Information Systems Planning and the Database Design Process

Ray R. Larson
University of California, Berkeley
School of Information
I 257: Database Management
Lecture Outline

• Review
  – Database Terms
  – Database Types

• Database Life Cycle

• Information Systems Planning

• Information Systems Architecture

• Information Engineering

• Database Design
Announcements

• Yiming office hours:
  – Thursdays 11-12 in the CoLab (or by arrangement

• My office hours:
  – Tue, Thu 2-3 in 207B South Hall or by arrangement
Lecture Outline

• Review
  – Database Terms
  – Database Types

• Database Life Cycle

• Information Systems Planning

• Information Systems Architecture

• Information Engineering

• Database Design
Terms and Concepts

- **Database activities:**
  - *Create*
    - Add new data to the database
  - *Read*
    - Read current data from the database
  - *Update*
    - Update or modify current database data
  - *Delete*
    - Remove current data from the database
Terms and Concepts

• **Enterprise**
  – Organization

• **Entity**
  – Person, Place, Thing, Event, Concept...

• **Attributes**
  – Data elements (facts) about some entity
  – Also sometimes called fields or items or domains

• **Data values**
  – instances of a particular attribute for a particular entity
Terms and Concepts

• **Records**
  – The set of values for all attributes of a particular entity
  – AKA “tuples” or “rows” in relational DBMS

• **File**
  – Collection of records
  – AKA “Relation” or “Table” in relational DBMS
Terms and Concepts

• **Key**
  – an attribute or set of attributes used to identify or locate records in a file

• **Primary Key**
  – an attribute or set of attributes that *uniquely* identifies each record in a file
Terms and Concepts

• **Models**
  – (1) Levels or views of the Database
    • Conceptual, logical, physical
  – (2) DBMS types
    • Relational, Hierarchic, Network, Object-Oriented, Object-Relational
Models (1)

Conceptual Model

Application 1
   External Model

Application 2
   External Model

Application 3
   External Model

Application 4
   External Model

Logical Model

Internal Model

Conceptual requirements

Application 1

Conceptual requirements

Application 2

Conceptual requirements

Application 3

Conceptual requirements

Application 4

Conceptual requirements
Data Models(2): History

- Hierarchical Model (1960’s and 1970’s)
  - Similar to data structures in programming languages.

```
Books (id, title)
  Authors (first, last)
  Publisher
  Subjects
```
Data Models(2): History

• Network Model (1970’s)
  – Provides for single entries of data and navigational “links” through chains of data.
Data Models(2): History

- **Relational Model (1980’s)**
  - Provides a conceptually simple model for data as relations (typically considered “tables”) with all data visible.

<table>
<thead>
<tr>
<th>Book ID</th>
<th>Title</th>
<th>pubid</th>
<th>Author id</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>The history</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>New stuff a</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Another title</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>And yet m</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pubid</th>
<th>pubname</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Harper</td>
</tr>
<tr>
<td>2</td>
<td>Addison</td>
</tr>
<tr>
<td>3</td>
<td>Oxford</td>
</tr>
<tr>
<td>4</td>
<td>Que</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Authorid</th>
<th>Author nan</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Smith</td>
</tr>
<tr>
<td>2</td>
<td>Wynar</td>
</tr>
<tr>
<td>3</td>
<td>Jones</td>
</tr>
<tr>
<td>4</td>
<td>Duncan</td>
</tr>
<tr>
<td>5</td>
<td>Applegate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Book ID</th>
<th>Subid</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subid</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cataloging</td>
</tr>
<tr>
<td>2</td>
<td>history</td>
</tr>
<tr>
<td>3</td>
<td>stuff</td>
</tr>
</tbody>
</table>
Data Models(2): History

• Object Oriented Data Model (1990’s)
  – Encapsulates data and operations as “Objects”
Data Models(2): History

• Object-Relational Model (1990’s)
  – Combines the well-known properties of the Relational Model with such OO features as:
    • User-defined datatypes
    • User-defined functions
    • Inheritance and sub-classing
NoSQL Databases

• Started as a reaction to the overhead in more conventional SQL DBMS
• Usually very simple key/value search operations
• Usually very fast, with low storage overhead, but often lack security, consistency and other features of RDBMS
• May use distributed parallel processing (grid/cloud, e.g. MongoDB + Hadoop)
• Semantic Web “TripleStores” are one type
Lecture Outline

• Review
  – Database Terms
  – Database Types

• Database Life Cycle

• Information Systems Planning
• Information Systems Architecture
• Information Engineering
• Database Design
Database System Life Cycle

- Physical Creation
- Conversion
- Integration
- Operations
- Growth, Change, & Maintenance
- Design

Steps:
1. Design
2. Physical Creation
3. Conversion
4. Integration
5. Operations
6. Growth, Change, & Maintenance
The “Cascade” View

See Hoffer, p. 41
1. Design

- Determination of the needs of the organization
- Development of the Conceptual Model of the database
  - Typically using Entity-Relationship diagramming techniques
- Construction of a Data Dictionary
- Development of the Logical Model
2. Physical Creation

• Development of the Physical Model of the Database
  – data formats and types
  – determination of indexes, etc.
• Load a prototype database and test
• Determine and implement security, privacy and access controls
• Determine and implement integrity constraints
3. Conversion

- Convert existing data sets and applications to use the new database
  - May need programs, conversion utilities to convert old data to new formats.
4. Integration

• Overlaps with Phase 3
• Integration of converted applications and new applications into the new database
5. Operations

• All applications run full-scale
• Privacy, security, access control must be in place.
• Recovery and Backup procedures must be established and used
6. Growth, Change & Maintenance

• Change is a way of life
  – Applications, data requirements, reports, etc. will all change as new needs and requirements are found
  – The Database and applications and will need to be modified to meet the needs of changes
Another View of the Life Cycle

Physical Creation

Conversion

Design

Integration

Operations

Growth, Change

1

2

3

4

5

6
Lecture Outline

- Review
  - Database Terms
  - Database Types

- Database Life Cycle

- Information Systems Planning
- Information Systems Architecture
- Information Engineering
- Database Design
Information Systems Planning

• Scope of IS is now the entire organization
• Sometimes called “enterprise-wide” computing or “Information Architecture”
• Problem: isolated groups in an organization start their own databases and it becomes impossible to find out who has what information, where there are overlaps, and to assess the accuracy of the information
Information Systems Planning

• To support enterprise-wide computing, there must be enterprise-wide information planning

• One framework for thinking about and planning for enterprise-wide computing is an *Information Systems Architecture* or ISA

• Most organizations do **NOT** have such an architecture
Lecture Outline

• Review
  – Database Terms
  – Database Types

• Database Life Cycle

• Information Systems Planning

• Information Systems Architecture
  • Information Engineering
  • Database Design
• An ISA is a “conceptual blueprint or plan that expresses the desired future structure for information systems in an organization”

• It provides a “context within which managers throughout the organization can make consistent decisions concerning their information systems”

  – Quotes from McFadden (Modern Database Management, 4th edition), Ch. 3
• Benefits of ISA:
  – “Provides a basis for strategic planning of IS
  – Provides a basis for communicating with top management and a context for budget decisions concerning IS
  – Provides a unifying concept for the various stakeholders in information systems.
  – Communicates the overall direction for information technology and a context for decisions in this area
  – Helps achieve information integration when systems are distributed (increasing important in a global economy)
  – Provides a basis for evaluating technology options (for example, downsizing and distributed processing)”
    – Quotes from McFadden (Modern Database Management, 4th edition), Ch. 3
• Zachman ISA Framework components
  – Data
    • The “What” of the information system
  – Process
    • The “How” of the information system
  – Network
    • The “Where” of the information system
  – People
    • Who performs processes and are the source and receiver of data and information.
  – Events and Points in time
    • When processes are performed
  – Reasons
    • Why: For events and rules that govern processing
Six roles or perspectives of the **Data**, **Process** and **Network** components

- Business scope (Owner)
- Business model (Architect)
- Information systems model (Designer)
- Technology model (Builder)
- Technology definition (Contractor)
- Information system (User)
### Zachman Framework

<table>
<thead>
<tr>
<th>Abstractions</th>
<th>Data</th>
<th>Function</th>
<th>Network</th>
<th>People</th>
<th>Time</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perspectives</td>
<td>What</td>
<td>How</td>
<td>Where</td>
<td>Who</td>
<td>Why</td>
<td>Who</td>
</tr>
<tr>
<td>Scope</td>
<td>List of Things Important to the Business</td>
<td>List of Processes - the Business Performs</td>
<td>List of Locations - In which the Business Operates</td>
<td>List of Organizations - Important to the Business</td>
<td>List of Events - Significant to the Business</td>
<td>List of Business Goals and Strategies</td>
</tr>
<tr>
<td>Contextual</td>
<td>Entity = Class of Business Thing</td>
<td>Function = Class of Business Process</td>
<td>Node = Major Business Location</td>
<td>People = Class of People and Major Organizations</td>
<td>Time = Major Business Event</td>
<td>Ends/Means = Major Business Goal/Critical Success Factor</td>
</tr>
<tr>
<td>Enterprise Model</td>
<td>Owner</td>
<td>e.g., Semantic Model</td>
<td>e.g., Business Process Model</td>
<td>e.g., Logistics Network</td>
<td>e.g., Work Flow Model</td>
<td>e.g., Business Plan</td>
</tr>
<tr>
<td>System Model</td>
<td>Designer</td>
<td>e.g., Logical Data Model</td>
<td>e.g., Application Architecture</td>
<td>e.g., Distributed System Architecture</td>
<td>e.g., Human Interface Architecture</td>
<td>e.g., Procesing Structure</td>
</tr>
<tr>
<td>Technology Constrained Model</td>
<td>Builder</td>
<td>e.g., Physical Data Model</td>
<td>e.g., System Design</td>
<td>e.g., Technical Architecture</td>
<td>e.g., Presentation Architecture</td>
<td>e.g., Control Structure</td>
</tr>
<tr>
<td>Detailed Representations</td>
<td>Subcontractor</td>
<td>e.g., Data Definition</td>
<td>e.g., Program</td>
<td>e.g., Network Architecture</td>
<td>e.g., Security Architecture</td>
<td>e.g., Timing Definition</td>
</tr>
<tr>
<td>Functioning Enterprise</td>
<td>Implementation</td>
<td>Implementation</td>
<td>Implementation</td>
<td>Implementation</td>
<td>Implementation</td>
<td>Implementation</td>
</tr>
</tbody>
</table>

**John A. Zachman, Zachman International**

<table>
<thead>
<tr>
<th>I257 - Fall 2015</th>
</tr>
</thead>
</table>

UC Berkeley School of Information 2015.09.01 - SLIDE 35
1. Enterprise Scope
(Owner)

Data
List of entities important to the business

Process
List of processes or functions that the business performs

Network
List of locations in which the business operates
2. Enterprise Model (Architect)

- Data: Business entities and their relationships
- Process: Function and process decomposition
- Network: Communications links between business locations
3. Information System Model (Designer)

Data
Model of the business data and their relationships (ERD in Database design)

Process
Flows between application processes

Network
Distribution Network
4. Technology Constrained Model (Builder)
Information Systems Architecture

Data
- Database Schema and subschema definition

Process
- Program Code and control blocks

Network
- Configuration definition/Network Architecture

5. Technology Definition/Detailed Representations (Contractor)
6. Functioning Enterprise (User)

Data
Implemented Database and information

Process
Implemented Application Programs

Network
Current System Configuration
Lecture Outline

• Review
  – Database Terms
  – Database Types

• Database Life Cycle

• Information Systems Planning

• Information Systems Architecture

• Information Engineering

• Database Design
Information Engineering

• A formal methodology that is used to create and maintain information systems
• Starts with the Business Model and works in a Top-Down fashion to build supporting data models and process models for that business model
Information Engineering

Planning

1. Identify Strategic Planning Factors
   a. Goals
   b. Critical Success Factors
   c. Problem Areas
2. Identify Corporate Planning Objects
   a. Org. Units
   b. Locations
   c. Business Functions
   d. Entity types
3. Develop Enterprise Model
   a. Function decomposition
   b. Entity-Relationship Diagram
   c. Planning Matrices

Analysis

1. Develop Conceptual Model
   (detailed E-R Diagram)
2. Develop Process Models
   (data flow diagrams)

Design

1. Design Databases
   (normalized relations)
2. Design Processes
   a. Action Diagrams
   b. User Interfaces: menus, screens, reports

Implementation

1. Build database definitions
   (tables, indexes, etc.)
2. Generate Applications
   (program code, control blocks, etc.)
One more recent, and very popular, development methods is RAD Prototyping.

1. Identify Problem
   - Conceptual data modeling
   - Initial requirements
2. Develop Prototype
   - Logical data modeling
   - New Requirements
3. Implement and use Prototype
   - Physical database design and definition
   - Working Prototype
4. Convert to Operational System
   - Problems
5. Next Version
   - Revise and enhance Prototype
   - Next Version
Lecture Outline

• Review
  – Database Life Cycle
• Information Systems Planning
• Information Systems Architecture
• Information Engineering

• Database Design
Lecture Outline

• Review
  – Database Terms
  – Database Types
• Database Life Cycle
• Information Systems Planning
• Information Systems Architecture
• Information Engineering
• Database Design
Database Design Process

Conceptual Model

Application 1
External Model
Conceptual requirements

Application 2
External Model
Conceptual requirements

Application 3
External Model
Conceptual requirements

Application 4
External Model
Conceptual requirements

Logical Model

Internal Model

Application 1
Conceptual requirements

Application 2
Conceptual requirements

Application 3
Conceptual requirements

Application 4
Conceptual requirements
Stages in Database Design

1. Requirements formulation and analysis
2. Conceptual Design -- Conceptual Model
3. Implementation Design -- Logical Model
4. Physical Design -- Physical Model
Database Design Process

- Requirements formulation and analysis
  - Purpose: Identify and describe the data that are used by the organization
  - Results: Metadata identified, Data Dictionary, Conceptual Model-- ER diagram
Database Design Process

• Requirements Formulation and analysis
  – Systems Analysis Process
    • Examine all of the information sources used in existing applications
    • Identify the characteristics of each data element
      – numeric
      – text
      – date/time
      – etc.
    • Examine the tasks carried out using the information
    • Examine results or reports created using the information
Database Design Process

• Conceptual Model
  – Merge the collective needs of all applications
  – Determine what **Entities** are being used
    • Some object about which information is to maintained
  – What are the **Attributes** of those entities?
    • Properties or characteristics of the entity
    • What attributes uniquely identify the entity
  – What are the **Relationships** between entities
    • How the entities interact with each other?
Database Design Process

• Logical Model
  – How is each entity and relationship represented in the Data Model of the DBMS
    • Hierarchic?
    • Network?
    • Relational?
    • Object-Oriented?
Database Design Process

• Physical (AKA Internal) Model
  – Choices of index file structure
  – Choices of data storage formats
  – Choices of disk layout
Database Design Process

• External Model
  – User views of the integrated database
  – Making the old (or updated) applications work with the new database design
Developing a Conceptual Model

• Overall view of the database that integrates all the needed information discovered during the requirements analysis.

• Elements of the Conceptual Model are represented by diagrams, *Entity-Relationship or ER Diagrams*, that show the meanings and relationships of those elements independent of any particular database systems or implementation details.
An Entity is an object in the real world (or even imaginary worlds) about which we want or need to maintain information:

- Persons (e.g.: customers in a business, employees, authors)
- Things (e.g.: purchase orders, meetings, parts, companies)
Attributes

• Attributes are the significant properties or characteristics of an entity that help identify it and provide the information needed to interact with it or use it. (This is the Metadata for the entities.)

- Employee
  - Name
    - First
    - Middle
    - Last
  - SSN
  - Age
  - Birthday
  - Projects
Relationships

• Relationships are the associations between entities. They can involve one or more entities and belong to particular relationship types
Relationships

- Student
  - Attends
  - Class

- Project
  - Supplies
    - Supplier
      - Supplies project parts
    - Part
Types of Relationships

- Concerned only with cardinality of relationship

Chen ER notation
Other Notations

"Crow’s Foot"
Other Notations

IDEFIX Notation

Employee \(\rightarrow\) Assigned \(\rightarrow\) Truck

Employee \(\rightarrow\) Assigned \(\rightarrow\) Project

Employee \(\rightarrow\) Assigned \(\rightarrow\) Project
More Complex Relationships

Manager

Employee

SSN

Date

Project

4(2-10)

Assigned

Manages

Manages

Employee

n/n(1)

Evaluation

n/n(1)

1/n/n

1/1/1

Project

Employee

Employee

Manages

Is Managed By

1

n

1

4(2-10)

1

n

n/n(1)
Weak Entities

- Owe existence entirely to another entity
Supertype and Subtype Entities

- Employee
  - Manages
  - Is one of
    - Sales-rep
      - Sold
      - Invoice
    - Clerk
    - Other

Supertype and Subtype Entities.
Many to Many Relationships

- Employee
- Project
- Project Assignment
- Assigned
- SSN
- Proj#
- Hours
- Is Assigned
- Project
Next Time

- **THURSDAY:**
  - More on ER modelling
  - Designing the Conceptual Model for the Diveshop Database
  - Assignment 1
  - Using MySQL for Assignment 1