

Нормализация данных

В реляционных базах данных структуры данных должны быть нормализованы, чтобы организация данных в таблицах была корректной.

Нормализованные структуры данных - это такие структуры, из которых удалены повторяющиеся элементы данных.

В реляционной базе данных в каждую ячейку можно поместить только один элемент данных, т.е. в одну ячейку нельзя записать несколько повторяющихся элементов данных.

Пример с повторяющимися элементами данных, при которых у одного человека может быть несколько автомобилей (данные не нормализованы):

Isiku_nr (pr. key)	Nimi	Auto
1	Lauri Tamm	678TCA 542AUN 887PPP
2	Teet Mägi	433REP

Цели нормализации данных

- уменьшение избыточности данных;
- избежать аномалий при изменении данных;
- улучшение качества данных.

Формы нормализации данных

- Первая нормальная форма (1NF);
- Вторая нормальная форма (2NF);
- Третья нормальная форма (3NF);
- Четвертая нормальная форма (4NF);
- Пятая нормальная форма (5NF);
- Шестая нормальная форма (6NF)

и т.д.

С чего начать?

- Нормализацию начинают с приведения данных к первой нормальной форме, и при необходимости продолжают нормализацию.
- Требования предыдущих нормальных форм уже содержатся в последующих.
- В реляционных базах данных рекомендуется проводить нормализацию как минимум до третьей формы, что и требуется в рамках данного предмета.
- Для начала нормализации необходимо найти все атрибуты объектов. Например, ISIK (perenimi, eesnimi, sünniaeg, sugu).

Первая нормальная форма

- Удалите из таблиц повторяющиеся группы данных;
- Создайте отдельную таблицу для каждой группы связанных между собой данных;
- Добавьте первичный ключ;
- Для повторяющихся данных создайте отдельные записи:

Повторяющиеся данные:

Isiku_nr (pr. key)	Nimi	Auto
1	Lauri Tamm	678TCA 542AUN 887PPP
2	Teet Mägi	433REP

В ячейке максимально одно значение (1NF):

Isiku_nr (pr. key)	Nimi	Auto
1	Lauri Tamm	678TCA
1	Lauri Tamm	542AUN
1	Lauri Tamm	887PPP
2	Teet Mägi	433REP

В данном случае возникает избыточность данных, которые занимают в базе данных слишком много места. Для этого необходимо избавиться от избыточности.

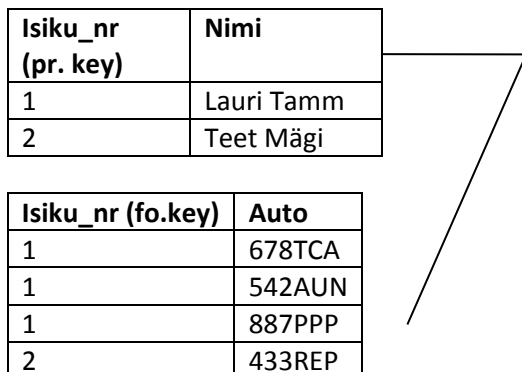
Ликвидация групп с повторяющимися столбцами

Необходимо создать отдельную таблицу:

Пример таблицы "Isik":

Isiku_nr (pr. key)	Isik	Auto1	Auto2	Auto3
1	Lauri Tamm	678TCA	542AUN	887PPP
2	Teet Mägi	433REP		

В таблице "Isik" есть столбцы "auto1", "auto2" и "auto3", которые можно объединить в отдельную таблицу "Isiku_Auto" и связать таблицы между собой через внешний ключ.



Ликвидация данных, состоящих из нескольких частей

Например, в таблице „Isik“ элемент данных "nimi" состоит из компонентов "eesnimi" ja "perenimi", которые следует разделить на два столбца.

Isiku_nr (pr. key)	Perenimi	Eesnimi
1	Tamm	Lauri
2	Mägi	Teet

Вторая нормальная форма

- Создайте отдельную таблицу для тех данных, которые используются в нескольких записях;
- Свяжите эти таблицы через внешний ключ.

Пример таблицы первой нормальной формы:

Isiku_nr (pr. key)	Perenimi	Eesnimi	Auto_regnr
1	Tamm	Lauri	678TCA
1	Tamm	Lauri	542AUN
1	Tamm	Lauri	887PPP
2	Mägi	Teet	433REP

В данном примере значения в первых 3 рядах в первых 3 столбцах одинаковые. Т.е. имя и фамилию человека можно идентифицировать по isiku_nr, Рег.номер машины по auto_nr.

Новая таблица „Isik“:

Isiku_nr (pr. key)	Perenimi	Eesnimi
1	Tamm	Lauri
2	Mägi	Teet

Новая таблица „Isiku_Auto“

Isiku_nr (fo. key)	Auto_nr (pr. key)	Auto_regnr
1	1	678TCA
1	2	542AUN
1	3	887PPP
2	4	433REP

Чтобы вся информация была в целости и сохранности, идентификатор isiku_nr необходим в обеих таблицах, чтобы показывать связь между ними.

Isiku_nr (pr. key)	Perenimi	Eesnimi
1	Tamm	Lauri
2	Mägi	Teet

1
|
*

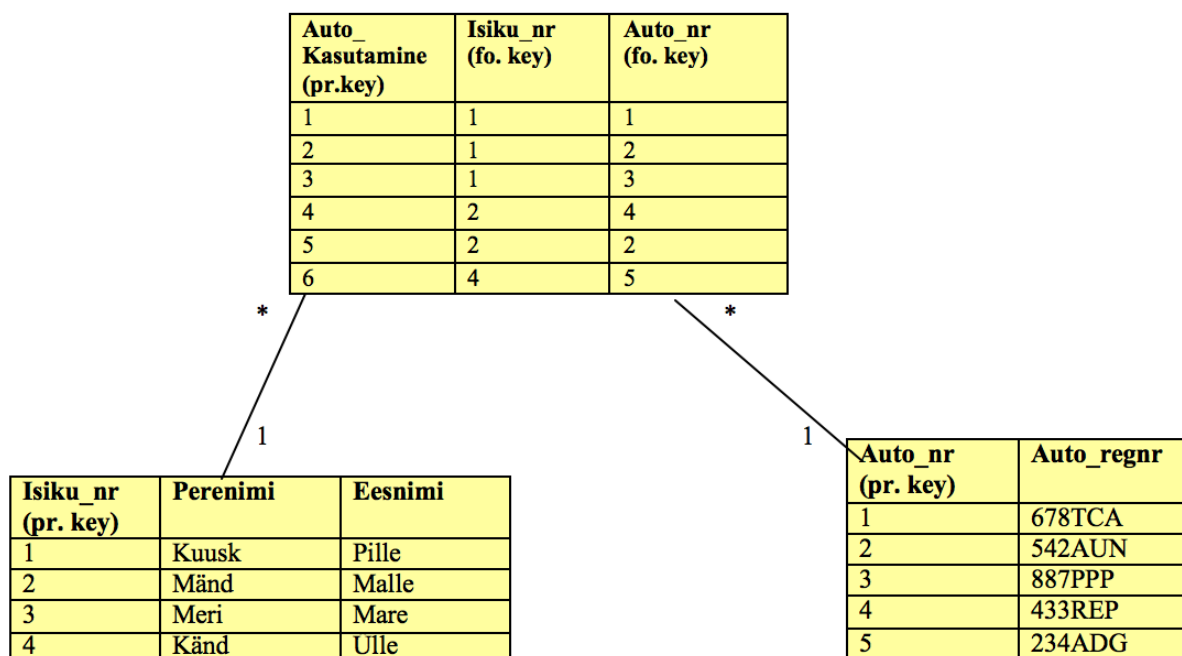
Isiku_nr (fo. key)	Auto_nr (pr. key)	Auto_regnr
1	1	678TCA
1	2	542AUN
1	3	887PPP
2	4	433REP

Вторая нормальная форма позволяет

- Уменьшить избыточность данных;
- Сохранить целостность данных (невозможно добавить новую машину человеку, кого нет в таблице „Isik“);
- Добавить/изменить, например, данные человека, не меняя данные в других таблицах (например, в таблице „Isiku_Auto“).

Третья нормальная форма

Удалите поля, которые не зависят от ключа. Если какая-то группа данных может содержаться более чем в одной записи таблицы, то стоит создать для нее отдельную таблицу.



Задание: Нормализация данных в базе данных

Дополните при необходимости модель данных и базу данных для Вашего проекта.