Лекция 6

Характеристики функциональных зависимостей

Функциональная зависимость является смысловым (семантическим) свойством атрибутом отношений. Она определяется только содержательно.

Если B зависит от A: если две строки имеют одно и то же значение атрибута A, то они обязательно имеют одно и тоже значение атрибута B, но не наоборот.

Пусть R является переменной отношения, а x и y произвольными подмножествами множества атрибутов переменной отношения R. Y функционально зависимо от x тогда и только тогда, когда для любого допустимого значения переменной отношения R существует единственное значение другой переменной. Если два кортежа переменной отношения R совпадают по значению x, то они также совпадают и по y. В этом случае подмножество x называется детерминант, а y – зависимая часть.

Содержательно функциональная зависимость является тривиальной, если она остаётся справедливой при любых условиях.

Функциональная зависимость тривиальна тогда и только тогда, когда её правая (зависимая) часть является подмножеством её левой части, т.е. детерминантом.

ФЗ является свойством абстрактной структуры т.е. реляционной схемы, а не свойством конкретного экземпляра.

В процессе проектирования рассматриваются только те ФЗ, которые: определяют связь один к одному между атрибутами левой и правой частей выражения зависимости, остаются справедливыми при любых условиях, являются не тривиальными.

На основании рассмотрения этих тезисов можно сузить этих ключей до первичного ключа с помощью этого правила: все атрибуты, которые не входят в состав первичного ключа, должны быть ФЗ от этого ключа.

Аксиомы Армстронга. Минимальное множество ФЗ.

Полный набор ФЗ для определённого отношения как правила является очень большим. Необходим подход, который позволит уменьшить размеры необходимого множества ФЗ. Основы такого подхода – некоторые ФЗ можно вывести из других ФЗ.

Замыкание x\* - это множество всех ФЗ, которые могут быть выведены из заданного множества ФЗ x. Для вычисления x\* используется набор правил вывода, называемы Аксиомы Армстронга.

1) Рефлексивность. Если B – подмножество A, то A -> B

2) Дополнение. A -> B, то (A, C) -> (B, C)

3) Транзитивность. A -> B, B -> C => A -> C

Свойства набора аксиом

1) Каждое из этих правил можно обосновать непосредственно из определения ФЗ

2) Набор является полным, т.е. все ФЗ производные от x можно вывести из x с помощь этих правил.

3) Набор является не противоречивым.

Следовательно эти правила могут применяться для получения x\*, т.е. замыкания.

4 дополнительных аксиомы:

4) Самоопределение A -> A

5) Декомпозиция A -> (B, C) => A -> B и A -> C

6) Объединение A -> B и A -> C => A -> (B, C)

7) Композиция A -> B и C -> D => (A, C) -> (B, D)

Правило 1+4 вместе вырабатывают тривиальные зависимости.

Правило 2 дополнение одного и того же атрибута к левой и право части приводит к получению к действительной зависимости

Правило 5 позволяет разложить ФЗ на ряд более простых.

Правило 6 – объединение, позволяет сделать обратную операцию, объединить простые таблицы в сложную

Правило 7 – указывает, что для получения ещё одной действительной зависимости можно объединять ряд неперекрывающихся зависимостей.

Множество ФЗ Y покрывается множеством ФЗ x, если каждая ФЗ из y присутствует также в замыкании x\*.

Множество ФЗ x является минимальным, если три условия:

1) Каждая зависимость x имеет единственный атрибут в правой части.

2) Ни одну зависимость A -> B нельзя заменить зависимостью C -> B, где С – собственное подмножество А, и при этом получить множество зависимости, эквивалентных x.

Все избыточные атрибуты в левой части зависимости должны быть удалены.

3) Из множества x нельзя удалить ни одной зависимости, и при этом получить множество зависимостей эквивалентные x. Исключаются зависимости, из которых могут быть выведены из остальных ФЗ множества x

К сожалению, может существовать несколько множеств минимальной зависимости.

Минимально покрытие множества ФЗ x – это минимальное множество зависимостей x(min), эквивалентное x.

Процесс нормализации

Нормализация – это формальный метод анализа отношений на основе первичного ключа (потенциальных ключей) и существующих ФЗ.

Уровней нормализации существует 5-6.

Общая идея: если некоторое требование, противоречащее данному уровню нормализации, не удовлетворяет, то противоречащее должно быть разделено на несколько отношений, каждое из которых в отдельности удовлетворяет требованиям нормализации.

Чаще всего нормализация осуществляется в виде нескольких последовательных этапов, выполняемых повышением уровня. Минимальными необходимы требования первым нф. Все остальные формы по желанию проектантов. В подавляющем большинстве существенно важно избежать аномалий обновления, т.е. проводить нормализацию до 3 нф.

Основной информации для нормализации является пара

1) множества ФЗ для каждого отношения

2) В каждом отношении назначен первичный ключ

Ненормализованная форма – это таблица, содержащая одну или несколько повторяющихся групп данных.

Повторяющаяся группа – группа, состоящая из одного или нескольких атрибутов таблицы, в которой возможно несколько значений для единственного значения ключевого атрибута.

1 НФ – это отношение, в котором на пересечении каждой строки и каждого столбца содержится одно и только одно значение.

Существуют два способа исключения повторяющихся групп:

1) Пустые места заполняются дубликатами неповторяющихся данных – выравнивание таблицы.

2) Один атрибут или группа назначаются ключом ненормализованной таблицы, затем повторяющиеся группы изымаются и помещаются в отдельные отношения вместе с копией ключа.

Оба способа являются одинаково обоснованными, однако второй приводит к меньшей избыточности данных.

2 НФ основана на понятии полной функциональной зависимости

Функциональная зависимость A -> B является полной, если удаление какого-либо атрибута из А приводит к утрате этой зависимости

Функциональная зависимость A -> B является частичной, если в а есть некий атрибут, при удалении которой зависимость сохраняется

2 НФ применяется к отношениям с составными ключами, т.е. к отношениям, у которых первичный ключ состоит из более чем одного атрибута.

2 НФ – это отношение, которое уже находится в 1 НФ и каждый атрибут которого, не входящий в состав первичного ключа, характеризуется полной функциональной зависимостью от этого первичного ключа.

Нормализация отношений 1 НФ до 2 НФ состоит в устранении частичных зависимостей.

Если в отношении между атрибутами существует частичная зав, то функционально-зависимые атрибуты удаляются из него и помещаются в новое вместе с копией их детерминанта.

Перевод отношений в 3 НФ имеет целью устранение транзитивных зависимостей.

Транзитивная зависимость состоит в следующем: если для атрибутов А, В, С, существует зависимость A -> В и B -> C, это означает, что A транзитивно зависит от C через B. При условии, что атрибут A функционально не зависит ни от B, ни от C.

3 НФ – это отношение, которое уже находится в 2 НФ и не имеет атрибутов, не входящих в первичный ключ, которые находились бы в транзитивной функциональной зависимости от этого первичного ключа.

Удаляем транзитивно-зависимые атрибуты со своим детерминантом, помещаем в новые таблицы вместе с копией детерминанта.

При переводе в 2 НФ и 3 НФ необходимо рассматривать не только назначенный первичный ключ, но и все потенциальные ключи.

На практике чаще всего это является избыточным.

Нормальная форма Бойса-Кодда

Отношения находятся в НФБК тогда и только тогда, когда каждый его атрибут (детерминант) является потенциальным ключом.

Использование НФБК может приводить к проблемам, в частности можем потерять некие важные зависимости. Особенно проявляется, когда делаем соединения.

4 НФ призвана устранить многозначные зависимости.

Многозначная зависимость – это такая зависимость между атрибутами А, В, С, если выполняются A -> B, A -> C, но множество значений для B и С не зависят друг от друга.

Многозначная зависимость может быть тривиальной или не тривиальной.

Многозначная зависимость является тривиальной, если атрибут B или C является подмножеством атрибута A.

Для предотвращения аномалий отношения декомпозируем.

Если к исходной переменной отношения добавить атрибут функционально зависящий от потенциального ключа, то традиционная декомпозиция не подходит.

5 НФ. При любой декомпозиции отношения на два других результат обладает свойством соединения без потерь, но иногда требуется сделать декомпозицию на более, чем два отношения.

Зависимость соединения без потерь – это свойство декомпозиции, которое гарантирует отсутствие фиктивных строк при восстановлении первоначального отношения с помощью операции естественного соединения.

Зависимость соединения без потерь – это свойство содержательное. Т.е. оно определяется не механически, а из анализа предметной области.

Отношение находится в 5 НФ (в проекционно-соединительно нормальной форме) тогда и только тогда, когда каждая не тривиальная зависимость соединения в нём определяется потенциальными ключами этого отношения.

Содержательная трактовка нормальных форм

1 НФ характеризуется 8 пунктами:

1) В исходной реляционной таблице каждое значение любого атрибута делается атомарным.

2) В таблице отсутствуют одинаковые строки

3) Каждый столбец уникально поименован именем атрибута, локализованным в таблице, и содержит текущее значение этого атрибута.

4) Каждый атрибут ассоциирован с определённым доменом. (Абстрактным типом данных)

5) Если в реляционной таблице имеются такие функциональные зависимости, которые решено учесть при работе с этой таблицей, то эти зависимости формируются в условии целостности БД

6) Записи упорядочены явно или неявно. Известным или неизвестным пользователю способом.

7) Каждый из компонентов первичного ключа не долен быть пустым, т.е. не должен иметь значение null

8) Если ключи определяются на уровне 1 НФ, то при их определении должно соблюдаться условие ссылочной целостности для внешних ключей.

2 НФ характеризуется одним пунктом:

1) Выполняются все требования 1 НФ+ полная функциональная зависимость каждого из не ключевых атрибутов от первичного либо потенциального ключа. Для этого все частичные зависимости путём факторизации выделяются в отдельные таблицы и рассматриваются отдельно.

3 НФ (классическая форма) характеризуется одним пунктом:

1) Все требования 2 НФ + отсутствие транзитивной зависимости между атрибутами.

НФБК одним пунктом:

1) Все требования 3 НФ +отсутствие зависимостей ключевых атрибутов от не ключевых атрибутов

Важный комментарий: может возникнуть необходимость коррекции первичного ключа при коррекции не ключевых атрибутов.

4 НФ – удовлетворяется, если в реляционной таблице отсутствуют привнесённые зависимости

G (A, B, C, D, E, F)

A -> B

A -> C

B -> C

DA -> E

DAF -> E

По аксиоме Армстронга

A -> B

A -> C

DB -> E

Атрибут F «вылетает» из БД, он ни от чего не зависит, от него ничего не зависит, тогда последняя зависимость привнесённая.

Замечание: один и тот же атрибут в одной и той же таблице может рассматриваться как привнесённый, так и не привнесённый. Это определяется только в условии целостности всей БД.

5 НФ. Реляционная таблица находится в 5 НФ если и только если любая не тривиальная полная её декомпозиция на проекции по потенциальным ключам при соединении даёт её саму. (Если таблица разложима без потерь)