Final Project for CIMT 611

Haisong Ye

Summer I 2013

Indiana State University

Dr. Larry Tinnerman

Introduction

This final project is to provide assessment guidelines for an instructional unit. The project includes organizational standards, learning targets, a sample plan of assignments and assessments, an assessment sample, rubric for the assessment sample, and item analysis sheet. The learning goal of the unit is that learners can construct correct SQL query statement to select required data from a relational database.

Standards

The following standards are selected from a course of the Microsoft certifications program (Microsoft, 2013b). The course is 70-461: Querying Microsoft SQL Server 2012. The selected unit is Work with Data (Microsoft, 2013a).

- Query data by using SELECT statements
- Use the ranking function to select top(X) rows for multiple categories in a single query; write and perform queries efficiently using the new (SQL 2005/8->) code items such as synonyms, and joins (except, intersect); implement logic which uses dynamic SQL and system metadata; write efficient, technically complex SQL queries, including all types of joins versus the use of derived tables; determine what code may or may not execute based on the tables provided; given a table with constraints, determine which statement set would load a table; use and understand different data access technologies; case versus isnull versus coalesce
- Implement sub-queries
- Identify problematic elements in query plans; pivot and unpivot; apply operator; cte statement; with statement

- Implement data types
- Use appropriate data; understand the uses and limitations of each data type; impact of GUID (newid, newsequentialid) on database performance, when to use what data type for columns
- Implement aggregate queries
- New analytic functions; grouping sets; spatial aggregates; apply ranking functions

Learning Targets

Instructional Unit: Working with SQL query statement (learning targets with bold fonts and underline are mentioned in the standards).

1. Knows typical data types in SQL
    1.1. Defines each data type in own words
    1.2. Identifies possible data types for a given data
    1.3. Determines which data types are best for a given situation
    1.4. Converts data to different data types
2. Knows keywords in SQL statement
    2.1. Identifies each keyword in SQL statement
    2.2. Defines the function of each keyword
    2.3. Determines which keyword is best for a given situation
3. Implements simple data query
    3.1. Constructs data query in one table
    3.2. Constructs data query from multiple tables
    3.3. Constructs a query with sub-query

3.4. Uses ranking functions in query

4. Filters data

4.1. Filters data by using logical operations

4.2. Filters data by using built-in functions

4.3. Filters data by using multiple conditions

5. Applies aggregate clauses in data query

5.1. Explains the reasons to group data

5.2. Describes the function of each aggregate clause

5.3. Aggregates data in query

Assignments and Assessments Plan

Table 1

Table of assignment and assessments

	Pretesting	Quizzes	Unit test
Time	Beginning of unit	Periodically during instruction	End of unit
Learning targets	Prior knowledge in relational database	Learning targets of each lesson	Unit learning targets
Assessment type	Formative	Formative	Summative

Table 2

Table of specifications for summative assessment (percentage of points)

	Knows	Understand	Application	Total percentage of

				points
Data types	4	6	6	16
Keywords in SQL	4	6	6	16
Data query	6	8	10	24
Data filtering	4	8	10	22
Aggregation	4	8	10	22
Total Percentage of points	22	36	42	100

Assessment

Unit Test (90 minutes)

I. Multiple Choice Questions (30 points, 2 points each)

Direction: Please choose one best answer for each of following questions.

1. ( ) Why is it important to use the appropriate type for attributes?
   A. Because the type of your attribute enables you to control the formatting of the values
   B. Because the type constrains the values to a certain domain of supported values
   C. Because the type prevents duplicates
   D. Because the type prevents NULLs
2. ( ) Which one of the following data types can store binary data?
   A. numeric
   B. text
   C. image
   D. bit
3. ( ) Which statement is **NOT** true?
   A. The default precision of decimal type in SQL Server 2012 is 18.
   B. money and smallmoney are used to represent monetary or currency values
   C. nchar can store more data than char.
   D. nchar can store unicode character strings

4. ( ) What are the mandatory clauses in a SELECT query, according to T-SQL?C
   A. The FROM and SELECT clauses
   B. The SELECT and WHERE clauses
   C. The SELECT clause
   D. The FROM and WHERE clauses
5. ( ) Which of the following correctly represents the logical query processing order of the various query clauses? A
   A. SELECT > FROM > WHERE > GROUP BY > HAVING > ORDER BY
   B. FROM > WHERE > GROUP BY > HAVING > SELECT > ORDER BY
   C. FROM > WHERE > GROUP BY > HAVING > ORDER BY > SELECT
   D. SELECT > ORDER BY > FROM > WHERE > GROUP BY > HAVING
6. ( ) Which of the following practices are considered bad practices? (Choose all that apply.)
   A. Aliasing columns by using the AS clause
   B. Aliasing tables by using the AS clause
   C. Using [] to embrace table name or column name
   D. Using * in the SELECT list
7. ( ) What is the difference between the simple CASE expression and the searched CASE expression?
   A. The simple CASE expression is used when the database recovery model is simple, and the searched CASE expression is used when it's full or bulk logged.
   B. The simple CASE expression compares an input expression to multiple possible expressions in the WHEN clauses, and the searched CASE expression uses independent predicates in the WHEN clauses.
   C. The simple CASE expression can be used anywhere in a query, and the searched CASE expression can be used only in the WHERE clause.
   D. The simple CASE expression can be used anywhere in a query, and the searched CASE expression can be used only in query filters (ON, WHERE, HAVING).
8. ( ) What does the term three-valued logic refer to in T-SQL?
   A. The three possible logical result values of a predicate : true, false, and NULL
   B. The three possible logical result values of a predicate : 1, 0, and NULL

C. The three possible logical result values of a predicate : true, false, and unknown

D. The three possible logical result values of a predicate : -1, 0, and 1

9. ( ) Which of the following predicates are search arguments?

A. DAY(orderdate) = 1

B. companyname LIKE 'A%'

C. companyname LIKE '%A%'

D. companyname LIKE '%A'

10. ( ) When a query doesn't have an ORDER BY clause, what is the order in which the rows are returned?

A. Arbitrary order

B. Primary key order

C. Clustered index order

D. Insertion order

11. ( ) You want result rows to be sorted by orderdate descending, and then by orderid, descending. Which of the following clauses gives you what you want?

A. ORDER BY orderdate, orderid DESC

B. ORDER BY DESC orderdate, DESC orderid

C. ORDER BY orderdate DESC, orderid DESC

D. DESC ORDER BY orderdate, orderid

12. ( ) What is the restriction that grouped queries impose on your expressions?

A. If the query is a grouped query, you must invoke an aggregate function.

B. If the query has an aggregate function, it must have a GROUP BY clause.

C. The elements in the GROUP BY clause must also be specified in the SELECT clause.

D. If you refer to an element from the queried tables in the HAVING, SELECT, or ORDER BY clauses, it must either appear in the GROUP BY list or be contained by an aggregate function.

13. ( ) What is the difference between the COUNT (*) aggregate function and the COUNT (<*expression*>) general set function?

A. COUNT(*) counts rows; COUNT(<*expression*>) counts rows where <*expression*> is

B. Not NULL.

C. COUNT (*) counts columns; COUNT (<*expression*>) counts rows.

D. COUNT (*) returns a BIGINT; COUNT (<*expression*>) returns an INT.
E. There is no difference between the functions.

14. ( ) What do the RANK and DENSE_RANK functions compute? B
A. The RANK function returns the number of rows that have a lower ordering value (assuming ascending ordering) than the current; the DENSE_RANK function returns the number of distinct ordering values that are lower than the current.
B. The RANK function returns one more than the number of rows that have a lower ordering value than the current; the DENSE_RANK function returns one more than the number of distinct ordering values that are lower than the current.
C. The RANK function returns one less than the number of rows that have a lower ordering value than the current; the DENSE_RANK function returns one less than the number of distinct ordering values that are lower than the current.
D. The two functions return the same result unless the ordering is unique.

15. ( ) Why are window functions allowed only in the SELECT and ORDER BY clauses of a query?
A. Because Microsoft didn't have time to implement them in other clauses.
B. Because you never need to filter or group data based on the result of window functions.
C. Because they are supposed to operate on the underlying query's result, which is achieved when logical query processing gets to the SELECT phase.
D. Because in the other clauses, the functions are considered door functions (also known as backdoor functions).

II. Short Answers (35 points)

Direction: Please answer the questions according to the given scenario.

One company has a database to store their employees' information. There are three tables related to employment,

Person (PersonId (int), FirstName (varchar), LastName (varchar))

Department (DepartmentId (int), Name (varchar))

EmployeeDepartmentHistory (PersonId (int), DepartmentId (int), StartDate (date))

According to the three tables, please construct ONE SQL query statement for each of following questions

1. Select all persons' information in table person. (5 points)

2. Select all persons who are/were in "Engineering" department. (5 points)

3. Select all persons who are in "Engineering" department and whose name starts with "J". (5 points)

4. Select all persons with their current department name and start date. (10 points)

5. Count the number of employees in each department now and listed in decreased order. (10 points)

III. Essay questions (35 points)

Direction: Please answer the following questions with no more than 150 words.

FINAL PROJECT

1. Provide examples for queries with deterministic and nondeterministic ordering. Describe in your own words what is required to get deterministic ordering. (10 points)

2. Describe the difference between ROW_NUMBER and RANK. (10 points)

3. You are presenting a session about set operators in a conference. At the end of the session, you give the audience an opportunity to ask questions. Answer the following questions presented to you by attendees:

1). In our system, we have a number of views that use a UNION operator to combine disjoint sets from different tables. We see performance problems when querying the views. Do you have any suggestions to try and improve the performance? (7 points)

2). Can you point out the advantages of using set operators like INTERSECT and EXCEPT compared to the use of inner and outer joins? (8 points)

Analysis of Item Formats

The assessment uses objective items (multiple choice), short answer questions, and essay questions. The form of short answer question in this assessment is a complex version. It is adapted from a combination of short answer and interpretive exercise. The performance based items are not applied in the assessment, because the learning targets focus on products more than process in this unit. The learning goal of this unit is that learners can construct correct SQL query statement to select required data from a relational database. It does not require learners to implement a data query in a particular database system and there are no any hands-on labs in this unit. Therefore, there is no skill need to be assessed by performance based item. Questions are selected and adapted from Mata-Toledo & Cushman (2000), Elmasri & Navathe (2007), and Ben-Gan, Sarka, & Talmage (2012).

Rubrics/Grading Criteria

I. Multiple Choice Questions (30)

1) B	2) C	3) C	4) C	5) A
6) D	7) B	8) C	9) B	10) A
11) C	12) D	13) A	14) B	15) C

II. Short Answers (35)

1. Select all persons' information in table person. (5)

*select * from Person*

2. Select all persons who are/were in *Engineering* department. (5)

*Select * from Person inner join department on person.personid=department.departmentid where name='Engineering'*

3. Select all persons who are in *Engineering* department and whose name starts with *J*. (5)

*Select * from Person inner join department on person.personid=department.departmentid where name='Engineering' and name like 'J%'*

4. Select all persons with their current department name and start date. (10)

select p.PersonID,p.FirstName,p.LastName, d.Name, max(edh.StartDate) as StartDate from Person as p inner join HumanResources.EmployeeDepartmentHistory as edh on p.PersonId=edh.PersonId inner join Department as d on edh.DepartmentID=d.DepartmentID

group by p.PersonId, p.FirstName, p.LastName, d.Name

5. Count the number of employees in each department now and listed in decreased order. (10)

Select d.Name, Count(p.PersonID) as Number, max(edh.StartDate) as StartDate from Person as p inner join HumanResources.EmployeeDepartmentHistory as edh on p.PersonId=edh.PersonId

inner join Department as d on edh.DepartmentID=d.DepartmentID

group by d.Name order by Number desc

III. Essay Questions (35)

1. Provide examples for queries with deterministic and nondeterministic ordering. Describe in your own words what is required to get deterministic ordering. (10)

*Nondeterministic ordering is when an ordering column isn't unique and the ordering of rows of the same values is not guaranteed. For example, select * from employees order by city. The city isn't unique, so this is a nondeterministic ordering.*

*Deterministic ordering is when an ordering column is unique and the ordering of rows is guaranteed. For example, select * from employees order by employeeid. The employeeid is unique, so this is a deterministic ordering.*

2. Describe the difference between ROW_NUMBER and RANK. (10)

The ROW_NUMBER function isn't sensitive to ties in the window ordering values. Therefore, the computation is deterministic only when the window ordering is unique. When the window ordering isn't unique, the function isn't deterministic. The RANK function is sensitive to ties and produces the same rank value to all rows with the same ordering value. Therefore, it is deterministic even when the window ordering isn't unique.

3. You are presenting a session about set operators in a conference. At the end of the session, you give the audience an opportunity to ask questions. Answer the following questions presented to you by attendees:

1). In our system, we have a number of views that use a UNION operator to combine disjoint sets from different tables. We see performance problems when querying the views. Do you have any suggestions to try and improve the performance? (7)

The UNION operator returns distinct rows. When the unified sets are disjoint, there are no duplicates to remove, but the SQL Server Query Optimizer may not realize it. Trying to remove duplicates even when there are none involves extra cost. So when the sets are disjoint, it's important to use the UNION ALL operator and not UNION. Also, adding CHECK constraints that define the ranges supported by each table can help the optimizer realize that the sets are disjoint. Then, even when using UNION, the optimizer can realize it doesn't need to remove duplicates.

2). Can you point out the advantages of using set operators like INTERSECT and EXCEPT compared to the use of inner and outer joins? (8)

Set operators have a number of benefits. They allow simpler code because you don't explicitly compare the columns from the two inputs like you do with joins. Also, when set operators compare two NULLs, they consider them the same, which is not the case with joins. When this is the desired behavior, it is easier to use set operators. With join, you have to add predicates to get such behavior.

Item Analysis Sheet

Following is an item analysis sheet for multiple-choice question.

Course _____

Objective _____

Unit _____

Item

Date used	Students	Alternatives					Omits	Difficulty	Discrimination
		A	B	C	D	E			
	Upper 10								
	Lower 10								

Comment:

Following is an item analysis sheet for short answer question or essay question.

Course _____

Objective _____

Unit _____

Item

Date used	Students	Highest Score	Lowest Score	Average Score	Omits	Difficulty	Discrimination
	Upper 10						
	Lower 10						

Comment:

References

Ben-Gan, I., Sarka, D., & Talmage, R. (2012). T*raining Kit (Exam 70-461): Querying Microsoft SQL Server 2012*. Sebastopol, CA: O'Reilly Media, Inc.

Elmasri, R., & Navathe, S. B. (2007). *Fundatmentals of Database System* (5th ed.). Boston, MA: Pearson Addison Wesley.

Mata-Toledo, R. A., & Cushman, P. K. (2000). *Fundamentals of Relational Databases*. New York: McGRAW-HILL.

Microsoft. (2013a). *Exam 70-461 Querying Microsoft SQL Server 2012*. Retrieved from http://www.microsoft.com/learning/en/us/exam-70-461.aspx#fbid=pXCfbRSMQTi

Microsoft. (2013b). *Find the right IT certification*. Retrieved from IT Certification: http://www.microsoft.com/learning/en/us/certification-overview.aspx#fbid=pXCfbRSMQTi