Лекция 3

Как узнать то, что нужно: SQL

Мощным преимуществом реляционной модели данных является наличие строгого математического аппарата. Этот аппарат – реляционная алгебра. Алгебра устроена таким образом, что все манипуляции с реляционной базой выполняются с помощью одного оператора – SELECT.

SELECT [ALL/DISTINCT] {\*/ [имя столбца] [AS новое имя]} [1...n] FROM имя таблицы [[AS] псевдоним] [1...n] [WHERE <условие поиска>] [GROUP BY имя столбца [1...n]] [HAVING <критерии выбора группы>] [ORDER BY имя столбца [1...n]].

Обработка операции SELECT выполняется в последовательности:

1. FROM – определяет имена исходных таблиц.
2. WHERE – фильтр строк объекта в соответствии с указанным условием.
3. GROUP BY – группирует строки, имеющие одно и тоже значение в указанном столбце.
4. HAVING – фильтрация группы с указанным условием.
5. SELECT – выбрать какие столбцы в выходных данных.
6. ORDER BY – упорядочивание результата.

Структура:

1. SELECT – порядок предложений не может быть изменен.
2. SELECT и FROM – обязательные, остальные – нет.
3. SELECT – закрытая операция в смысле замкнутого множества (результат представляет собой отдельную таблицу).

Отдельную проблему составляет ситуация, когда нужно делать выборку из нескольких таблиц. Для этого в предложение FROM дописывается JOIN:

FROM [имя таблицы 1] {INNER, LEFT, RIGHT} JOIN [имя таблицы 2] ON <условие присоединения>

Конкатенация – базовая операция соединения (декартово произведение).

|  |  |
| --- | --- |
| Отношение S | |
| S.b1 | S.b2 |
| 1 | H |
| 2 | G |
| 3 | H |

|  |  |  |  |
| --- | --- | --- | --- |
| Отношение RxS | | | |
| R.a1 | R.a2 | S.b1 | S.b2 |
| A | 1 | 1 | H |
| A | 1 | 2 | G |
| A | 1 | 3 | H |
| A | 2 | 1 | H |
| A | 2 | 2 | G |
| A | 2 | 3 | H |
| … | … | … | … |

|  |  |
| --- | --- |
| Отношение R | |
| R.a1 | R.a2 |
| A | 1 |
| A | 2 |
| B | 1 |
| B | 3 |
| B | 4 |

В результирующей таблице количество строк равно произведению количества строк первой таблицы на количество строк во второй таблице. Как правило, декартово произведение формирует большую и избыточную таблицу, поэтому обязательно указываются условия соединения:

SELECT R.a1, R.a2, S.b1, S.b2

FROM R

INNER JOIN S

ON R.a2=S.b1

|  |  |  |  |
| --- | --- | --- | --- |
| RxS, R.a2=S.b1 | | | |
| R.a1 | R.a2 | S.b1 | S.b2 |
| A | 1 | 1 | H |
| A | 2 | 2 | G |
| B | 3 | 3 | H |
| B | 1 | 1 | H |

Естественное соединение – соединение по эквивалентности двух отношений R и S, выполненное по всем общим атрибутам, из результатов которого исключаются по одному экземпляру каждого общего атрибута. Для естественного соединения нет зарезервированного слова, его надо задавать самостоятельно.

Основная терминология

Атрибут – именованный столбец отношений.

Домен – набор допустимых значений одного или нескольких атрибутов. Каждый атрибут реализуется в базе на некотором домене.

|  |  |  |  |
| --- | --- | --- | --- |
| Атрибут | Имя Домена | Содержание домена | Определение Домена |
| Brunch № | Номер отдела | Множество всех допустимых номеров отдела | Символьный размер – 4  Диапазон от Brunch 111 до 999 |
| Street | StreetName | Множество всех допустимых названий улиц | Символьный размер - 30 |
| DateOfBirth | Дата рождения | Множество всех допустимых дат | Диапазон дат в разных форматах |

В результате введения понятия система получает больше информации, что позволяет избежать семантически некорректных операций.

Кортеж – строка отношений, строки могут идти в любом порядке.

Заголовок – описание структуры отношений вместе со спецификацией домена и любыми другими ограничивающими возможности атрибутами.

Степень отношения – количество атрибутов, которое содержится в отношении. Определение входит в определение заголовка.

Отношение с одним атрибутом называется унарным отношением или одноименным кортежем.

Кардинальность – количество кортежей, которые содержатся в отношении.