Georgia State University
J. Mack Robinson
COLLEGE of BUSINESS

Final Exam Study Notes

[Not guaranteed to be all-inclusive.

This is a summary of the main topics covered throughout the course. Please see the lecture notes, the assignment descriptions, and the assigned readings from the text.

Topics

- Conceptual modeling representations (Chen's, Crow's Feet)
  - Be able to read/interpret either
- Relational model and translation of conceptual models into logical models
  - Know the transformation rules. Application of them was part of Assignment 3.
- Data mining: Why is it important? What results might you get? What are its challenges?
- SQL queries: DDL vs DML [Be sure to know the difference between these.]
- Disruptive technologies. Sustaining vs disruptive and role of data.
- Big data: why is there big data? Why is its management important?
- Privacy – Ways to manage privacy. Challenges of data breaches.
- Implications of proper data management. Different ways to ensure data integrity.

There will be a variety of questions on the final. Some will be conceptual in nature, so be sure you understand the main concepts from the course. There will be SQL questions, similar to those in Quiz 2. Besides writing the queries, you will be asked to interpret the questions in terms of real-world applications.

ER-Relational Model

- Translation of ER Model to Relational Model
  - Every entity becomes a separate relation
  - For relationships there are two options:
    - Foreign key: for 1:N relationships (with some variations for optional relationships)
    - Separate relation: with the key of the relation the concatenation, or joining together, of the two keys of the corresponding entities. Relationship attributes become non-key attributes.
  - Reverse engineering. Logical model to conceptual model. Given a logical model, can you answer questions about the corresponding conceptual model from which it came? Why is this important? Analyze existing relational models to identify missing concepts.

SQL

- Understand what a query language is
  - Why are query languages important for data management?
  - Helps us build and retrieve useful information from the database
  - SQL – Structured Query Language, nonprocedural language, tell what to retrieve, not how to do so. Used for data administration, data manipulation and to query a database

- SQL – DDL (data definition language) and DML (data manipulation language)

  - Understand the Create Statement and its usefulness (create the tables)

  - Appreciate how to populate a database. Ensure referential integrity in the data.

- SQL – DML for Basic SQL queries
  - Be able to write an SQL query that involves multiple tables and multiple joins (e.g., chef example)
  - Be able to interpret an SQL query. That is, given an SQL query, be able to provide a corresponding business interpretation of it.
  - Basic form of command: Select – From – Where

  - Know how to insert data into a set of tables

  - Understand the requirements for specifying a data type for each attribute

  - See examples in lecture notes and text

- SQL queries on single versus multiple tables
  - See two different sets of lecture notes and the examples posted and reviewed in class.
  - Understand the concept of "join" on common attributes when queries involve more than one table.
- Be able to answer short answer questions like those that appeared on Quiz 2, the inclass examples, and the lecture notes. This includes providing real-world interpretations of results.

Data Warehouses and Data Mining

- Know the difference between query processing, OLTP and OLAP

    - Understand how / why data is represented or considered as an OLAP cube to handle the multi-dimensional aspects of it. Think of each piece of data as represented as one piece of a (multi-valued) cube.
    - Appreciate that data is input from multiple sources into a data warehouse.
- Data mining applications in: customer segmentation, marketing and promotion targeting, market basket analysis, collaborative filtering, customer churn, fraud detection, financial modeling, and hiring and promotion. Recall also the separate example on market segmentation. This is the matrix of customers, which are categorized based on their status as customers. The managerial implications are that this helps you know whether you should put resources into trying to retain a certain type of customer.

Data and Databases

- Data is an important asset in any organization. Understand the difference and uses of public versus private data.
- Why must you ensure data consistency? Data integrity? Data updates?
    - Data supports decision making; must be correct.

Disruptive technologies

- Still require data. Many other related topics on this including cybersecurity (another course).
- Know characteristics of a true disruptive technology; see examples.

Privacy

- What is it? The right to be left alone.
- Courts: personal privacy balanced with society's "right to know"
- Why is digital privacy an issue? Data can be easily shared. Inferences can be made from data collected from different places (mosaic of data)
- Issues: anonymity, control, sharing with 3rd parties, personal identifying information
- Policy protection: personal, technical, legal or policy
- Laws (many), vary by country and challenging for international organizations
- Many privacy and security concerns
    - Need to protect against them
    - Security is evolving problem
    - Notions of privacy change over time (agree?)
- Why is privacy of great concern in our digital world? Relate this to the need to design and build "good" databases. (This is one of the objectives of this course.)

Big data

Large amounts of data being collected, stored, and used. Why do we have and need to manage big data? Advances in database and other technologies; automated collection of data from many sources; trend to data-driven, real time decision making.

- Reliable telecommunications enable us to share data easily.
- Value from big data includes identifying patterns from which predictions and decisions can be made. Also identify operation concerns (e.g., Walmart cookie example).
- Effective use of data for data mining can require large databases, statistics/mathematics, professional, e.g., data analysts working with domain-specific experts (finance, marketing, human resources). Saw some data mining in this class.

Sample final exam questions

Consider the following sets of relations. This is from the Chef specialty example.

Dish: (DishName, description, cuisine-type)
Chef: (Emp#, name, email, kitchen-where-trained, specialty)
Creates: (DishName, Emp#, expertise-level)

Answer the following query.

What are the names of the chefs who cook American food, what is the name of the food, and what is the chef's level of expertise? Note that this is a written description of the data to be retrieved from a database.

We have seen this example before. Note that the Select command identifies what data will be shown to the user and in what order. The From command shows which tables are needed. The Where command is used for the joins and to provide the restrictions.

Answer:

```
Select name, dishname, expertise-level

From Dish, Chef, Creates

Where Dish.DishName = Creates.DishName

And Chef.Emp# = Creates.Emp#

and cuisine-type = 'American';
```

What kind of query is this? This is a multi-table query that involves joins on three relations. Therefore, the From clause specifies three relations.

What did the corresponding conceptual model look like? It was a many to many relationship between Chef and Dish with expertise-level as a relationship attribute.

You can practice with a similar set of relations:

Skill: (Skill-ID, description, min-yrs-experience-required)

Student: (Student#, name, email, university-graduated-from, major)
Acquires-Skill: (Student#, Skill-ID#, date-completed)

Consider a user's perception of society in today's world. When a user provides data, can they always know exactly what data they are providing. Do you agree or disagree and why? Ans. The user does not always know because there are many ways data is collected and shared with a third party. Users need to be aware of this and minimize the data they provide. Of course, it is necessary to provide certain types of data (name, address, etc.) to conduct online transactions.

Data analytics and business intelligence have been credited with driving discovery and innovation. Why do companies invest in data analytics and business intelligence? Ans. Gaining insights form large databases or patterns that they can act upon.

Reporting tools are used for assessment whereas data mining tools are often used for prediction (T/F). Ans. T.

Good luck on the exam! Please let me know if you have any questions.