in as SSIS package instead of using an XMLA script in a SQL Server Agent job step. An SSIS package provides a more flexible platform for using variables and procedural logic than a SQL Server Agent job.

Question: If the volume of data to load and process was significantly larger, or the time period available for performing the ETL load was shorter, how might you change the solution that you created in the lab?

> **Answer:** To reduce the time taken for the load, you might not update indexes and statistics with every load, and instead perform these tasks on a separate schedule that occurs at a different time. You might see a more significant reduction in the time required to process the cube with large volumes of data if you perform incremental processing of the dimensions and measure group partitions instead of processing the entire cube.