

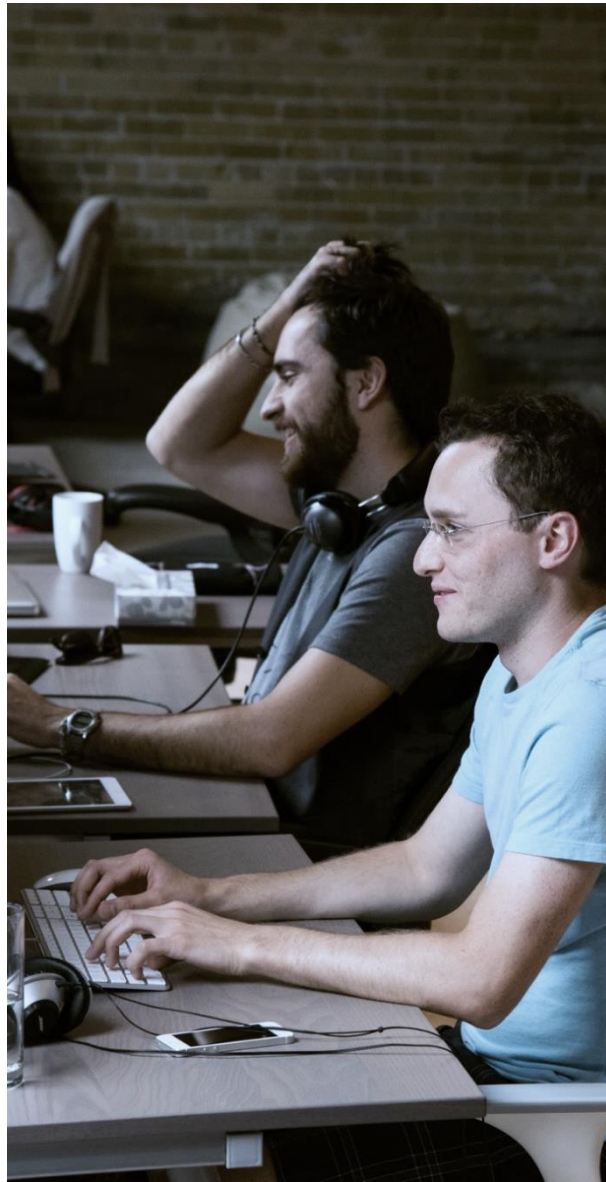


УНИВЕРСИТЕТ

искусственного интеллекта

НОРМАЛИЗАЦИЯ ДАННЫХ

1-я, 2-я и 3-я формы



ПЛАН

- Основные термины
- 1-я нормальная форма
- 2-я нормальная форма
- 3-я нормальная форма
- 3.5, 4-я, 5-я и 6-я НФ
- Плюсы и минусы нормализации





Основные термины

Атрибут — свойство некой сущности (поле таблицы)

Домен атрибута — множество допустимых значений, которые может принимать атрибут.

Кортеж — конечное множество допустимых значений атрибутов, которые вместе описывают некоторую сущность (строка таблицы).

Отношение — конечное множество кортежей (таблица).

Схема отношения — конечное множество атрибутов, определяющих некоторую сущность (структура таблицы)

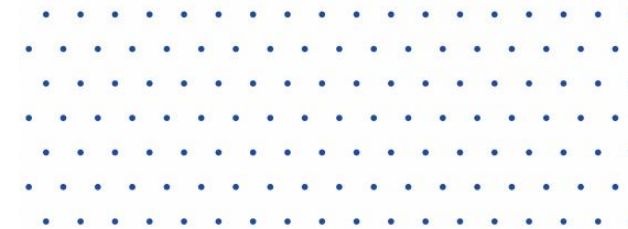
Проекция — отношение, полученное из заданного путём удаления и/или перестановки некоторых атрибутов

Функциональная зависимость между атрибутами X и Y означает, что для любого допустимого набора кортежей в данном отношении: если два кортежа совпадают по значению X , то они совпадают по значению Y .

Нормальная форма — требование, предъявляемое к структуре таблиц в теории реляционных баз данных для устранения из базы избыточных функциональных зависимостей между атрибутами (полями таблиц).

Метод нормальных форм (НФ) состоит в сборе информации о объектах решения задачи в рамках одного отношения и последующей декомпозиции этого отношения на несколько взаимосвязанных отношений на основе процедур нормализации отношений.

1-я нормальная форма

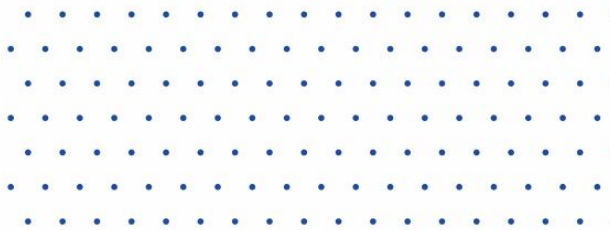


Отношение находится в 1НФ, если все его атрибуты являются простыми, все используемые домены должны содержать только скалярные значения. Не должно быть повторений строк в таблице.

Не нормализованные данные:

ID сотрудника	ФИО	Должность	Проекты
1	Илон Иванович Маск	Программист	ID: 123; Название: Система Электронных платежей PayPal; Дата сдачи: 01.01.1998 ID: 222; Название: Система автопилота в электромобиле Tesla; Дата сдачи: 01.09.2016 ID: 666; Название: Секретный проект «Костюм робота Алеши»; Дата сдачи: 12.12.2018

1-я нормальная форма



Приведенные к 1НФ данные:

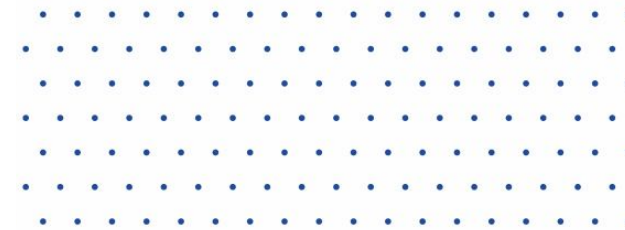
ID сотру- дника	ФИО	Должность	Код проекта	Название	Дата сдачи
1	Илон Иванович Маск	Программист	123	Система Электронных платежей PayPal	01.01.1998
1	Илон Иванович Маск	Программист	222	Система автопилота в электромобиле Tesla	01.09.2016
1	Илон Иванович Маск	Программист	666	Секретный проект «Костюм робота Алеши»	12.12.2018

1-я нормальная форма

- **Аномалия вставки.** В отношении нельзя добавить информацию о поставщике, который ещё не поставил ни одного товара.
- **Аномалия удаления.** Если от поставщика была только одна поставка, то при удалении информации о ней будет удалена и вся информация о поставщике.
- **Аномалия обновления.** Если необходимо изменить какую-либо информацию о поставщике (например, поставщик переехал в другой город), то придётся изменять значения атрибутов во всех записях о поставках от него.

Код поставщика	Город	Статус города	Код товара	Количество
1	Москва	20	1	300
1	Москва	20	2	400
1	Москва	20	3	100
2	Лондон	10	4	200
3	Инсбрук	30	5	300
3	Инсбрук	30	6	400
4	Нью-Йорк	15	7	100

2-я нормальная форма



Отношение находится во 2НФ, если оно находится в 1НФ и каждый не ключевой атрибут неприводимо зависит от Первичного Ключа(ПК).

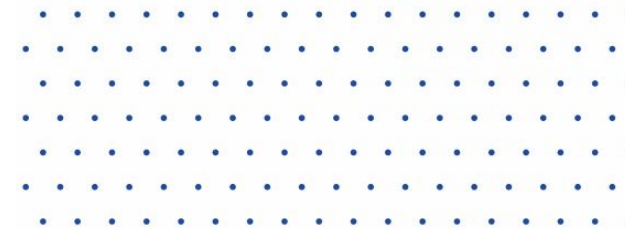
Код поставщика	Код товара	Количество
1	1	300
1	2	400
1	3	100
2	4	200
3	5	300
3	6	400
4	7	100

Код поставщика	Город	Статус города
1	Москва	20
2	Лондон	10
3	Инсбрук	30
4	Нью-Йорк	15

Разносим уникальные ключи {Код поставщика, Код товара} по двум таблицам

- {Код поставщика, Код товара} -> {Количество}
- { {Код поставщика} -> {Город},
{Код поставщика} -> {Статус},
{Город} -> {Статус} }

2-я нормальная форма



Транзитивная зависимость:

Пусть X , Y , Z - три атрибута некоторого отношения. При этом $X \twoheadrightarrow Y$ и $Y \twoheadrightarrow Z$, но обратное соответствие отсутствует, т.е. $Z \not\rightarrow Y$ и $Y \not\rightarrow X$. Тогда Z транзитивно зависит от X .

Фирма	Склад	Объем Товара
1	10	1000
2	10	20000
3	20	100
4	30	12000

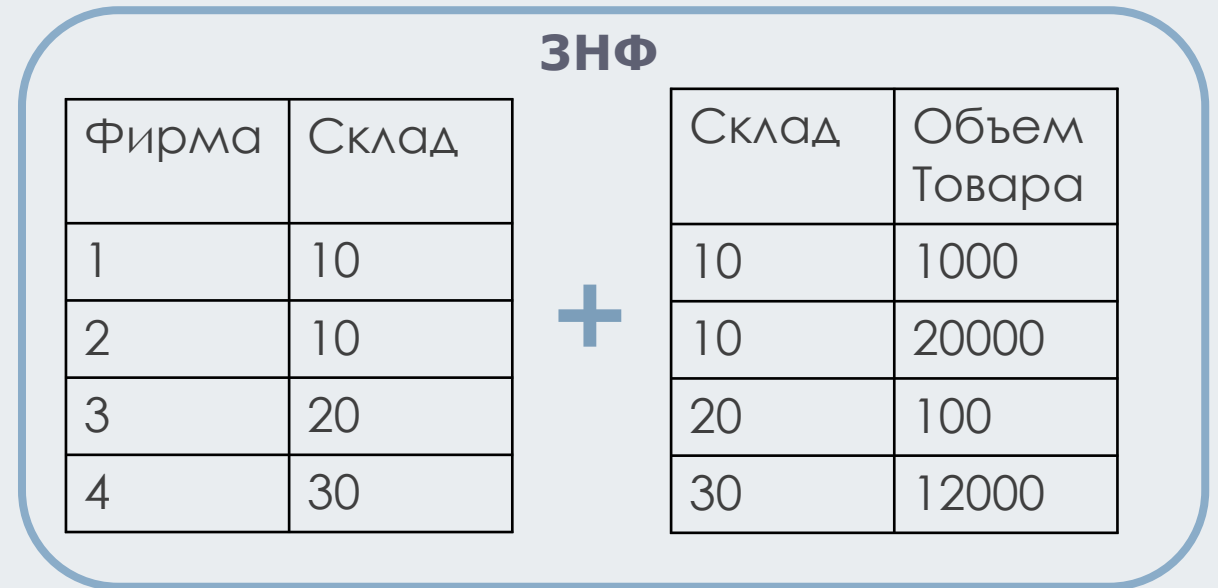
Аномалии:

- Если в данный момент ни одна фирма не получает товар со склада, то в базу данных нельзя ввести данные о его объеме (т.к. не определен ключевой атрибут)
- Если объем склада изменяется, необходим просмотр всего отношения и изменение кортежей для всех фирм, связанных с данным складом.

3-я нормальная форма

3НФ - Отношение находится в 3НФ, если оно находится во 2НФ и каждый неключевой атрибут нетранзитивно зависит от первичного ключа.

Фирма	Склад	Объем Товара
1	10	1000
2	10	20000
3	20	100
4	30	12000



3.5-я нормальная форма

3.5НФ(BCNF - нормальная форма Бойса-Кодда) - Отношение находится в BCNF, если оно находится во 3НФ и в ней отсутствуют зависимости атрибутов первичного ключа от неключевых атрибутов.

- 1) отношение имеет два или более потенциальных ключа;
- 2) два и более потенциальных ключа являются составными;
- 3) они пересекаются, т.е. имеют хотя бы один атрибут.

Номер стоянки	Время начала	Время окончания	Тариф
1	09:30	10:30	Бережливый
1	11:00	12:00	Бережливый
1	14:00	15:30	Стандарт
2	10:00	12:00	Премиум-В
2	12:00	14:00	Премиум-В
2	15:00	18:00	Премиум-А

<u>Тариф</u>	Номер стоянки	Имеет льготы
Бережливый	1	Да
Стандарт	1	Нет
Премиум-А	2	Да
Премиум-В	2	Нет

<u>Тариф</u>	<u>Время начала</u>	Время окончания
Бережливый	09:30	10:30
Бережливый	11:00	12:00
Стандарт	14:00	15:30

4/5/6-я нормальные формы

- Отношение находится в 4НФ, если оно находится в 3.5НФ и все нетривиальные многозначные зависимости фактически являются функциональными зависимостями от ее потенциальных ключей.
- Отношения находятся в 5НФ, если оно находится в 4НФ и отсутствуют сложные зависимые соединения между атрибутами.
Если «Атрибут_1» зависит от «Атрибута_2», а «Атрибут_2» в свою очередь зависит от «Атрибута_3», а «Атрибут_3» зависит от «Атрибута_1», то все три атрибута обязательно входят в один кортеж.
- Отношения находятся в шестой нормальной форме тогда и только тогда, когда она удовлетворяет всем нетривиальным зависимостям соединения. Из определения следует, что переменная находится в 6НФ тогда и только тогда, когда она неприводима, то есть не может быть подвергнута дальнейшей декомпозиции без потерь.

Плюсы и минусы нормализации



Плюсы

Снижения дублирования данных :

На каждом уровне нормализации происходит снижения дублирования данных, хранящихся в БД

Повышение качества и надежности данных:

Нормализация сокращает шанс введения некорректных данных в БД

Простота в реализации требований бизнеса:

Снижение вероятности возникновения аномалий при операциях над данными.

Увеличение скорости работы с данными

Нормализация делит данные на небольшие порции хранимые отдельно, где каждая сущность может быть оптимизирована индивидуально.

Минусы

При большом размере БД и большом количестве сущностей, пользователям приходится писать сложные запросы по объединению множества таблиц (что так же влияет на скорость выполнения).
Эффективность нормализации напрямую зависит от понимания разработчиками БД бизнеса организации и от способности бизнес пользователей четко изложить требования к БД