Curriculum Development Centers Program

Awardee of The Office of the National Coordinator for Health Information Technology

Component 4: Introduction to Information and Computer Science

Unit 6:
Databases and SQL
Lecture 6

This material was developed by Oregon Health & Science University, funded by the Department of Health and Human Services, Office of the National Coordinator for Health Information Technology under Award Number IU24OC000015.

Topic VI: Create simple querying statements for the database

- The SELECT statement
- Clauses
- Functions
- Joins
- Subqueries
- Data manipulation

Getting Data Out of the Database

- The SQL SELECT statement is the common way to retrieve data
- Statements invoked to retrieve data are called queries
- The general form of the basic standard for the SELECT statement is:

```
SELECT attributename1, attributename2, . . .
FROM tablename;
```

Example SELECT Statement

This query returns all the InstName values and associated InstContact values from the named table:

```
SELECT InstName, InstContact
FROM ClinicalTrialTestingInstitution;
```


The WHERE Clause

This query returns the InstName and InstContact for only those rows where the contact is "7218823843":

```
SELECT InstName, InstContact
FROM ClinicalTrialTestingInstitution
WHERE InstContact = '7218823843';
```


The ORDER BY Clause

The above statement will output the values for InstName and InstContact for rows with an institution contact of "7218823843" in alphabetical order on InstName:

```
SELECT InstName, InstContact
FROM ClinicalTrialTestingInstitution
WHERE InstContact = '7218823843'
ORDER BY InstName;
```

Many More Clauses and Operators (these are for SQL Server)

DISTINCT
Arithmetic (**+, -, *, /, %/Modulo)**
LIKE
Sign
UNION
NULL and **IS NULL**
INTERSECT
=, <, <=, >=, >, <> or !=
Underscore and **%** wildcards
TOP
Concatenation **(+)**
GROUP BY
HAVING
AND and **OR**
NOT
IN and **BETWEEN**
(and more)

Functions

This query returns a count of all the rows in the table (since the primary key is InstName, this is the count of how many different institutions are in the table):

```
SELECT COUNT(*)
From ClinicalTrialTestingInstitution;
```


There Are Many Different Functions (these are for SQL Server)

Convert
Cast
Sum
Avg
Max, Min
Variance or Varp
Stddev or stdev
Date and Time
Months Between
DateName
ABS
Ceiling/Ceil and Floor
Trig functions
Exp
Log, Log10 and LN
Power (and many more)

Getting Data From More Than One Table

- The join of two or more tables by using the primary-to-foreign key relationship allows a query to get data from all tables that have been joined.
- Inner Joins
- Equi-Join
- Natural Join
- Outer Joins

Inner Join

```
SELECT T.TrialCode,
       T.DrugNameFK,
       C.InstName,
       C.InstContact
FROM ClinicalTrialTestingInstitution C, Trial T
WHERE C.InstName = T.InstNameFK
AND T.TrialCode < 4000;
```


Subqueries

- One query's results can be the input to another query.
- A query is nested within another query
- More than two levels of nesting are allowed

Example Subquery

- Find the names of institutions in Denver, Colorado with a trial cost resource of "NSF"
- We could write two SELECT statements and then manually compare the two outputs
- If we combine the two queries the output should be just what we want.

The Subquery

```
SELECT C.InstName
FROM ClinicalTrialTestingInstitution C
WHERE C.City = 'Denver'
AND C.State  = 'CO'
AND C.InstName IN
      (SELECT T.InstNameFK
       FROM Trial T
       WHERE T.TrialCostResource = 'NSF');
```


Manipulation of Data Within the Database

- INSERT

```
INSERT INTO Trial
(TrialCode, TrialStartDate, DrugNameFK, InstNameFK)
VALUES
(39984, 09/20/2010, 'Alaxamine', 'Acme Pharmaceuticals');
```

- UPDATE

```
UPDATE Trial
   SET TrialCostResource = 'NSF'
   WHERE TrialCode = 43895;
```

- DELETE

```
DELETE FROM Trial WHERE TrialCode = 58340;
```

Transaction Processing

- Multiple SQL statements executed as a unit . . . . all or nothing
- Ability to back out changes within a transaction process
  - ROLLBACK
  - COMMIT

Summary

- A database has significant storage, efficiency and security advantages over other forms of storage.
- Data in a database is received, stored and retrieved via a Structured Query Language (SQL) also called a data sublanguage
- The database, tables, attributes, keys and relationships are created with SQL
- SQL can be placed in a transaction process and stored to be executed whenever appropriate

Summary Continued

- Data modeling is a process in the development of a database design
- The entity relationship model shows entities, attributes and relationships.
- Primary and foreign keys are used to connect database tables together making retrieval of data from multiple tables possible

Summary Continued

- Various anomalies are addressed in a database by splitting data into multiple tables.
- There are many normal forms that can be used in the normalization of a database, but typically only the first three are used.

Summary Continued

- The Database Management System (DBMS) is used for maintaining the database and carrying out SQL statements
- There are six phases of database development: Specification gathering, design, testing, implementation, maintenance and modification