

Целостность данных (Data integrity)

Разделы: [Бизнес-задачи](#)

В теории баз данных под целостностью понимают свойство соответствия структуры и содержания базы данных предметной области. В реляционной модели данных определяются **два основных требования**, при которых обеспечивается целостность данных: **целостность сущностей** и **целостность ссылок**.

Каждый объект или наблюдение представляется в реляционной базе как группа взаимосвязанных элементов данных — кортеж некоторого отношения. Требование целостности сущностей заключается в том, чтобы каждый кортеж любого отношения отличался от другого кортежа этого отношения (т.е. любое отношение должно обладать первичным ключом).

Очевидно, что если данное требование будет нарушено (т.е. кортежи в рамках одного отношения не уникальны), то в базе будет храниться противоречивая информация об одном и том же объекте.

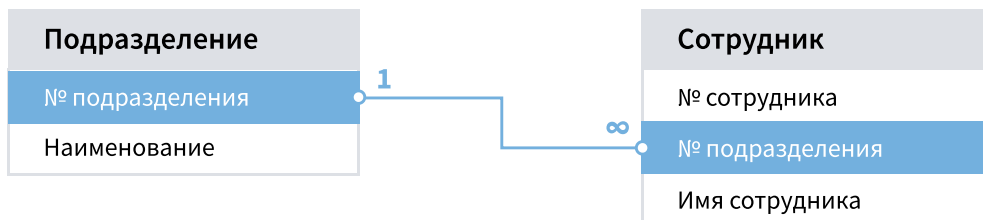
Поддержание целостности сущностей обеспечивается средствами системы управления базой данных (СУБД) с помощью двух ограничений:

- при добавлении записей в таблицу проверяется уникальность их первичных ключей;
- запрещено изменять значения атрибутов, входящих в первичный ключ.

Требование ссылочной целостности состоит в том, что для каждого значения внешнего ключа, появляющегося в дочернем отношении, в родительском должен найтись кортеж с таким же значением первичного ключа.

Например, даны отношения «ПОДРАЗДЕЛЕНИЕ» («N ПОДРАЗДЕЛЕНИЯ», «НАИМЕНОВАНИЕ ПОДРАЗДЕЛЕНИЯ») и «СОТРУДНИК» («N СОТРУДНИКА», «N ПОДРАЗДЕЛЕНИЯ», «ИМЯ СОТРУДНИКА»), в которых хранятся сведения о работниках организации и подразделениях, где они работают. Отношение «ПОДРАЗДЕЛЕНИЕ» в данной паре является **родительским**, поэтому его первичный ключ «N ПОДРАЗДЕЛЕНИЯ» присутствует в дочернем отношении «СОТРУДНИК».

Требование ссылочной целостности означает в данном случае, что в таблице «СОТРУДНИК» не может присутствовать кортеж со значением атрибута «N ПОДРАЗДЕЛЕНИЯ», которое не встречается в таблице «ПОДРАЗДЕЛЕНИЕ». Если такое значение в отношении «ПОДРАЗДЕЛЕНИЕ» отсутствует, значение внешнего ключа в отношении «СОТРУДНИК» считается **неопределенным**.



Как правило, поддержание ссылочной целостности также является функцией СУБД. Например, она может запретить пользователю добавлять запись, содержащую внешний ключ с несуществующим (неопределенным) значением.

Целостность данных является важнейшим фактором их качества. Нарушение целостности не позволит производить их корректную интеграцию в хранилище данных и осуществлять их дальнейший анализ. Поэтому при реализации аналитических проектов необходимо использовать механизмы автоматического контроля и поддержания целостности данных.

Решение Loginom Data Quality обеспечивает качество клиентских данных и создает единое представление о клиентской базе компании. Статья «Очистка данных. Кейс коллекторского агентства ЭОС» повествует об эффективности очистки и стандартизации данных для ключевых бизнес-процессов.