

# Understanding RDBMS

IT4GIS  
Keith T. Weber, GISP  
GIS Director  
ISU-GIS Training and Research Center

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

# FUNDAMENTALS

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

# RDBMS

- Relational Database Management System
- The "I" in GIS (*Information*)

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

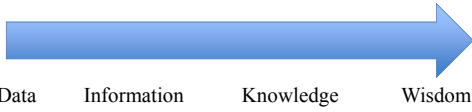
---

---

---

## BTW

- The Wisdom Pathway



Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Database software...

- Light Duty
- Medium Duty
- Heavy Duty

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Database software...

- Light Duty
- Medium Duty
- Heavy Duty



Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

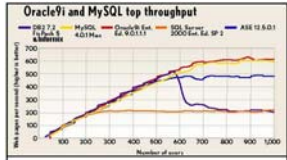
---

## IBM DB2 UDB

- The GIS Center's heavy hitter- IBM DB2, Universal



DB2 Data Management Software



Article found in 25 February 2002 edition of eWeek



Pocatello | Idaho Falls | Meridian | Twin Falls

UNIVERSITY

---

---

---

---

---

---

---

---

## Spreadsheets vs. Databases

- Integrity!
- Structure

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State UNIVERSITY

---

---

---

---

---

---

---

---

## RDBMS CONCEPTS AND TERMS

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State UNIVERSITY

---

---

---

---

---

---

---

---

## Independence

- Physical
- Logical



Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Integrity

- Important for consistency and transaction management.
- Types:
  - Domain: all values come from predefined domains or are null
  - Redundancy: problems occur as a result of repetitive storage that is not consistently updated and from stored data that is derived from other stored data. Redundant info must be consistent.

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Integrity (cont'd)

- Constraint: Business integrity. Stored data must not violate business rules.
- Entity: Every record must be uniquely identifiable (index field or ObjectID)
- Referential: Relationships must not be ambiguous. Two types...
  - Cascading or non-cascading

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Key Fields



- Unique Identifiers (?)
  - Primary key
  - Foreign key
- AKA- Relate fields.

---

---

---

---

---

---

---

---

## RDBMS STRUCTURE

---

---

---

---

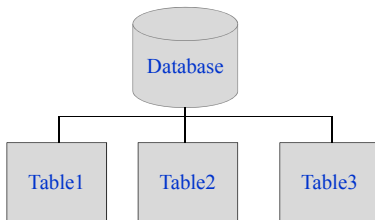
---

---

---

---

## Database Tables



---

---

---

---

---

---

---

---

## Table Structure

	COLUMN 1 (FIELD OR ATTRIBUTE)	COLUMN 2
ROW 1 (RECORD OR ENTITY)	VALUE	
ROW 2		

---

---

---

---

---

---

---

---

---

---

## Data Value Types

Type Name	Storage Occupied/ data value	Valid Domain Range
Short Integer	2 bytes	-32768 to 32767
Long Integer	4 bytes	-2147483648 to 2147483647
Float	4 bytes	Any number from $n^{-45}$ to $n^{38}$
Double	8 bytes	Any number from $n^{-254}$ to $n^{108}$
Text (string)	10 + max. length = bytes	Any alphanumeric characters
Date	8 bytes	Jan 1, 100 to Dec. 31 9999
LOB (variant)	22 + max. length = bytes	Any alphanumeric characters

---

---

---

---

---

---

---

---

---

---

## (btw) Raster Data Types Worth Knowing

- **1\_BIT**—A 1-bit unsigned integer. The values can be 0 or 1.
- **2\_BIT**—A 2-bit unsigned integer. The values supported can be from 0 to 3.
- **4\_BIT**—A 4-bit unsigned integer. The values supported can be from 0 to 15.
- **8\_BIT\_UNSIGNED**—An unsigned 8-bit data type. The values supported can be from 0 to 255.
- **8\_BIT\_SIGNED**—A signed 8-bit data type. The values supported can be from -128 to 127.
- **16\_BIT\_UNSIGNED**—A 16-bit unsigned data type. The values can range from 0 to 65,535.
- **16\_BIT\_SIGNED**—A 16-bit signed data type. The values can range from -32,768 to 32,767.
- **32\_BIT\_UNSIGNED**—A 32-bit unsigned data type. The values can range from 0 to 4,294,967,295.
- **32\_BIT\_SIGNED**—A 32-bit signed data type. The values can range from -2,147,483,648 to 2,147,483,647.
- **32\_BIT\_FLOAT**—A 32-bit data type supporting decimals.
- **64\_BIT**—A 64-bit data type supporting decimals.

---

---

---

---

---

---

---

---

---

---

## Making Sense of all this...

- Recall, there are 8-bits in 1-byte
- Cross-reference
  - 8-bit is byte data
  - 16-bit is short integer (2 bytes)
  - 32-bit (signed or unsigned) is long integer (4 bytes)
  - 32-bit (float) is single-precision floating point (4 bytes)
  - 64-bit is double-precision floating point (8 bytes)

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## DATABASE DESIGN

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Basic Steps in Database Design



- Understand and document the business' needs.
  - Problem statement
  - Business object types
  - Business relationships
  - Business constraints
- Create an ERM
- Data and process inventory
- Develop tuple types
- Tuple types to tables
- Integrity
- Populate the database

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## A Scenario...

- Develop a GIS-Based Tourism database for Southeast Idaho.



Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Document the business needs

- What problem or issue is this database going to address?
- This is a **business statement**



Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## READING A BUSINESS STATEMENT

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---



## Identify Candidate Classes

- A *candidate* class may or may not remain a class throughout the design process
- A *candidate* class may or may not become a table
- Do not think about tables and relationship classes at this point

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Think Object-Oriented

- Classes are nouns
- A noun is a “person, places, and things”

**SCHOOLHOUSE ROCK**



Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## And now...Verbs

- *Candidate* methods are verbs
  - They show action
  - They are behaviors

**SCHOOLHOUSE ROCK**



Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Methods

- Identifying *candidate* methods allows us to better understand how the business operates and how the Enterprise uses GIS data.
- A method is a behavior...a relationship between classes
- The candidate methods will describe an inheritance, aggregation, or dependency relationship

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## DATABASE DESIGN (CONT'D)

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## The Preliminary ERM

- Symbolized.
  - Standard Representation
  - Attribute Representation
  - Entity Instance Representation

<b>DINING</b>
<b>K Restaurant Number: 126</b>
Name: Burger King
Type of food: Fast



Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Relationships

- Determine the relationships between your entity types.
- Add these to the ERM

---

---

---

---

---

---

---

---

## Define the List

- Database Dictionary
  - Restaurant\_Name
    - The name of the restaurant
  - Food\_Type
    - Categories of food (e.g., 1 = Continental, 2 = Fast food, etc.)
  - Cost\_Mean
    - The average cost of all regular menu items



---

---

---

---

---

---

---

---

## Develop Tuple Types

- Use your ERM with relationships
- Perform a “Walk-through” exercise
  - Simulate information is being added/used in your database.
- Symbolize using Attribute Representation

---

---

---

---

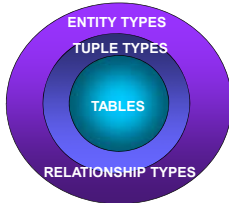
---

---

---

---

## Tuple Types to Tables



Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Normalization

- First-Fifth Form Normal (1FN , 2FN ,...5FN)
- Academic
- Applied

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## 1FN

- All values are atomic
  - Single cell contains single data value
- Eliminate repeating groups
  - Puppy\_Trick1, Puppy\_Trick2, ...

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Check this (1FN)...

Field Name	Data Type	Length	Description	Examples
OWNER1	Text	100	Owner of Parcel	John Smith
OWNER2	Text	100	Additional owner of parcel	Mary Smith
MAIL_ADD1	Text	100	Mailing address of owner	1234 S Paper Rd
MAIL_ADD2	Text	100	Additional mailing address of owner	apt 5C
MAIL_CITY	Text	100	Mailing city of owner	Andover, Poo.
MAIL_STATE	Text	2	Mailing state of owner	AK
MAIL_ZIP	Text	10	Mailing U.S. zip code of owner	99721-0000
MAIL_COUNTRY	Text	4000	Mailing country of owner	USA
SITE_ADD	Text	100	Site address of property	6700 W Slinger Ave
SITE_CITY	Text	100	City of property	King
SITE_ZIP	Text	10	Zip code of property	16005-0000
CATEGORY1	Text	2	Assessed land use	01
CATEGORY2	Text	2	Assessed land use	02
CATEGORY3	Text	2	Assessed land use	03
CATEGORY4	Text	2	Assessed land use	04
CATEGORY5	Text	2	Assessed land use	05
CATEGORY6	Text	2	Assessed land use	06
CATEGORY7	Text	2	Assessed land use	07
BRW_ACRES	Double	7	Irrigated acreage by land use category	7.50
DRY_ACRES	Double	7	Unirrigated acreage by land use category	1.68
ZONING	Text	10	Zoning category	R3W
DESC1	Text	100	Property description	PUR 0000 of S20E4
DESC2	Text	100	Property description	S&C 25 IN II
DESC3	Text	100	Property description	#44930-5
DESC4	Text	100	Property description	
DESC5	Text	100	Property description	
SUBVNT	Text	100	Subdivision name	Happy Valley
VALUATION	Integer	7	Net assessed value of property	100,000

Pocatello | Idaho Falls | M...

**Idaho State UNIVERSITY**

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

## 2FN

- Satisfy 1FN and...
- Redundant **data** must be eliminated
  - How?
  - Example: Puppy\_ID, Trick\_ID, Trick\_Name

Pocatello | Idaho Falls | Meridian | Twin Falls

**Idaho State UNIVERSITY**

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

## Check this (2FN)...

Field Name	Data Type	Length	Description	Examples
OWNER1	Text	100	Owner of Parcel	John Smith
OWNER2	Text	100	Additional owner of parcel	Mary Smith
MAIL_ADD1	Text	100	Mailing address of owner	1234 S Paper Rd
MAIL_ADD2	Text	100	Additional mailing address of owner	apt 5C
MAIL_CITY	Text	100	Mailing city of owner	Andover, Poo.
MAIL_STATE	Text	2	Mailing state of owner	AK
MAIL_ZIP	Text	10	Mailing U.S. zip code of owner	99721-0000
MAIL_COUNTRY	Text	4000	Mailing country of owner	USA
SITE_ADD	Text	100	Site address of property	6700 W Slinger Ave
SITE_CITY	Text	100	City of property	King
SITE_ZIP	Text	10	Zip code of property	16005-0000
CATEGORY1	Text	2	Assessed land use	01
CATEGORY2	Text	2	Assessed land use	02
CATEGORY3	Text	2	Assessed land use	03
CATEGORY4	Text	2	Assessed land use	04
CATEGORY5	Text	2	Assessed land use	05
CATEGORY6	Text	2	Assessed land use	06
CATEGORY7	Text	2	Assessed land use	07
BRW_ACRES	Double	7	Irrigated acreage by land use category	7.50
DRY_ACRES	Double	7	Unirrigated acreage by land use category	1.68
ZONING	Text	10	Zoning category	R3W
DESC1	Text	100	Property description	PUR 0000 of S20E4
DESC2	Text	100	Property description	S&C 25 IN II
DESC3	Text	100	Property description	#44930-5
DESC4	Text	100	Property description	
DESC5	Text	100	Property description	
SUBVNT	Text	100	Subdivision name	Happy Valley
VALUATION	Integer	7	Net assessed value of property	100,000

Pocatello | Idaho Falls | M...

**Idaho State UNIVERSITY**

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

### 3FN

- Satisfy 1NF and 2FN and...
- No non-key attributes are dependent on other non-key attributes.
  - Example: Appointment\_ID, Name, Date, Time, Species

---

---

---

---

---

---

---

---

### After Normalization

- New tuple types will be created.
- New tables will be planned.
- Many-many relationships will be handled using associative tables (bridge tables).



---

---

---

---

---

---

---

---

### De-Normalization

- What? Is this heresy?



---

---

---

---

---

---

---

---

## Designing the Actual RDBMS

- Visual modeling based upon your ERM and Tuple type model.
- Implementation of integrity rules based upon your business constraints.



Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## Populate...

- Questions and concerns to revisit
  - Null data
  - Reporting discrepancies and variations
  - Measuring or estimating methods
  - Client utility/efficiency

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

## The Last Step?

# Validation!

Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---

---

Questions?



Pocatello | Idaho Falls | Meridian | Twin Falls

Idaho State  
UNIVERSITY

---

---

---

---

---

---

---