DATABASE MANAGEMENT SYSTEMS

DATA AND INFORMATION

- Data — raw facts describing characteristics of an event
- Information — data converted into meaningful and useful context
- Database — maintains information about various types of objects (inventory), events (stacentrals), people (employees), and places (warehouses)

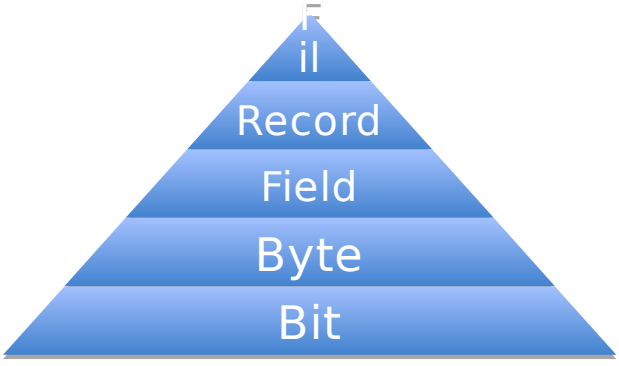
DATA VS INFORMATION

- Data: shares in a stock market
- Information: how did a certain stock perform in the past? How does the performance compare to other stocks?
- Firms need to make their data easily accessible by managers and employees

TRANSACTIONAL DATA AND ANALYTICAL INFORMATION

- Transactional Data (packing slips, sales receipts etc) are put into a database which then provides Analytical information (trends, product stats, sales projections etc.). Used by analysts. Executives use analytical data
- Analysts = *Fine Transactional Data*, *Executives = Coarse Analytical Data*.

TRADITIONAL FILE ORGANIZATION



- **File** — Group of records of same type.
- **Record** — group of related fields.
- **Field** — group of words or a complete number.

- **Byte** — group of bits that represent a single character.
- **Bit** — smallest unit of data; binary digit (0,1)
- **Database** — group of related files
- **Entity** — person, place, thing, event about which information is maintained
- **Attribute** — description of a particular entity
- **Key field** — identifier field used to retrieve, update, or sort a record
- **Primary key** – each record in a database needs an attribute to uniquely identify it so that the record can be retrieved, updated and stored.
    - **Secondary key**: nonunique fields that have some identifying information.
    - **Foreign key**: keys whose purpose is to link 2 or more tables together.

- Traditional file processing allows for functional areas to develop specialized applications.
    - Applications have master file which will inturn have subsets.
        - Subsets can contain redundancies, processing inflexibilities, and wasted storage resources…

VALUE OF QUALITY DATA AND INFORMATION

- Low quality information; data that is:
    - Inaccurate, incomplete, wrong, missing, duplicate
- Sources:
    - Customers intentionally enter inaccurate information for privacy.
    - Different system data/information has different standards and formats
    - Abbreviated or erroneous information by accident or out of convenience
    - Third party and external information

PROBLEMS WITH TRADITIONAL FILE ORGANIZATION

- Data redundancy — duplicate data in multiple data files
- Data inconsistency — same attributes have different values
- Program-Data Dependence — changes in program require changes to data
- Lack of Flexibility
- Poor Security — unknown who is accessing/changing data
- Lack of data sharing and availability — information cannot flow freely across different parts of the firm.

COSTS OF POOR INFORMATION

- Inaccurately track customers
- Unable to identify valuable customers
- Inability to identify selling opportunities
- Marketing to non-existent customers
- Difficulty tracking revenue
- Unable to build strong customer relationships

DATABASE APPROACH

- Database — Collection of data. Centralizes data and controls redundancies.
- Database Management: structured approach for capturing, storing, processing, integrating, distributing, securing and archiving data effectively throughout its life cycle.
- Three general data principles that gives the importance of data life cycle perspective:
    1. **Principle of diminishing data value**: the value of the data diminishes as the data ages. The more recent data are the most valuable.
    2. **Principle of 90/90 data use**: being able to act on real time. According to this principle, 90% of the data is seldom accessed after 90 days (except auditing). Much of the data loses its value after 3 months.
    3. **Principle of data in context**: the ability to capture, process, format and distribute data in near real time or faster=huge investment in data management infrastructure to link remote POS systems to data storage. End users need to see the data in a meaningful format and context if the data is to guide their decisions and plans.
        - To format in meaningful form, businesses use **data visualization-**presents data that is easier and quicker to interpret.
- Database Management System — software that creates and maintains databases. Interfpartionace. Separates logical and physical views of data
- **Enterprise portals:** set of software applications that consolidate, manage, analyze and transmit data to users through web-based interface.
- **Master data management (MDM):** Is a process whereby companies integrate data from various sources or enterprise applications to provide a more unified view of the data. Cannot create a single unified version of the data. **Master reference file** feeds data back to the applications, thereby creating accurate and consistent data across the

enterprise. Based on **data entities**, anything real or abstract about which the company wants to collect and store data.
        - **Master data entities:** main entities of a company, such as customer products, suppliers, employees and assets.
- **Centralized Database vs Distributed Database:** centralized database stored all related files in one physical location while distributed databases are in multiple locations.
    - Distributed databases can include:
        - **Replicated databases:** stores complete copies (replicas) of entire database into multiple locations.
        - **Partitioned database:** database that is divided up so each location has a portion of the entire database.

TYPES OF DATABASES

RELATIONAL DBMS

- Data is represented as 2D tables.
- Examples, DB2, Oracle, MS SQL Server

ENTITIES, ENTITY CLASSES, ATTRIBUTES

- Entity — person, place, thing, transaction or event about which information is stored
- Entity class (Table) — collection of similar entities; CUSTOMER, ORDER, ORDER LINE, DISTRIBUTOR, etc.
- Attributes — characteristics or properties of an entity class; Customer ID, Name, Phone #

KEYS AND RELATIONSHIPS

- Primary key — field that uniquely identifies a given entity in a table
- Foreign key — primary key that appears as an attribute in another table and provides in relationship between two tables

THREE OPERATIONS IN A RELATIONAL DBMS

- **Select** — creates subset of rows that meet specific criteria
- **Join** — combines relational tables to provide users with information
    - Inner join: includes rows in the query only when the joined field matches records in both tables.
    - Outerjoin: includes all the rows from one table in the query results and only those rows from other tables that match the join field in the first table.
    - Left outer join, right out join, cross join, unequal join

- **Project** — enables users to create new tables containing only relevant information

HIERARCHICAL DBMS

- Data is organized tree-like
- Supports one-to-many parent-child relationships

NETWORK DBMS

- Logical many-to-many relationships between data

DISADVANTAGES OF THSE TWO: OUTDATED, LESS FLEXIBLE, AND LACK OF SUPPORT FOR PLAIN LANGUAGE QUERIES.

OBJECT-ORIENTED DBMS

- Stores data and procedures as objects. Able to be shared and retrieved automatically.

DESIGNING DATABASES

- Conceptual design: Abstract model of database from business perspective.
- Physical design: Detailed description of business information needs.
- Entity-relationship diagram: illustrate relationships between database entities
- Normalization: creating smaller stable data structures form complex groups of data

DISTRIBUTED DATA

- Database stored in multiple physical locations
- Reduces vulnerability
- Increase service and responsiveness
- Can run on smaller/less expensive computers
- Depends on high quality telecom liens
- Can be decentralized through:
    - Partition — parts of db are stored elsewhere
    - Replication — duplicate entire database

ENSURING DATA QUALITY

- Audit — structured survey of accuracy and level of completeness of data in an information system
- Cleansing — detecting and correcting data in a database or file

DATABASE TRENDS

DATA WAREHOUSE

- Copy of transaction data specifically structured for querying, analysis, reporting and rigorous data mining.
- Massive database stores current and historical data, which is standardized into common data model
- Consolidates data for management analysis and decision making
- RDMS is used for data warehousing
- Example, marketing, sales, inventory, billing data all gets created into information that stored in data warehouse then transferred into a Marketing Data Mining.

DATA WAREHOUSE FUNDAMENTALS

- Extraction, Transformation, Loading (ETL) — extracts information from internal and external databases, transforms the information using a common set of enterprise definitions and loads information into a data warehouse
- Data mart — contains subset of data warehouse information. Summarized portion of data for specified business unit

INFORMATION CLEANSING OR SCRUBBING

- Information cleansing or scrubbing — fixes/discards inconsistent, incorrect, incomplete information

MULTIDIMENSIONAL DATA ANALYSIS

- Online Analytical Processing (OLAP — interactive, exploratory analysis of multidimensional data.
    - CUBE.
    - Each key in a table is a “dimension”

DATABASE AND THE WEB

- Hypermedia Database
  - Organizes data as network of nodes
  - Text, graphics, sound, video, .exe support
- Wikipedia
  - Search queries are related to a database, that return search results

DATA MINING

- Deep analysis of large pools of data
  - Find hidden patterns, predictions, infers decision-making rules
  - These techniques use Data Warehouse data
- Data-mining analysis capabilities:
  - Cluster analysis
  - Classification
  - Association detection

CLUSTER ANALYSIS

- Division into mutually exclusive groups, each member is as close together as possible to eachother and the groups are far apart.

DATA MINING AND TARGET MARKETING

- Data mining helps: Identify good prospects, choosing appropriate communication channels, picking suitable messages

DATA MINING AND CUSTOMER RELATIONSHIP MANAGEMENT

- Matching campaigns to customers
- Customer segmentation
- Reducing exposure to credit risk — avoid bad customers

ASSOCIATION DETECTION

- Degree to which variables are related and the nature and frequency of these relationships in the information.
- Market basket analysis — analyzes items to detect customer buying behavior and predicts future behaviours.

BUSINESS INTELLIGENCE

- Business intelligence — information people use to support decision-making efforts.

MGMT OPPORTUNITIES: ORGANIZATIONAL PERFORMANCE CAN IMPROVE BY MAKING BETTER USE OF DATA. (INVEST IN DATA MINING AND CRM TECHNOLOGY)

MGMT CHALLENGES: SIGNIFICANT CHANGE IN DATABASE ENVIRONMENT OF A FIRM CAN BE EXPENSIVE AND TIME CONSUMING.

QUERIES

- **Query:** set of instructions used for working with data. Creating a query is like asking the database a question.
    - Select query: most basic query on Access. Creates subsets of data that you can use to answer questions.
    - Query criterion: rule that identifies the records that you want to include in the query result.
    - Parameter query: the user interactively specifies one or more criteria values.
    - Crosstab query: calculates a sum, average, count, or other type of total on darrecords and then groups the results by two types of information