MIS2502: Exam 2 Study Guide (Spring 2018)

Instructor: Jing Gong

The exam will be a combination of multiple-choice and short-answer questions. It is a closed-book, closed-notes exam. You will not be able to use a computer during the exam.

The following is a list of items that you should review in preparation for the exam. Note that *not every item on this list may be on the exam, and there may be items on the exam not on this list.*

SQL Out: Advanced Queries (Joins, Subselects)

- Given the schema of a database, be able to create the SQL statements that
    - Require a join of multiple tables
    - Contain a Subselect statement
      (i.e., determine the customers with the highest sales)

SQL In (CREATE, ALTER, INSERT, UPDATE, and DELETE)

- Given the schema of a database, be able to create the SQL statements that
    - Create a table based on a list of its metadata/schema using CREATE TABLE
        - Know how to specify primary keys and foreign keys in CREATE TABLE statements
    - Change the structure of a table using ALTER TABLE
    - Delete a table using DROP TABLE
    - Add a record to a table using INSERT INTO
    - Update an existing record in a table using UPDATE
    - Delete a record from a table using DELETE FROM
- Be familiar with using WHERE conditional statements in the UPDATE and DELETE FROM statements
    - The safest way is to use primary keys in WHERE conditions
- Be familiar with MySQL data types (INT, DECIMAL, VARCHAR, BOOLEAN, DATE/DATETIME, etc.)
    - Know when to use single quotes
- Identify how to add records to a table created from a many-to-many relationship so that the new record associates two existing records in the associated tables
  (i.e., add a record to a film_actor table that associates a particular film with a particular actor)

ETL

- What is it? Why is it important?
- Explain the purpose of each component (Extract, Transform, Load)
- ETL in Excel
    - Understand absolute references versus relative references
    - Understand the syntax and purpose of the Excel functions VLOOKUP, CONCATENATE, LEN

Dimensional Data Modeling

- What is the difference between a data warehouse, a data mart, and a data cube?
- What is a data cube? How does it aggregate data?
    - Give an example of "slicing" or "dicing" the data
- What is the star schema?
    - Understand fact table and dimension tables
- Identify facts, dimensions, and associated data fields that address a business question
- Kimball's four step process for dimensional data modeling
    - What is granularity? Why is it important?
- Advantages and disadvantages of data cubes
    - Understand the "non-volatility" of data cubes

Pivot Table Analysis

- Understand how Pivot Tables relate to data cubes
    - The fields in the ROWS box correspond to dimensions in a data cube
    - The fields in the VALUES box correspond to measured facts in a data cube
- Given a question about a set of data, be able to identify the fields required to create a pivot table
    - Identify which fields are assigned as VALUES and which ones are assigned as ROWS
    - Identify the correct function for aggregation: e.g., SUM, COUNT, AVERAGE, MAX, MIN
- Understand how to use sorting and label filter when creating a pivot table

Data Visualization

- Be able to assess a visualization by applying data visualization principles.
    - Tell a story
    - Graphical integrity (lie factor)
    - Minimize graphical complexity (table versus chart, data ink, chartjunk, moire effect)
- Explain how a visualization can be improved based on those principles.
- Understand basic chart types. Be able to choose an appropriate chart type given a scenario.
- Understand the issues with 3D charts.

**Advanced Analytics and R will be covered in Exam 3.