Database Management Systems: Design and Implementation

CS564

Ambuj Shatdal

Johnson Controls
and
UW - Madison

What is Data?

- Anything we decide to keep track of or collect
  - Track on a vinyl record vs CD vs SmartPhone
  - What you ate
  - What you are wearing
- Some things we must keep track of
  - Banking
  - Airline reservation
- Some things we probably don’t need to keep track of
  - Web clicks
  - Steps taken

Two dimensions of Data

- How valuable?
- How structured?

How valuable?

- Very valuable, e.g. banking data
- …
- …
- …
- …
- Almost worthless (but easy to collect), e.g. web-clicks

How structured?

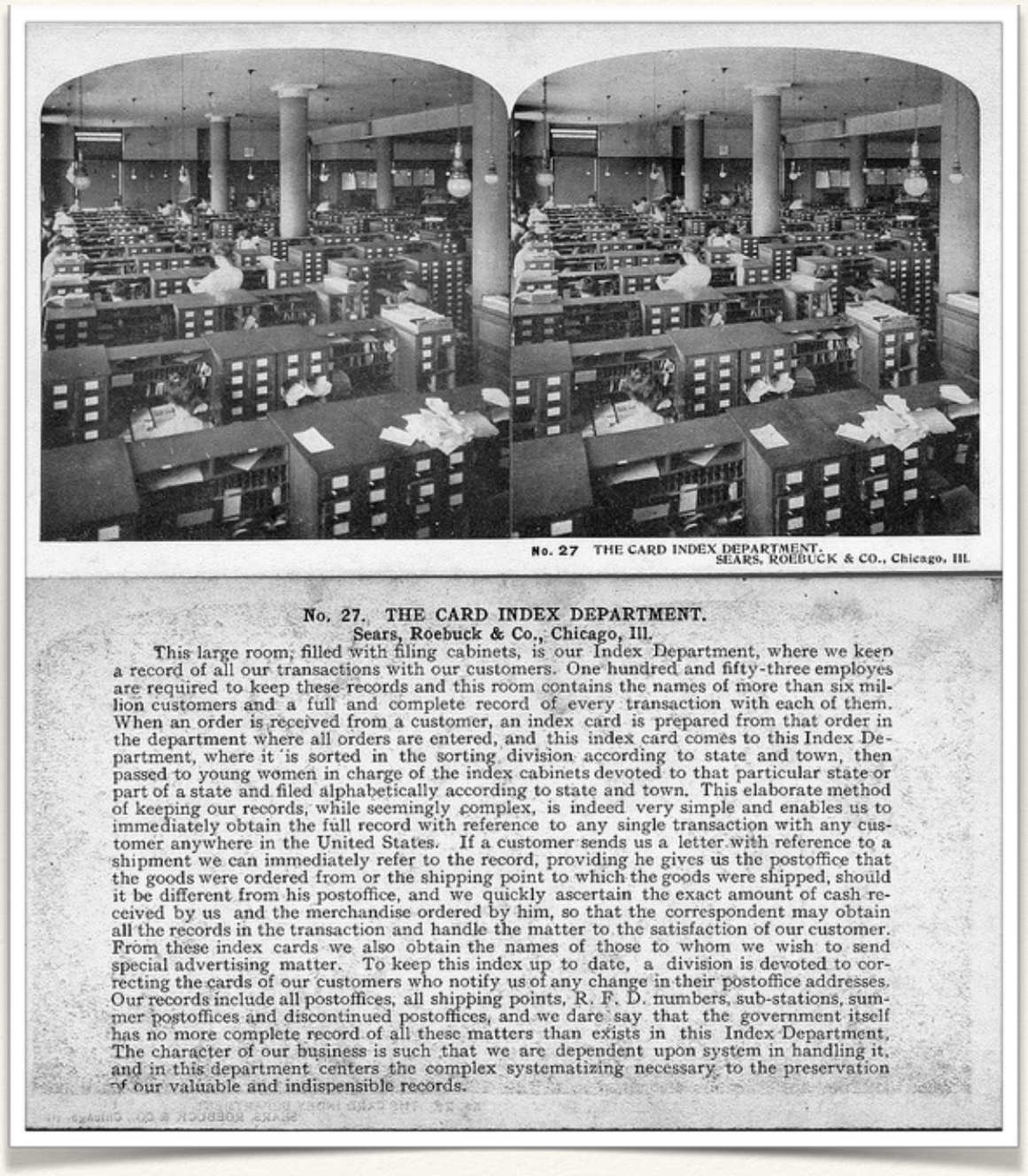
- Structured
  - bank records, check out receipt
- Semi-structured
  - email, wikipedia page, clickstream
- Unstructured
  - audio file, image (without any “meta” information)
  - (generally stored with “meta” info => semi-structured)

What's a Database?

- Storage / retrieval / analysis of valuable data
- Valuable =>
  - Worth it to understand data and its characteristics
    - structure of the data itself - metadata - data modeling
    - analysis of data in the dbms
  - keep it secure and consistent => correct
  - keep it available (in mostly shared environments)
- Data usually structured but can be semi-structured

DBMS

- Software that enables efficient storage, retrieval, analysis of data while ensuring consistency and correctness
- End users (mostly via applications) interact with this system to achieve their data management goals
- Focus on valuable structured-semi-structured data



No. 27 THE CARD INDEX DEPARTMENT.
SEARS, ROEBUCK & CO., Chicago, Ill.

No. 27. THE CARD INDEX DEPARTMENT.
Sears, Roebuck & Co., Chicago, Ill.

This large room, filled with filing cabinets, is our Index Department, where we keep a record of all our transactions with our customers. One hundred and fifty-three employes are required to keep these records and this room contains the names of more than six million customers and a full and complete record of every transaction with each of them. When an order is received from a customer, an index card is prepared from that order in the department where all orders are entered, and this index card comes to this Index Department, where it is sorted in the sorting division according to state and town, then passed to young women in charge of the index cabinets devoted to that particular state or part of a state and filed alphabetically according to state and town. This elaborate method of keeping our records, while seemingly complex, is indeed very simple and enables us to immediately obtain the full record with reference to any single transaction with any customer anywhere in the United States. If a customer sends us a letter with reference to a shipment we can immediately refer to the record, providing he gives us the postoffice that the goods were ordered from or the shipping point to which the goods were shipped, should it be different from his postoffice, and we quickly ascertain the exact amount of cash received by us and the merchandise ordered by him, so that the correspondent may obtain all the records in the transaction and handle the matter to the satisfaction of our customer. From these index cards we also obtain the names of those to whom we wish to send special advertising matter. To keep this index up to date, a division is devoted to correcting the cards of our customers who notify us of any change in their postoffice addresses. Our records include all postoffices, all shipping points, R. F. D. numbers, sub-stations, summer postoffices and discontinued postoffices, and we dare say that the government itself has no more complete record of all these matters than exists in this Index Department. The character of our business is such that we are dependent upon system in handling it, and in this department centers the complex systematizing necessary to the preservation of our valuable and indispensible records.

Rich701's photostream-Flickr

Why study DBMS?

- One of the most successful applications of computers
- Part of daily-life but mostly in the background
- Learning about DBMS makes for clearer understanding of other types of data-related technologies

Design and Implementation

- Fundamental concepts and theoretical underpinning
- How to use one
- How to build one

Logistics

- Prereq: CS367, CS354
- Project(s): Programming required (C++)
- Recommended Textbook: Database Management Systems by Ramakrishnan and Gehrke, 3rd Ed
  - Any other database textbook/older edition is OK
- Best Option: Attend lectures, take additional notes, consult books as needed

Logistics (contd.)

- Two 75 minute lectures (Monday and Wednesday)
- Friday session
  - Discussions: lead by Akshat, one of the TAs
  - Occasional makeup lecture
- Mid-terms 3/12/19 5:30-7:00 PM
- Finals 5/5/19 5:05 PM-7:05 PM
- Homework and Project

Grading

- Mid-term 25%
- Finals 35%
- Homework and Project 40%

Database Management System

- Database
  - an organized store of data for computer processing (Shorter OED)
  - A collection of data stored on a computer storage medium, such as a disk, that can be used for more than one purpose (Barron's)
- DBMS
  - To manage and make use of a database
  - Typically a client-server architecture but others possible

Basic Architecture

- Application Layer - what most users see, talks SQL
- Parsing/Planning Layers - the intelligence
- Runtime or execution Layer - the brawn
- Storage Layer - where data resides, may include simple access layer

Some DBMSs

- IBM DB2
- Oracle
- Microsoft SQL Server
- Teradata
- Amazon Redshift
- Postgres
- MySQL
- SQLite