

Избыточное дублирование данных

Избыточность данных в базе данных относится к нежелательным явлениям, поскольку ведет к увеличению объема памяти, необходимого для физического хранения отношений и времени обработки. Избыточность вызывается, прежде всего, дублированием данных.

Избыточное дублирование данных – это повторение данных в базе, вызванное наличием зависимостей между ними. При избыточном дублировании повторяются не просто данные, а информация о предметной области.

Пример отношения, содержащего нежелательную избыточность

Отношение Студент

Номер зачетной книжки	ФИО студента	Группа	ФИО старосты	Куратор
210-22	Иванов С.И.	Т-11	Рябов В.С.	Сергеева А.Ю.
211-22	Петров Я.С.	Т-12	Сизов О.Г.	Докина П.Р.
212-22	Рябов В.С.	Т-11	Рябов В.С.	Сергеева А.Ю.
213-22	Семина О.Л.	Т-11	Рябов В.С.	Сергеева А.Ю.

Данные о ФИО старосты и кураторе в данном случае являются избыточными

Различают простое (не избыточное) и избыточное дублирование данных. Первое из них вполне естественно, второе может приводить к проблемам при обработке данных.

Пример не избыточного дублирования данных представляет отношение Сотрудник – Телефон с такими же атрибутами.

Отношение Сотрудник - телефон

Сотрудник	Телефон
Иванов	4159
Петров	4121
Сидоров	4121
Егоров	4121

Для сотрудников, находящихся в одном помещении, номера телефонов совпадают. Номер телефона 412 встречается несколько раз, хотя для каждого служащего номер телефона уникален. Поэтому ни один из номеров не является избыточным. Действительно, при удалении одного из номеров телефонов будет утеряна информация о том, по какому номеру можно дозвониться до одного из служащих.

Пример избыточного дублирования представляет отношение Сотрудник – Телефон – Комната.

Это отношение Сотрудник – Телефон, дополненное номером комнаты. Естественно предположить, что все сотрудники в одной комнате имеют один и тот же телефон, следовательно, в рассматриваемом отношении имеется избыточное дублирование данных. Так, в связи с тем, что Сидоров и Егоров находятся в той же комнате, что и Петров, то их номер можно узнать из кортежа со сведениями о Петрове.

Отношение Сотрудник – Телефон – Комната

Сотрудник	Телефон	Номер комнаты
Иванов	4159	115
Петров	4121	111
Сидоров	4121	111
Егоров	4121	111

Возможный способ выхода из данной ситуации: два отношения Сотрудник – Комната и Комната – Телефон. Первое из них содержит информацию о номерах комнат, в которых располагаются сотрудники, а второе – информацию о номерах телефонах в каждой комнате. Увольнение одного из сотрудников и, следовательно, удаление такого сотрудника из базы данных не приведет к утере информации о номере телефона в 111 комнате.

Отношение Сотрудник - Комната

Сотрудник	Номер комнаты
Иванов	115
Петров	111
Сидоров	111
Егоров	111

Отношение Комната - Телефон

Номер комнаты	Номер комнаты
115	4159
111	4121

Избыточное дублирование данных создает проблемы при обработке кортежей отношения, названные Эдгаром Коддом аномалиями обновления отношений.

Аномалиями называют такую ситуацию в таблицах базы данных, которая приводит к противоречиям в базе данных либо существенно усложняет обработку данных. Аномалия обновления заключается в том, что для внесения изменения в данные об одном факте (событии, объекте) необходимо выполнить множественные изменения в базе данных.

Проблемы, возникающие при избыточном дублировании, устраняются путем нормализации таблиц.