

CIS4301 Notes: Database Design

Ryan Roden-Corrent

March 17, 2014

Note: Some class resources available here:

Database Systems Textbook

Inference Rules

1 Functional Dependencies and Normalization

1.1 Outline

- Informal Design Guidelines for Relation Schemas
- Functional Dependencies
- normal forms based on primary keys

1.2 Outline

- Levels at which we can discuss **goodness** of relational schemas
 - Logical (conceptual)
 - Implementation (physical storage)
- Approaches to DB design
 - Top down: start with large table, break down to details
 - Bottom Up: start with details, merge to create larger structure

1.3 Informal Design Guidelines

1.3.1 Measures of Quality

- make attribute semantics clear
- Reduce redundant info in tuples

- Reduce NULL values in tuples
- Disallow possibility of creating spurious tuples

1.3.2 Natural Join Example

Schema:

EMP_LOCS(Ename, Plocation)

EMP_PROJ(SSN, Pnumber, Hours, Pnme, Plocation)

Want to use natural join to find out how many hours each employee worked at a location
see this and this for the textbook problems

note that the natural join introduces duplicate information **meaning is ambiguous**.

update anomaly: need to update data in more than one spot. Sometimes these are unavoidable, so must be **documented well**.

Triggers are useful in these, but can get messy.

1.3.3 Redundant Infor and Update Anomalies

- Types: Insertion, Deletion, Modifications
- Result of storing natural joins of base relations
- Significant effect on storage space

Deletion Anomaly Example:

Team Name	Player	Playerid	pts
NULL	12	12	2

Player 12 kicked off team, but still exists

Insertion Anomaly Example:

Team Name	Player
Gators	
Louisville	
?	12

Must know team name before inserting player.

1.4 Why are NULLs bad?

- Way group many attributes together into a "fat" relation
- Wasted storage space (reserves space in every column for that row)

Suppose an essay relation has a text field that is a char(40000). A null essay will still reserve this space.

1.4.1 Guideline 4

- Design relation schemas to be joined with equality conditions on attributes that are appropriately related
- guarantees no spurious tuples generated

1.5 Summary

- anomalies cause redundant work
- NULL wastes space

2 Functional Dependencies

denoted by $X \rightarrow Y$

for any two tuples t_1 and t_2 in r that have $t_1[X] = t_2[X]$, there must be a mapping $t_1[Y] = t_2[Y]$

	W	X	Y	Z
t_1		billy d	gators	
t_2		billy d	gators	

- formal tool for analysis of relational schemas
- detect and describe some of the above problems in precise terms
- theory of functional dependency

2.1 normal forms based on primary keys

- normalization process
- Approaches for relational schema design
- takes a relation schema through a series of tests
 - certify whether it satisfies a certain normal form
 - proceeds in top-down fashion
- ideally only want functional dependencies on primary keys (but there is some leeway)

Example: