

**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 assigned readings from the text.

See also the summary lecture notes (course wrap up) that were posted.

There will be a variety of questions on the final exam. Some will be conceptual in nature, so be sure you understand the main concepts from the course. There will be SQL questions. Besides writing the queries, you will be asked to interpret the questions in terms of real-world applications. You should also be able to read and interpret conceptual models. See examples done in class and from the text.

Topics

* Conceptual modeling representations
  + What is it and why is it important?
  + Representation: 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 (what they are, why they are needed, how to apply them)
* Data mining: Why is it important? What results might you get? What are its challenges?
  + Understand the inclass examples (e.g., diaper, coffee store).
* SQL queries: DDL vs DML (When is each used?)
* Disruptive technologies. Sustaining vs disruptive and role of data.
* Big data: why does big data exist? Why is its management important? How is such data acquired? Example of customer segmentation.
* Privacy – Ways to manage privacy. Challenges of data breaches.
* Implications of proper data management (from both technical and managerial points of view).

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? This is important to identify whether existing relational models are missing concepts.
  + Conceptual modeling is important because it dictates the structure of the final relational model.

SQL

* Understand what a query language is
  + Why are query languages important for data management?
  + Helps us to retrieve useful data 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 SQL DML 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 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 similar to those reviewed in class. This includes providing real-world interpretations of results and the implications for managerial decision making.

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, who are categorized based on their status as customers. The managerial implications are that this provides insights into whether you should put resources for obtaining or retaining 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?
* Understand how data supports decision making; must be correct and database must be properly designed.

Disruptive technologies

* Still require data. Many other related topics including cybersecurity (another course).
* Know characteristics of a true disruptive technology; see examples.
* Just because a technology is based on large amounts of data, does not mean it is disruptive and vice versa.
* Examples of truly disruptive technologies (e.g., impact an industry or industries)

Privacy

* What is it? The right to be left alone.
* Courts: personal privacy balanced with society’s right to know (legal)
* Why is digital privacy an issue? Data can be easily shared. Inferences can be made from data collected from different places.
* 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
  + Do notions of privacy change over time?
* Why is ethics of great concern in our digital world? Ease of sharing of data. Personal information may be explicitly or implicitly shared.

Big data

Large amounts of data being collected, stored, and used. Why do we have so much data and hence the need to store, manage, and use it? 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, to identify operation concerns or anomalies (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).

Good luck on the exam!