

MIS 420 Exam #2 Study guide

This document is used to guide your studying for MIS 420 exam #2. While it is useful to guide your studies, this document does not present an exact list or contract regarding exam contents. Rather this document should be used to guide your preparation, by highlighting the topics that are deemed important. To provide your indepth answers to this study guide by reading and interpreting the couse modules. Upload your study guide when taking your exam.

As this is a 400-level class you need to do more than prepare 3 sentence answers to these questions. Depth, insight, and perspective is required to earn higher grades. You just have to closely examine the module documentation to locate and craft your understanding and response. Go for deep understanding rather than surface recitation of content. You are warned NOT to find data on Google to answer the SQL questions, as you may read incorrect or unuseful data which can harm your grade. **Use scenarios to demonstrate your understanding.**

Module 5 - Sub-queries

1. What is a sub-query, how can you tell if your query has a sub-query? **What were some of the problems mentioned in the module that were solved by using sub-queries**
2. It was demonstrated that it is hard to join different fact tables together. When the business requirement is to create calculated columns of analytics from different fact tables, how did sub-queries allow this functionality, where prior GROUP BY() could not?
3. Explain the process and write an example of a SQL query that has a dimension table attribute and uses sub-queries to pull and aggregate **data from fact tables** into columns **in your dataset.**
4. How were sub-queries used to enhance a WHERE clause filter? Write a SQL query example.
5. What is **your** preferred method, joining tables to access data and creating more columns of metrics, or using sub-queries to add more columns of metrics. Discuss and justify your answer.
6. What is the difference in the rows you would receive from a dimension and fact table, amongst INNER JOINS, LEFT JOINS and FULL JOINS, discuss an example. Was there a problem with FULL JOINS?
7. **What does SELECT (DISTINCT([Country])) produce? When is code like this it needed?**
8. **What does SELECT COUNT(DISTINCT([Country])) produce? When is code like this it needed?**
9. Write an example SQL subquery
10. Describe what the following DAX code produces:
SUMX(RelatedTable(factResellerSales, [Quantity]))
What SQL functionality is the DAX code similar to?
11. SQL Programming – write a set of SQL Statements that declare a numeric decimal variable, gives it a value of 10, changes the value of the variable by tripling it, and then displays the final answer.

Module 6 - Arrays

1. Recap the history and purpose of using SQL table arrays in data management. Why is data moved from secondary storage into primary RAM arrays to process data, only to later copy array data back to database tables?
2. A) Describe how SQL arrays are used in ETL processes, include the benefits from doing so, and give some examples of how they are used. Give an original scenario that demonstrates your understanding.

B) Describe how SQL arrays are used to build datasets for analytics, include the benefits from doing so, and give some examples of how they are used. Give an original scenario that demonstrates your understanding.

New content was added to pages 4 and 5 of module 6, review that to answer these questions

3. A) Mod 6 documentation presents an example of how to copy tables of data from one database (Featherman_analytics) to another (MF01***) using the SELECT * INTO [yourdatabase] FROM syntax. Be able to write code like this.

Write an original SQL statement that uses this SELECT * INTO FROM data processing to copy rows of transaction data from a fact table to archives (a second database table with the same schema).

B) The SQL pattern INSERT INTO SELECT * FROM was used to copy rows of data into an array. An example is presented. Write your understanding of the following code from page 9 of module 6.

```
INSERT INTO @ArrayTable2
(SELECT Column1, Count(Column2)
FROM SomeDatabaseTable
GROUP BY Column1)
```

C) Create a table in your word processing document. Write a few rows of data that could be the output from this query.

D) Write your explanation of each line of code 1-10.

```
1. USE [AdventureWorksDW2012];
2. DECLARE @Sales TABLE ([Product] int, [Reseller Units] decimal, [Inet Units] decimal )
3. INSERT INTO @Sales ([Product], [Reseller Units] )
4. SELECT [ProductKey], SUM([OrderQuantity])
5. FROM [dbo].[FactResellerSales]
7. GROUP BY [ProductKey]
8. UPDATE @Sales SET [Inet Units] = (SELECT SUM([OrderQuantity])
9. FROM [dbo].[FactInternetSales]
10. WHERE [dbo].[FactInternetSales].[ProductKey] = [Product] ).
```

4. Write code that can create an array and fill its columns
5. What is a UNION SELECT query? What is it used to accomplish? Does it add more rows or more columns? What is a pre-requisite to use this type of query?
6. How was the UNION SELECT query used in module 6? Give a scenario.

7. What functionality limitations were solved with array tables in module 6?
8. If data is saved into database tables, why are array tables needed?
9. What is the alchemy of a business analyst, and how is this related to arrays? What are arrays considered to be a canvas to paint data upon?
10. Regarding calculation and presentation of metrics, describe when to use the following data type in array based calculations - decimal, decimal(5,2), nvarchar(6), int
11. How are UPDATE SET commands used inside arrays? Give a scenario.
12. Be able to write different UPDATE SET commands and CASE statements in combination.
13. What are some unique aspects and benefits of using SQL arrays as compared to other SQL we have used?

14. Write your explanation of the code to the right. Why would this be useful?
Why are some SQL words purplish or blue in the example provided here?

```
UPDATE @Sales SET [Reseller %] =
(CASE
WHEN [Reseller Profit] > 0 THEN ([Reseller Profit]/ [Total Profit])
WHEN [Reseller Profit] <=0 THEN 0
END)
```

15. Write your original example of a CASE statement being used inside an UPDATE array SET column = ().

Why is this type of code useful?

16. Refer to page 14 of module 6.

a) What is the granularity of the array? Write the line of code that tells you the granularity level.

b) for the first INSERT INTO array SELECT FROM GROUP BY code section for the query on Page 14 of Module 6, how many columns of data have data loaded into them?

Why is it useful to condense data before loading it into an array as shown in this first code section.

17. On the same page 14 example, after the INSERT INTO @Array SELECT FROM GROUPBY code section, the coding style changed to an UPDATE SET = (sub-query).

a) provide your explanation of that first UPDATE SET SQL code section, that generates the blue column of Internet Profit

b) give your belief of *why the coding style changed* from INSERT INTO SELECT FROM to UPDATE SET ColumnName = (sub-query)

18. Why is declaring variables in your query routines (whether using an array or not) useful? Give an example of functionality provided and be able to write your own unique example.

Module 7 – Window/Table Functions

1. What are table functions and why are they different than other SQL code that we produced thus far? Give an example of functionality that is solved by using table functions and what the alternative would be if we did not know how to leverage them.
2. What is the major problem solved with lag statements?
3. What are the common analytics output demonstrated in module 6 that are made available by using SQL WINDOW functions? Can these be created in SQL any other way?
4. What is a window and how are they created using the PARTITION BY and OVER statements?
5. Describe the huge importance and usage of LAG statements? What type of analysis do they enable?
6. Recap the eight different examples of window functions in module 7, giving a scenario of a problem that can be solved by each example.
7. Compare the usage of table functions in Tableau with window functions in SQL.
8. Explain how an IIF statement works, the example of how was used in the module. Be able to write your own IIF statement. Should IIF statement be used rather than CASE statements?
9. What is month-over-months, quarter-over-quarter, and year over year analysis, and what analysis of business performance can be understood from using these analyses? Give an original example.
10. What does the pseudo code look like to create a year over year % change column of metrics?
11. What is NTILE() useful to accomplish? How would you accomplish similar functionality, if you did not use NTILE?
12. What DAX functions provide similar functionality to LAG functions? Write one of these formulas.

Module 8 – Stored Procedures

1. What are the benefits of using stored procedures?
2. Why are stored procedures central to DA/Analyst productivity?
3. Why would a firm switching from Excel-based analytics to SQL/PowerBI analytics appreciate use of stored procedures?
4. Recap the business case of using stored procedures.
5. Discuss the usage of stored procedures in ETL vs analytics.