Concepts from 3.1-3.2

- Functional dependencies
- Keys & superkeys of a relation
- Reasoning about FDs
- Closure of a set of attributes
- Closure of a set of FDs
- Minimal basis for a set of FDs

Plan

- How can we use FDs to show that a relation has an anomaly (a potential problem)?
- How can we algorithmically fix the problem?

Projecting sets of FDs

- Suppose we have a relation R and set of FDs F
- Let S be a relation obtained by projecting R into a subset of the attributes of R $\pi_{Attributes}(R)$
- The **projection** F_S of F is the set of FDs that follow from F and hold in S
  - Involve only attributes of S

Projecting sets of FDs

- Algorithm for computing F_S :
  - Compute closure F^+
  - F_S is the set of all FDs in F^+ that involve only the attributes in S
- Book describes a different algorithm in section 3.2.8.
- Book's algorithm also shows how to compute a minimal basis of F_S

Projecting sets of FDs

- R(A, B, C, D); F = {A→B, B→C, C→D}
- Which FDs hold in S(A, C, D)?

F^+ is {A→B, B→C, C→D, A→C, A→D, B→D}

F_S is {C→D, A→C, A→D}

Anomalies

- An anomaly is a problem that arises when we try to add too many attributes to a single relation.
- Arises from ***redundancy***: information repeated unnecessarily.
  - When designing schemas, try to ensure you never repeat yourself!

title	year	length	genre	studio	star
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	Drama	MGM	Vivien Leigh
Wayne's World	1992	95	Comedy	Paramount	Dana Carvey
Wayne's World	1992	95	Comedy	Paramount	Mike Meyers

Anomalies

- Update anomaly: when you change information in one tuple but leave the same information in a different tuple unchanged.

title	year	length	genre	studio	star
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	Drama	MGM	Vivien Leigh
Wayne's World	1992	95	Comedy	Paramount	Dana Carvey
Wayne's World	1992	95	Comedy	Paramount	Mike Meyers

Anomalies

- Deletion anomaly: when deleting one or more tuples removes information that we didn't want to lose.

title	year	length	genre	studio	star
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	Drama	MGM	Vivien Leigh
Wayne's World	1992	95	Comedy	Paramount	Dana Carvey
Wayne's World	1992	95	Comedy	Paramount	Mike Meyers

Anomalies

- Insertion anomaly (left out of book): when storing a piece of information forces us to store an unrelated piece of information as well.

title	year	length	genre	studio	star
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	Drama	MGM	Vivien Leigh
Wayne's World	1992	95	Comedy	Paramount	Dana Carvey
Wayne's World	1992	95	Comedy	Paramount	Mike Meyers


DECOMPOSITION
a poem by Jordan
In the cold winter time, plant cells freeze and thaw;
This process is just part of nature's law.
The cell walls burst and get all squishy...
That's why this old pumpkin here is so mushy!
(Eventually this pumpkin, like all dead plants, will "decompose," or turn back into soil! The process is called "decomposition.")

Decomposing Relations

- Given a relation R(A1, A2…, An), two relations S(B1, B2…, Bm) and T(C1, C2…, Ck) form a decomposition of R if:

1. the attributes of S and T together make up the attributes of R, i.e., {A's} = {B's} U {C's}
2. the tuples in S are the projections into {B1…Bm} of the tuples of R i.e. $S \equiv \pi_{B1,B2,..,Bm}(R)$
3. the tuples in T are the projections into {C1…Ck} of the tuples of R i.e. $T \equiv \pi_{C1,C2,..,Ck}(R)$

title	year	length	genre	studio	star
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	Drama	MGM	Vivien Leigh
Wayne's World	1992	95	Comedy	Paramount	Dana Carvey
Wayne's World	1992	95	Comedy	Paramount	Mike Meyers

- Decompose into
  - Movies(title, year, length, genre, studio)
  - Stars(title, year, star)
- Are the anomalies removed? Is anything redundant? Why or why not? Do you see a connection to FDs?

BCNF

- Anomalies are guaranteed not to exist when a relation is in ***Boyce-Codd normal form*** (BCNF).
- A relation R is in BCNF iff whenever there is a nontrivial FD $A_1 \ldots A_n \rightarrow B_1 \ldots B_m$ for R, $\{A_1, \ldots, A_n\}$ is a superkey for R.
- Informally, the left side of every nontrivial FD must be a superkey.

Check for BCNF violations

- List all nontrivial FDs in R.
- Ensure left side of each nontrivial FD is a superkey.
- (First have to find all the keys!)

  Note: a relation with two attributes is always in BCNF.

title	year	length	genre	studio	star
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	Drama	MGM	Vivien Leigh
Wayne's World	1992	95	Comedy	Paramount	Dana Carvey
Wayne's World	1992	95	Comedy	Paramount	Mike Meyers

- Decompose into
  - Movies(title, year, length, genre, studio)
  - Stars(title, year, star)
- What are the new FDs and keys?

Example….

- Is Courses(Number, DepartmentName, CourseName, Classroom, Enrollment, StudentName, Address) in BCNF?
- FDs:
  - Number DepartmentName → CourseName
  - Number DepartmentName → Classroom
  - Number DepartmentName → Enrollment
- What is {Number, DepartmentName}$^+$ under the FDs?

{Number, DepartmentName, Coursename, Classroom, Enrollment}

- So the key is {Number, DepartmentName, StudentName, Address}
- So the relation is not in BCNF.

Decomposition into BCNF

- Suppose R is a relation schema that violates BCNF
- We can decompose R into a set *S* of new relations such that:
  - each relation in *S* is in BCNF and
  - we can "recover" R from the relations in *S*, i.e., we can reconstruct R exactly from the relations in *S*

Algorithm: Given relation R and set of FDs F:

- Check if R is in BCNF, if not, do:
- If there are FDs that violate BCNF, call one X -> Y. Compute X^+. Let R1 = X^+ and R2 = X and all other attributes not in X^+.
- Compute FDs for R1 and R2 (projection algorithm for FDs).
- Check if R1 and R2 are in BCNF, and repeat if needed.

title	year	length	genre	studio	star
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	Drama	MGM	Vivien Leigh
Wayne's World	1992	95	Comedy	Paramount	Dana Carvey
Wayne's World	1992	95	Comedy	Paramount	Mike Meyers

FDs: title year -> length genre studio
Key: {title, year, star}
This relation is not in BCNF (the single FD is a violation because the LHS (title, year) is not a superkey).
Decompose:
- Compute $\{title, year\}^+$ = {title, year, length, genre, studio}
- New relation: R1(title, year, length, genre, studio). Key = {title, year}
- New relation: R2(title, year, star). Key = {title, year, star}
- FDs for R1: Same FD as for original relation. FDs for R2: none
- No BCNF violations in R1 or R2. (LHS of FD in R1 is a superkey.)

- Schema is Courses(Number, DeptName, CourseName, Classroom, Enrollment, StudentName, Address)
- BCNF-violating FD is

Number DeptName → CourseName Classroom Enrollment

- What is {Number, DeptName}+ ?

{Number, DeptName, CourseName, Classroom, Enrollment}

- Decompose Courses into

Courses1(Number, DeptName, CourseName, Classroom, Enrollment)

and

Courses2(Number, DeptName, StudentName, Address)

Are there any BCNF violations in the two new relations?

Students and Profs

- Suppose we have one single relation with attributes:
  - R#
  - StudentName
  - ProfID (ID of professor teaching a class with the student)
  - ProfName
  - AdvisorID
  - AdvisorName