

# Нормализация и денормализация данных

## Цели

В этом разделе Вы будете изучать:

- Восходящий и нисходящий методы проектирования
- Избыточность данных
- Первые, вторые и третьи нормальные формы
- Нормальные формы Бойса - Кодда (BCNF)
- Понятие о необходимости денормализации

# Нормализация и денормализация данных

## СРС 1

1. Для ситуации, в которой студент может работать только с одним проектом, и никакой другой студент не может работать над этим же проектом, связь между студентом и проектом называется\_\_\_\_\_
  - a. Один-к-одному
  - b. Один-к-многим
  - c. Многие-к-одному
  - d. Многие-К-многим
2. Какое из следующих утверждений истинно?
  - a. Первичный ключ супертипа является первичным ключом подтипа.
  - b. Внешний ключ супертипа является первичным ключом подтипа.
    1. Первичный ключ супертипа является внешним ключом подтипа.
    2. Внешний ключ супертипа является внешним ключом подтипа.

# Нормализация и денормализация данных

## СРС 1 (продолжение..)

3. Ключ, который может быть, но не является первичным ключом, называется а(n) \_\_\_\_\_ ключом.
  - a. Возможный ключ
  - b. Внешний ключ
  - c. Альтернативный ключ
  - d. Составной ключ
4. Какие проблемы возникнут, если разрешить значения NULL для первичного ключа?
  - a. Затруднится уникальная идентификация строк.
  - b. Затруднится уникальная идентификация столбцов.
    - Затруднится объединение таблиц.
    - Затруднится определение внешнего ключа.
6. В \_\_\_\_\_ каждая сущность более высокого уровня должна быть сущностью низкого уровня.
  3. Обобщение
  4. Схема С/С
  5. Конкретизация
  6. Связь «многие-к-многим»

# Нормализация и денормализация данных

## Нисходящий и восходящий подход

- Существует два подхода к логическому проектированию базы данных:
  - Нисходящий подход
  - Восходящий подход
- Методика С/С моделирования является нисходящим подходом. Она включает определение сущностей, связей и атрибутов, вычерчивание схемы С/С и отображение схемы в.
- Нормализация является подходом снизу вверх. Это пошаговое разложение сложных записей на простые.
- Нормализация понижает избыточность, используя принцип расщепления.
- Расщепление представляет собой преобразование таблицы в таблицы меньшего размера без потери информации.
- Нисходящее приближение наилучшим образом подходит для тестирования существующих разработок.

# Нормализация и денормализация данных

## Избыточность данных

- Избыточность данных означает их повторяемость.
- Избыточность увеличивает время, затрачиваемое на обновление, добавление и удаление данных.
- Избыточность также повышает использование дискового пространства, и, как следствие, увеличивает число обращений к диску.
- Следствием избыточности могут быть:
  - Аномалии обновления — вставка, замена и удаление данных могут вызвать ошибки.
  - Противоречивость — вероятность появления ошибок возрастает при записи повторяющихся фактов.
  - Неоправданный расход дискового пространства.

# Нормализация и денормализация данных

## Необходимость нормализации

- Нормализация является научным методом расщепления сложных табличных структур на простые по определенным правилам.
- Пользуясь нормализацией, Вы можете понизить избыточность в таблице, устранить проблемы противоречивости и неоправданного расходования пространства на диске.
- Нормализация обеспечивает отсутствие потерь информации.
- Нормализация дает ряд преимуществ:
  - Ускоряет сортировку и создание индексов.
  - Помогает создавать групповые индексы.
  - Требуется меньше индексов на каждую таблицу.
  - Понижает количество значений NULL в таблице.
  - Делает базу данных компактной.

# Нормализация и денормализация данных

## Необходимость нормализации (продолжение)

- Производительность приложения прямо связана с тем, как спроектирована база.
- Для того, чтобы обеспечить качество проектирования, следует придерживаться некоторых правил :
  - Каждая таблица должна иметь идентификатор.
  - В каждой таблице должны храниться данные для одного типа сущности.
  - Следует избегать столбцов, принимающих значения NULL.
  - Следует избегать повторяющихся значений или столбцов.

# Нормализация и денормализация данных

## Нормальные формы

- Нормализация состоит в создании таблиц, удовлетворяющих установленным правилам и представляющих определенные нормальные формы.
- Нормальные формы применяют для того, чтобы предотвратить проникновение в базу данных каких-либо аномалий и противоречий.
- Табличная структура всегда находится в определенной нормальной форме.
- Наиболее важными и широко используемыми нормальными формами являются :
  - Первая нормальная форма (1NF)
  - Вторая нормальная форма (2 NF)
  - Третья нормальная форма (3 NF)
  - Нормальная форма Бойса-Кодда (BCNF)



# Нормализация и денормализация данных

## Функциональная зависимость

- Теория нормализации базируется на фундаментальном понятии функциональной зависимости.
- В отношении  $R$  атрибут  $A$  является функционально зависимым от атрибута  $B$ , если каждое значение  $A$  в  $R$  ассоциируется только с одним значением  $B$ .
- Атрибут  $B$  называется детерминантом.
- Все атрибуты в таблице должны быть функционально зависимыми от ключа. Однако, атрибут не должен быть ключом для того, чтобы функционально определять другие атрибуты .
- Функциональной зависимости можно дать и такое определение: В данном отношении  $R$  атрибут  $A$  функционально зависит от  $B$ , если для согласования двух кортежей, принадлежащих  $R$ , по их значениям  $B$ , они должны быть согласованы по значению  $A$ .
- Функциональными зависимостями представлены связи "многие-к-одному".

# Нормализация и денормализация данных

## Первая нормальная форма (1 NF)

- Таблица находится в 1 NF, если каждая ячейка таблицы одно определенное значение.

# Нормализация и денормализация данных

## Вторая нормальная форма (2 NF)

- Таблица находится в 2 NF, когда она в 1 NF и каждый атрибут в строке функционально зависит от ключа в целом, а не только от его составной части.
- Указания по преобразованию таблицы в 2 NF:
  - Найдите и удалите атрибуты, которые функционально зависят лишь от части ключа, а не от ключа в целом. Поместите их в отдельную таблицу.
  - Сгруппируйте оставшиеся атрибуты.

# Нормализация и денормализация данных

## Третья нормальная форма (3 NF)

- Таблица находится в 3 NF, если она в 2 NF и каждый атрибут, не являющийся ключом, функционально зависит только от первичного ключа.
- Указания по преобразованию таблицы в 3 NF:
  - Найдите и удалите не ключевые атрибуты, которые функционально зависят от атрибутов, не являющихся первичным ключом. Поместите их в отдельную таблицу.
  - Сгруппируйте оставшиеся атрибуты.

# Нормализация и денормализация данных

## Нормальная форма Бойса-Кодда (BCNF)

- Основное определение 3 NF становится неадекватным и неприемлемым для таблиц:
  - Имеющих множественные возможные ключи .
  - Имеющих составные возможные ключи.
  - Имеющих перекрывающиеся возможные ключи.
- Для нормализации таблицы при указанных условиях была предложена нормальная форма Бойса-Кодда (BCNF).
- Отношение находится в BCNF тогда и только тогда, когда каждый детерминант является возможным ключом.
- Указания по преобразованию таблицы в BCNF:
  - Найдите и удалите перекрывающиеся возможные ключи. Поместите часть возможного ключа и атрибут, от которого он функционально зависит, в отдельную таблицу.
  - Сгруппируйте оставшиеся элементы в таблицу.

# Нормализация и денормализация данных

## Денормализация

- Преднамеренный ввод избыточности в таблицу с целью повышения производительности запросов называется денормализацией.
- Решение осуществить денормализацию является компромиссом между производительностью и непротиворечивостью данных.
- Денормализация увеличивает используемое пространство на диске.

# Нормализация и денормализация данных

## Выводы

На этом занятии Вы узнали, что:

- Имеется два подхода к логическому проектированию базы данных:
  - Подход "сверху вниз"
  - Подход "снизу вверх"
- Методика С/С моделирования является подходом "сверху вниз", а нормализация - подходом "снизу вверх".
- Нормализация используется для упрощения табличных структур.
- Нормализация состоит в оформлении таблиц в соответствии с заданными условиями в виде определенных нормальных форм.
- Табличная структура всегда находится в определенной нормальной форме.

# Нормализация и денормализация данных

## Выводы (продолжение..)

- Наиболее важными и широко используемыми нормальными формами являются:
  - Первая нормальная форма (1 NF)
  - Вторая нормальная форма (2 NF)
  - Третья нормальная форма (3 NF)
  - Нормальная форма Бойса-Кодда (BCNF)
- Теория нормализации основана на фундаментальном понятии о функциональной зависимости. Функциональными зависимостями представлены связи "многие-к-многим".
- Таблица находится в 1 NF, если каждая ячейка таблицы содержит только одно значение.
- Таблица находится в 2NF, если она в 1 NF и каждый атрибут в строке зависит от всего ключа, а не от его части.
- Таблица находится в 3NF, если она в 2 NF и каждый не ключевой атрибут функционально зависит только от первичного ключа.



# Нормализация и денормализация данных

## Выводы (продолжение..)

- Основное определение 3NF становится неадекватным и не подходит для таблиц, у которых:
  - Имеются множественные возможные ключи.
  - Возможные ключи являются составными.
  - Возможные ключи перекрываются.
- Отношение находится в нормальной форме Бойса-Кодда (BCNF) тогда и только тогда, когда каждый детерминант является возможным ключом.
- Преднамеренный ввод избыточности в таблицу для повышения производительности запросов называется денормализацией.
- Денормализация является компромиссом между производительностью и непротиворечивостью данных.
- Денормализация увеличивает используемое пространство на диске.