Session One

Ryan Honea

8/29/2017

Summary

The evolution of databases is driven by issues in data dependancy, data redundancy, and file processing. Its evolution has created a much more efficient system for the utilization of data. This includes the production of analytics engines, software products, etc. The creation of databases and their associated systems follows the typical software production methodologies.

Objectives of Session

- Define terms
- Name limitations of conventional file processing
- Explain advantages of databases
- Identify costs and risks of databases
- List components of database environment
- Identify categories of database applications
- Describe databse system development life cycle
- Explain prototyping and agile development approaches
- Explain roles of individuals
- Explain the three-schema architecture for databases

Definitions

- **Databases:** Organized collection of logically related data
- **Data:** Stored representations of meaningful objects and events
  - *Structured:* Numbers, text, dates
  - *Unstructured:* Images, videos, documents
- **Information:** Data processed to increase knowledge in the perso using the data
- **Metadata:** data that describes the properties and context of user data
- **Database Management System:** software system that is used to create, maintain, and provide controlled access to user databases
- **Project:** A planned undertaking of related activities to reach an objective that has a beginning and an end

Introduction to Data and Databases

- Data in context helps users understand data
- Graphical displays turn data into useful information that managers can use for decision making and interpretation
- *Example metadata:* Descriptions of properties or characteristics of data, including data types, field sizes, allowable values, and data context
- Disadvantages of file processing
  - Program-data dependence
  - Duplication of data
  - Limited data sharing

  - Lengthy development times
  - Excessive Program Maintenance
- Problems with data dependency
  - Each application programmer must maintain his/her own data
  - Each application program needs to include code for the metadata of each file
  - Each application program must have its own processing routines for reading, inserting, updating, and deleting data
  - Lack of coordination and central control
  - Nonstandard file formats
- Problems with data redundancy
  - Waste of space to have duplicate data
  - Causes more maintenance headaches
  - Data changes in one file could cause inconsistencies
  - Compromises in data integrity
- Solution to above problems is the database approach
  - Central repository of shared data
  - Data is managed by a controlling agent
  - Stored in a standardized convenient form
  - Requires a Database Management System (DBMS)
- Advantages to database approach
  - Program-data independence
  - Planned data redundancy
  - Improved data consistency
  - Improved data sharing
  - Increased application development productivity
  - Enforcement of standards
  - Improved data quality
  - Improved data accessibility and responsiveness
  - Reduced program maintenance
  - Improved decision support
- Costs and Risks of Database Approach
  - New, specialized personnel
  - Installation and management cost and complexity
  - Conversion costs
  - Need for explicit backup and recovery
  - Organizational conflict
- Elements of the database approach
  - Data Models
    - Graphical system capturing nature and relationship of data
    - Enterprise Data Model - high-level entities and relationships for the organization
    - Project data model - more detailed view, matching data structure in database or data warehouse
  - Entities
    - Noun form describing a person, place, object, event, or concept
    - Composes of attributes
  - Relationships
    - Between entities
    - Usually one-to-many (1:M) or many-to-many (M:N)
  - Relational Databases
    - Database technology involved tables (relations) representing entities and primary/foreign keys representing relationships
- Components of the database environment
  - CASE Tools: Computer-aided software engineering
  - Repository: centralized storehouse of metadata
  - DBMS: Software for managing the database

  - Dataabse: Storehouse of the data
  - Application Programs: Text ad graphical displays to users
  - Data/Database Administrators: Personnel responsible for maintaining the database
  - System Developers: Personnel responsible for designing databases and software
  - End Users: People who use the applications and databases *Enterprise Data Model
  - First step in the database development Process
  - Specifies scope and general content
  - Overall picture of organizational data at high level of abstraction
  - Entity-relationship diagram
  - Descriptions of entity types
  - Relationships between entities
  - Business rules
- Range of database applications
  - Personal databases
  - Two-tier and N-Tier Client/Server databases
  - Enterprise applications
    * Enterprise resource planning (ERP) systems - integrates all enterprise functions (manufacturing, finance, sales, marketing, inventory, accounting, human resources)
    * Data warehousing implementations - integrated decision support system derived from various operational databases

Models for understanding databases

- ER Model assists in understanding how data interacts
- Relationship Types
  - One-To-Many Ex: A customer may place many orders, but each order is placed by one customer
  - Many-To-Many Ex: An order involves many products and one product is involved in many orders
- These relationships create the E-R Model
- A Function-To-Data Entity Matrix shows how different business function will use data entities within a database
- Database Schema
  - External Schema
    * User Views
    * Subsets of conceptual schema
    * Can be determined from business-function/data entity matrices
  - Conceptual Schema
    * E-R Models
  - Internal Schema
    * Logical Strucutres
    * Physical Structures

Approaches to Database Development

System Development Life Cycle

- Detailed, well-planned development process
- Time-consuming, buut comprehensive
- Long development cycle
- Life-Cycle
  - Planning
    * Purpose - preliminary understanding
    * Deliverable - request for study
    * Database Activity - Enterprise modeling and early conceptual data modeling

  - Analysis
    - Purpose - Thorough requirements analysis and structuring
    - Deliverable - Functional system specifications
    - Database Activity - Thorough and intengrated conceptual data modeling
  - Logical Design
    - Purpose - Information requirements elicitation and structure
    - Deliverable - Detailed design specification
    - Database Activity - Logical Database design (transactions, forms, display, views, data integrity, and security)
  - Physical Design
    - Purpose - Develop technology and organizational specifications
    - Deliverable - program/data structures, technology purchases, organizational redesigns
    - Database Activity - Physical database design (define DBMS, physical data organization, database processing programs)
  - Implementation
    - Purpose - Programming, testing, training, installation, documenting
    - Deliverable - Operational programs, documentation, training materials
    - Database Activity - Database implementation, including coded programs, documentation, installation, and conversion
  - Maintenance
    - Purpose - Monitor, repair, enhance
    - Deliverable - periodic audits
    - Database Activity - Database maintenance, performance analyis and tuning, error corrections

Prototyping

- Rapid applicaiton development (RAD)
- Cursory attempt at conceptual data modeling
- Define database during development of initial prototype
- Repeat implementation and maintenance activities with new prototype versions
- Methodology
  - Identify Problem
    - Analyze requirements
    - Develop preliminary data model
  - Develop initial prototype
    - Analyze requirements in detail
    - Integrate database views into conceptual data model
    - Define new database contents to DBMS
    - Decide on physical organization for new data
    - Design database processing programs
    - Code database processing
    - Install new database contents, usually from existing data sources
    - Analyze database to ensure it meets application needs
    - Fix errors in database
  - Implement and use prototype
    - Follow methodology for development of initial prototype
  - Revise and Enhance Prototype
    - Follow methodology for development of initial prototype
  - Convert to operational system
    - Tune database for improved performance
    - Fix errors in database

Project Management

- Projects are planned in planning stage of SDLC
- Executed during analysis, design, and implementation
- Closed at the end of implementation
- People involved in management of project:
  - Business analysts
  - Systems analysts
  - Database analysts and data modelers
  - Users
  - Programmers
  - Database architects
  - Data administrators
  - Project managers
  - Other technical experts

Evolution of database systems

- Driven by four main objectives:
  - Need for program-data independence which would give reduced maintenance
  - Desire to manage more complex data types and structures
  - Ease of data access for less technical personnel
  - Need for more powerful decision support platforms
- Types of database systems
- Flat Files (Legacy)
- Hierarchal (Legacy)
- Network (Legacy)
- Relational
- Object-oriented
- Object-relational
- Data warehousing