Data and Its Applications (CS4.301)

Monsoon 2021, IIIT Hyderabad
23 September, Wednesday (Lecture 10)

Taught by Prof. Kamal Karlapalem

Relational Data Model

Functional Dependencies (contd.)

Some additional inference rules are

- projectivity: if $X \to YZ$, then $X \to Y$ and $X \to Z$.
- additivity: if $X \to Y$ and $X \to Z$, then $X \to YZ$.
- pseudotransitivity: if $X \to Y$ and $WY \to Z$, then $WX \to Z$.

These are provable from the first three inference rules.

If X determines some other subset $Y \subseteq \{A_1, \dots, A_n\}$, but $X - A \to Y$ for some A, then Y is partially functionally dependent on X. Otherwise, Y is fully functionally dependent on X.

Further, $X \to Y$ is called a transitive dependency if there exists some $Z \subseteq \{A_1, \dots, A_n\}$ such that $X \to Z$ and $Z \to Y$.

Keys

Given a relation $R(A_1, \dots, A_n)$, we know that the tuple of attributes $(A_1, \dots, A_n)$ has a unique value for each row in the relation.

If some $X \subseteq \{A_1, \dots, A_n\}$ is a superkey, we know that $X \to (A_1, \dots, A_n)$ holds. Further, X cannot be minimised (it is a key) if there is no A such that $X - A \to (A_1, \dots, A_n)$ (*i.e.* removing any attribute from X makes it no longer a superkey). A relation can have multiple keys (which are all called candidate keys).

An attribute A is a prime attribute if it belongs to some key of R, and non-prime otherwise.

Normalisation

Normalisation is a systematic process of decomposing relation schemas into smaller ones that possess desirable properties.

There is a series of normal forms, each one satisfying a successively increasing set of rules.

First Normal Form (1NF)

A relation schema R is in 1NF if the values in the domain of each attribute R are atomic.

If some attribute has multiple values, we either split the tuple into multiple tuples, or move the attribute to a new relation schema.

Second Normal Form (2NF)

A relation schema R is in 2NF if every non-prime attribute A in R is fully functionally dependent on every key of R.

If some non-prime attribute has a partial FD on some key, we remove it from the schema and add it to a new schema with a key that it is fully functionally dependent on.

Third Normal Form (3NF)

A relation schema R is in 3NF if all non-trivial dependencies are of the form $X \to A$, where either
X is a superkey, or
A is a prime attribute.

Together, these two conditions ensure that there are no transitive dependencies. If they are not satisfied, we remove the dependency and create a new relation schema with only X and A.

Boyce-Codd Normal Form (BCNF)

A relation R is in BCNF if all non-trivial dependencies are of the form $X \to A$, where X is a superkey.

As before, we remove violating dependencies and add the attributes involved to a new relation schema.

Closures

The closure F^+ of a set F of functional dependencies can be computed by repeatedly checking for the sets of attributes determined by attributes within F, *i.e.*, it should satisfy $Y \in F^+$ for all $X \to Y$ such that $X \in F^+$.

Two closures are equivalent if $F^+ = G^+$.

For each FD $X \to Y$ in F, if $Y \in X^+$ (computed w.r.t. G), then G is said to cover F.