Advanced Data Structures

COMP282
Normal Forms

Motiviation

- Data can be organized a variety of ways
- Poor choices concerning organization can lead to database anomalies (corruption) or difficulty in keeping information accurate.
- Normal forms prevent certain anomalies from occuring.

First Normal Form (1NF)

- 1NF is rather easy.
  - simply stated: attributes can only contain values
  - attributes cannot consist of sets of values
- It might be useful to want to store information about songs
- Madonna produced Erotica and Ray of Light.
  - it would not 1NF if relation consisted of the attribute albums and it was assigned the set {Erotica, Ray of Light}

All modern RDMS are 1NF

- Oracle, MySQL, et al. all produce 1NF by default. The SQL language doesn't provide for non-1NF queries.

1NF anomalies

- 1NF databases have some problems:
- Most notable:
  - repetition of data
  - to change a department name all tuples of the relation need to be updated since the department name can exist in multiple rows.

Partial Dependencies

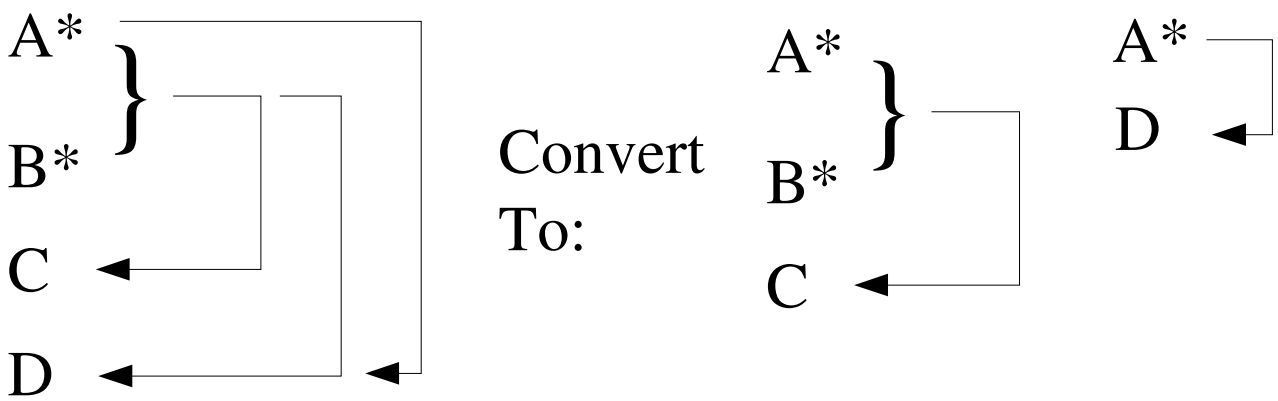
- if $X \rightarrow Y$ and $W \rightarrow Y$ and $W \subset X$
  then Y is said to be partially dependent on X
- Essentially You could restrict X a bit and still functionally determine Y.
- 2NF requires knowledge of this.

Second Normal Form (2NF)

- A relation r(R) is in Second Normal Form (2NF) iff:
  - r(R) is already in 1NF
  - each attribute that is not part of a candidate key is fully dependent on every candidate key

2NF example

- In the following example D is partially dependent on AB

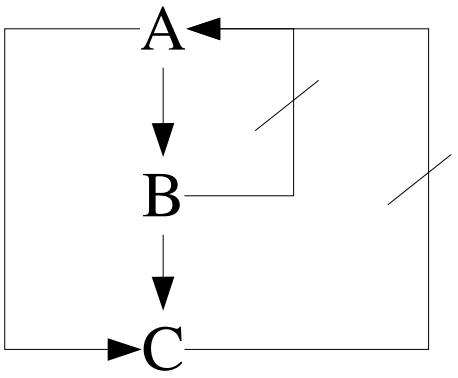


2NF Anomalies

- Assume that Emp-Dpt → Emp-Hrly-Rate
  - We cannot Set the Emp-Hrly Rate until an employee is added with that Dpt
  - We cannot delete the last employee of a department without losing the department's hourly rate.
  - If the department hourly rate changes the entire table must be edited and all employees working in that department need to be updated.

Transitive Dependencies

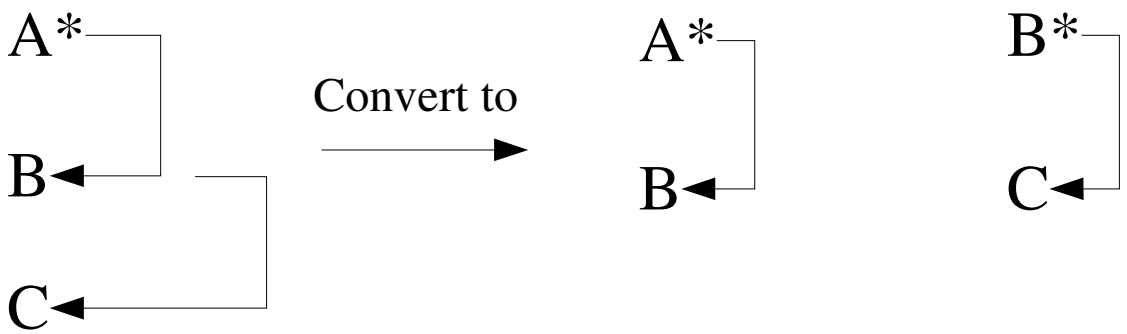
- This is the graphical depiction of a transitive dependency.



- Third normal form removes these from relations

Third Normal Form (3NF)

- A relation r(R) is in Third Normal Form (3NF) iff:
  - r(R) is already in 2NF
  - No nonprime attribute functionally determines any other nonprime attribute.



3NF Anomalies

East Coast	SQL 101	Introduction	8:00-10:00	Atlanta Educational Center
East Coast	SQL 101	Intermediate	10:00-12:00	New York Educational Center
West Coast	SQL 101	Advanced	8:00-10:00	Los Angeles Educational Center

Boyce-Codd Normal Form (BCNF)

- Boyce-Codd Normal Form requires:
  - The relation is (1NF)
  - for every FD $X \rightarrow A$, we have either $A \subset X$ or X is a superkey of r(R)
- The second criteria implies 3NF (All BCNF relations are also 3NF but not all 3NF relations are BCNF)