DATABASE ISSUES

CSE416-S01 - Software Engineering

1

Reference

- MongoDB Design
  https://www.mongodb.com/docs/manual/data-modeling/

2

Project Observations

- No requirement for enormous amounts of data
- Processing done one state at a time
- Most (all?) GUI data stored on server
- No graded requirement to support simultaneous multiple users (i.e., multiple simultaneous states) although your design should assume multiple simultaneous users
- Most client requests can be handled from memory
- Conclusions
  - Primary data requirement is for data persistence between invocations of your system
  - Importance of data caching

Your sequence diagrams should show approach to data caching

3

Project DB Approaches

- NoSQL (MongoDB)
  - Simple interface
  - JSON-like storage
  - Question of access time to load a state
  - Possible use of in-memory MongoDB
- Relational DB – good fit with an ORM (Object Relational Mapping) approach to simplify DB design
- File system, backed with a DBMS
  - Simple approach
  - DBMS provides pointers to files (easier maintenance)

4

DB Component of Design Review

You can store some data (e.g., GeoJSON) in your file system if you have a path to it stored in your DB

- 30-minute reviews begin the week of 11/17
- For a MongoDB, you will
  - Show structure of your data (complete) using a list of collections and a schema for each collection
  - Discuss your approach to queries from different collections
- For a relational DB, you will show
  - Structure of your data (complete) using a relational DB diagram
  - Sample DB in 3NF
  - Use of a persistence layer (e.g., JPA) in your sequence diagrams
- DB Data for at least one of your detailed states and one other state

5

Relational DB Approach

- Use MySQL or any other DBMS you are familiar with
- ORM allows you to structure your DB based on your Java class design
- You can use
  - CS server – might have performance problems at end of semester
  - Your own server (e.g., on your laptop)
  - External server – risk of non-availability during final presentation

6

MongoDB Design Issue

- Will each GUI use case return a single JSON with all the needed data?
- Or will you issue a series of requests for the use case
- Example – GUI-2 Display State
  - Use case requires state map, zoom level, center point (lat/long), EAVS geographical unit boundaries (for a detailed state), CVAP % (if political detailed state), and GUI-3, GUI-4, GUI-5

7

MongoDB

- In the Design Review, you need to be prepared to show all the MongoDB collections you have defined
- If you intend to aggregate data from multiple collections in response to a GUI request, you need to be prepared to discuss the technique you will use to retrieve and aggregate the data

8

MongoDB Schema - Collection Example …

```
const equipmentModelSchema = new Schema({
  modelName: { type: String, required: true, unique: true },
  equipmentType: {
    type: String,
    enum: ['BMD', 'OPTICAL_SCANNER', 'E_POLLBOOK', 'DRE_NO_VVPAT',
'DRE_WITH_VVPAT', 'HAND_COUNT'],
    required: true
  },
```

9

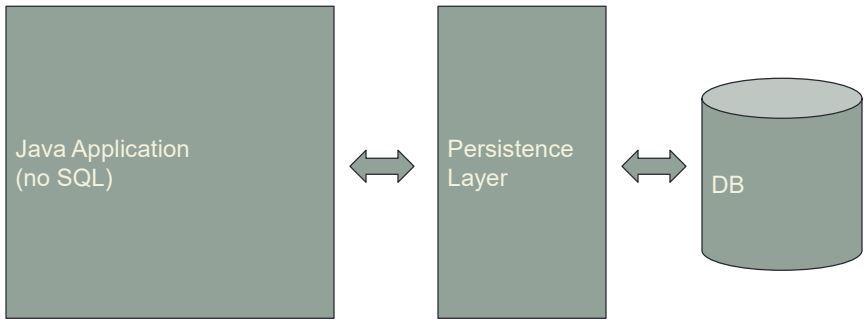
… MongoDB Schema - Collection Example

```
manufacturer: String,
  firstManufactured: Number,
  lastManufactured: Number,
  isActive: Boolean,
  OS: String,
  firmwareVersion: String,
  batteryLife: Number,      // hours (if relevant)
  scanningRate: Number,     // sheets/sec (if relevant)
  hasVVPAT: Boolean,        // for DREs (if relevant)
  paperCapacity: Number,    // sheets (if relevant)
  certificationLevel: String, // e.g. "federal", "state", "none"
  securityRisks: String,    // comments about security
  equipmentNotes: String
}, {});
```

10

 11

Relational Persistence Layer Architecture



You can use your Java annotation to 1) define your DB table structure or 2) adapt your object mapping to your existing DB structure

11

 CSE416 – Software Engineering 12

An Entity

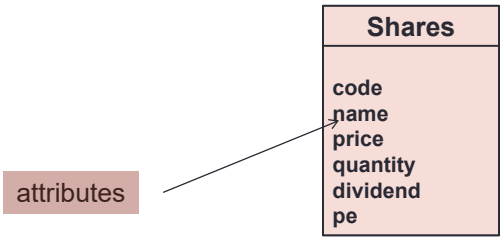
- Usually corresponds to something concrete in the domain of the application
- Represented by a rectangle
- An instance is a particular occurrence of an entity (corresponds to a row in a DB table)

12

 CSE416 – Software Engineering 13

Attributes

- Also referred to as properties
- An attribute is a discrete data element that describes an entity
- Attribute names should be meaningful

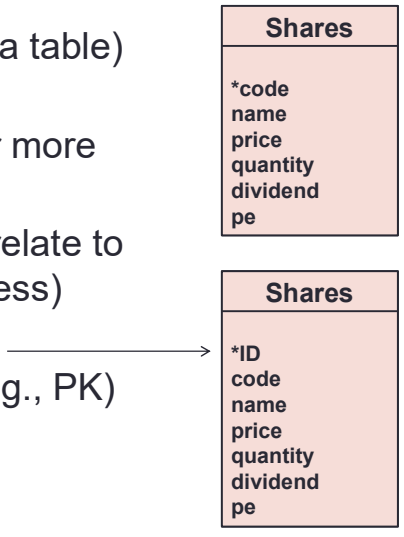


13

 CSE416 – Software Engineering 14

Identifiers (Primary Keys)

- Every instance of an entity (think row of a table) must be uniquely identified
- An identifier (primary key) can be one or more attributes
- Better to use an identifier that does not relate to a domain attribute (guaranteed uniqueness)
- A leading asterisk denotes an identifier (sometimes, another notation is used, e.g., PK)



14

DB Naming Conventions

- No universal standard
- Good to be consistent within a project
- Camel case used frequently

15

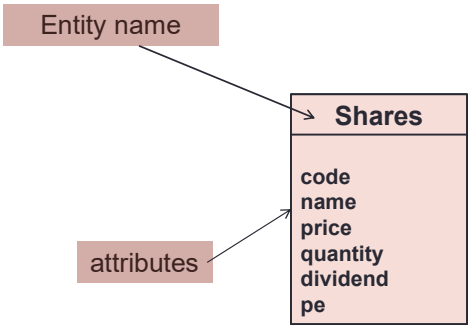
CSE416 DB Naming Conventions

- Options
  - Camel case for table names (upper cc) and column names (lower cc)
  - All caps for table names with underscore (_) as a separator
- Table names – plural (unlike OO convention)
- Column names – singular
- Primary key field – ID
- Avoid acronyms and abbreviations except where well known (e.g., PI for Principal Investigator)

16

The Building Blocks

- Entity
- Attribute
- Relationship

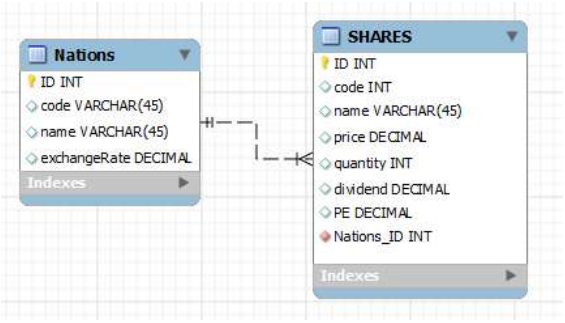


Remember our notation is plural entity name in upper camel case and singular attribute names in lower camel case

17

Relationships

- ERD and RDM show relationships between entities
  - 1-1
  - 1-many
  - Recursive
- ERD shows
  - Many-many
  - No foreign keys
- RDM usually shows
  - Associative entity (in-between table)
  - Foreign keys



Workbench uses more of a DB model style (not an ERD)

18

Normalization

- A theoretical foundation for the relational model
- Application of a series of rules that gradually improve the design
  - Minimize redundancy
  - Minimize dependency
- Objectives*
  - Free the collection of relations from undesirable insertion, update and deletion dependencies
  - Isolate data so that additions, deletions, and modifications of a field can be made in just one table and then propagated through the rest of the database

* Wikipedia

19

Normal Forms

- Based on rules about relationships among the columns of a table
- Removes data redundancies that can cause update anomalies
- A classification of relations
  - 1NF
  - 2NF
  - 3NF
  - BCNF
  - 4NF
  - 5NF

20

Data Redundancy

- Major aim of relational database design is to group columns into tables to:
  1. minimize data redundancy and
  2. reduce file storage space required by implemented base tables

Problems associated with data redundancy are illustrated in the example on the following slides

21

StaffDistributionCenters Table

- Note the details of a distribution center are repeated for every employee (not normal form)

StaffDistributionCenter

ID	StaffNo	staffName	position	salary	dCenterNo	dAddress	dTelNo
1	S1500	Tom Daniels	Manager	$48,000.00	D0001	8 Jefferson Way, Portland, OR 97201	503-555-3618
2	S0003	Sally Adams	Assistant	$30,000.00	D0001	8 Jefferson Way, Portland, OR 97201	503-555-3618
3	S0010	Mary Martinez	Manager	$51,000.00	D0002	City Center Plaza, Seattle, WA 98122	206-5556756
4	S3250	Robert Chin	Assistant	$33,000.00	D0002	City Center Plaza, Seattle, WA 98122	206-5556756
5	S2250	Sally Stern	Manager	$48,000.00	D0004	2 W. El Camino, San Francisco, CA 94087	822-555-3131
6	S0415	Art Peters	Manager	$42,000.00	D0003	14 - 8th Avenue, New York, NY 10012	212-371-3000

22

Update Anomalies

- Tables that contain redundant information may potentially suffer from update anomalies
- Types of update anomalies include:
  - Insertion – how do you insert details of a new distribution center that has no employees?
  - Deletion – when we delete the last employee in a distribution center, we lose the information about the distribution center
  - Modification – changes to a distribution center must be made for all records containing that distribution center

23

Better Design

StaffDistributionCenter — Staff

ID	staffNo	staffName	position	salary	dCenterID
1	S1500	Tom Daniels	Manager	$48,000.00	1
2	S0003	Sally Adams	Assistant	$30,000.00	1
3	S0010	Mary Martinez	Manager	$51,000.00	2
4	S3250	Robert Chin	Assistant	$33,000.00	2
5	S2250	Sally Stern	Manager	$48,000.00	4
6	S0415	Art Peters	Manager	$42,000.00	3

Staff — DistributionCenters

ID	dCenterNo	dAddress	dTelNo
1	D0001	8 Jefferson Way, Portland, OR 97201	503-555-3618
2	D0002	City Center Plaza, Seattle, WA 98122	206-5556756
3	D0003	14 - 8th Avenue, New York, NY 10012	212-371-3000
4	D0004	2 W. El Camino, San Francisco, CA 94087	822-555-3131

24

First Normal Form (1NF)

- All rows must have the same number of columns
- Single valued attributes only

No universal agreement as to what would disqualify a table from being in 1NF

Customer

Customer ID	First Name	Surname	Telephone Number
123	Robert	Ingram	555-861-2025
456	Jane	Wright	555-403-1659 555-776-4100
789	Maria	Fernandez	555-808-9633

Typical violation of 1NF

Resist the temptation to include repeated fields as CSV text

25

Example – Table not 1NF

Staff | DistributionCenters

ID	dCenterNo	dAddress	dTelNo
1	D0001	8 Jefferson Way, Portland, OR 97201	503-555-3618, 503-555-3619
2	D0002	City Center Plaza, Seattle, WA 98122	206-5556756
3	D0003	14 - 8th Avenue, New York, NY 10012	212-371-3000
4	D0004	2 W. El Camino, San Francisco, CA 94087	822-555-3131

Repeated field

26

Converting to 1NF

DistributionCenters

ID	dCenterNo	dAddress	dTelNo
1	D0001	8 Jefferson Way, Portland, OR 97201	503-555-3618, 503-555-3619
2	D0002	City Center Plaza, Seattle, WA 98122	206-5556756
3	D0003	14 - 8th Avenue, New York, NY 10012	212-371-3000
4	D0004	2 W. El Camino, San Francisco, CA 94087	822-555-3131

Replace a repeating group with a foreign key relationship

DistributionCenters1NF

ID	dCenterNo	dAddress
1	D0001	8 Jefferson Way, Portland, OR 97201
2	D0002	City Center Plaza, Seattle, WA 98122
3	D0003	14 - 8th Avenue, New York, NY 10012
4	D0004	2 W. El Camino, San Francisco, CA 94087

DistribtionCentersTelephones

ID	dCenterID	dTelNo
1	1	503-555-3618
2	1	503-555-3619
3	2	206-5556756
4	3	212-371-3000
5	4	822-555-3131

27

Second Normal Form (2NF)

- Violated when a non-key column is a fact about part of the primary key
- A column is not fully functionally dependent on the primary key
  - `customer-credit` in this case

Mainly applies to tables with multiple natural keys

order			
itemno	customerid	quantity	customer-credit
12	57	25	OK
34	679	3	POOR

28

Third Normal Form (3NF)

- Violated when a non-key column is a fact about another non-key column, restated as
  - A column is not fully functionally dependent on the primary key

stock		
stock code	nation	exchange rate
MG	USA	0.67
IR	AUS	0.46

Exchange rate is a fact about a nation

29

Example - not 3NF

Values in staffNo, staffName, position and salary are determined from ID

Values in dAddress and dTelNo can be determined from dCenterNo

StaffDistributionCenter

ID	StaffNo	staffName	position	salary	dCenterNo	dAddress	dTelNo
1	S1500	Tom Daniels	Manager	$48,000.00	D0001	8 Jefferson Way, Portland, OR 97201	503-555-3618
2	S0003	Sally Adams	Assistant	$30,000.00	D0001	8 Jefferson Way, Portland, OR 97201	503-555-3618
3	S0010	Mary Martinez	Manager	$51,000.00	D0002	City Center Plaza, Seattle, WA 98122	206-5556756
4	S3250	Robert Chin	Assistant	$33,000.00	D0002	City Center Plaza, Seattle, WA 98122	206-5556756
5	S2250	Sally Stern	Manager	$48,000.00	D0004	2 W. El Camino, San Francisco, CA 94087	822-555-3131
6	S0415	Art Peters	Manager	$42,000.00	D0003	14 - 8th Avenue, New York, NY 10012	212-371-3000

30

Conclusion

- Be prepared for your design review

31