

Использование технологии OLAP для аналитических исследований процесса обучения

Житкевич А.П., Пашенко Е.А.

Белорусский национальный технический университет

С середины прошлого десятилетия в мире стремительно растет интерес компаний к программным продуктам, позволяющим руководителям и аналитикам работать с большими объемами информации, накопленными в учетных системах и хранилищах данных, и извлекать из них полезные сведения. И сегодня можно говорить о появлении целого направления в области программного обеспечения - Business Intelligence (BI).

Любая учетная система, какой бы универсальной она ни была, не в состоянии покрыть всех потребностей обработки и анализа хранимой информации. Для ведения успешной деятельности необходимо иметь актуальные сведения о текущем состоянии дел, готовить материалы для дальнейших действий. Информационная составляющая превращается в важнейший элемент управления любым направлением деятельности, поэтому способность предприятий обеспечивать своих руководителей всеми необходимым для принятия взвешенных решений сведениями становится преимуществом любой организации.

При создании запросов на поиск информации по заданным критериям возникает проблема быстрого действия. Методы, используемые в системах оперативной обработки транзакций, осложняют создание даже простых отчетов, так как данные чаще всего распределены по множеству таблиц, и для их агрегирования необходимо выполнять сложные операции объединения. Как правило, попытки создания комплексных отчетов требуют больших вычислительных мощностей и приводят к потере производительности.

Кроме того, некоторые виды анализа требуют таких структурных изменений, которые недопустимы в текущей оперативной среде. Например, нужно выяснить, что произойдет, если у компании появятся новые продукты. На реальной базе такое исследование провести нельзя. Следовательно,

эффективный анализ редко удается выполнить непосредственно в учетной системе.

На рисунке 1 представлена схема базы данных, содержащая информацию из учетных карточек студентов. Данные в этой базе будут удовлетворять третьей нормальной форме, однако это повлечет за собой некоторые затруднения при просмотре и анализе содержимого.

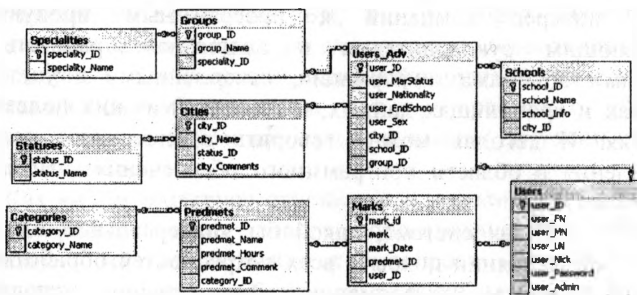


Рисунок 1. Схема нормализованной базы данных

Беглого взгляда на структуру достаточно, чтобы понять сложность извлечения информации даже из базы данных. Аналитические запросы потребуют объединения большого количества таблиц, что существенно увеличивает время выполнения запроса.

Для проведения аналитических исследований исходная база данных должна сначала преобразовываться в хранилище данных (ХД). В основе концепции ХД лежит идея разделения данных, используемых для оперативной обработки и для решения задач анализа. Такое разделение позволяет оптимизировать как структуры данных оперативного хранения для выполнения операций ввода, модификации, удаления и поиска, так и структуры данных, используемые для анализа. В разработанной системе данные, необходимые для определенных задач анализа и хранящиеся в OLTP-системе, сводятся в единую таблицу, представляющую хранилище данных.

Затем к информации, хранящейся в хранилище данных, должны быть применены технологии OLAP. Применение этой технологии обусловлено недостатками реляционной модели базы данных, в частности, невозможностью объединять,

просматривать и анализировать данные с точки зрения множественности измерений. В процессе анализа данных, для аналитика важно построить и изучить зависимости между различными параметрами. Кроме того, число таких параметров может варьироваться в широких пределах.

Для приведения данных к удобной аналитику форме необходимо выделение одного из атрибутов, на основании которого и будет производиться анализ. После чего из реляционного (двухмерного) представления данных, на основании выделенного параметра, формируется, трехмерная структура, называемая куб решений. В общем случае эта структура может представлять собой гиперкуб. Операции над результирующим гиперкубом позволяют агрегировать данные по определенным аналитическим критериям.

Для примера использования OLAP для анализа данных процесса обучения была поставлена следующая задача: исследовать зависимость успеваемости от пола студента. Полученные результаты представлены на рисунке 2.

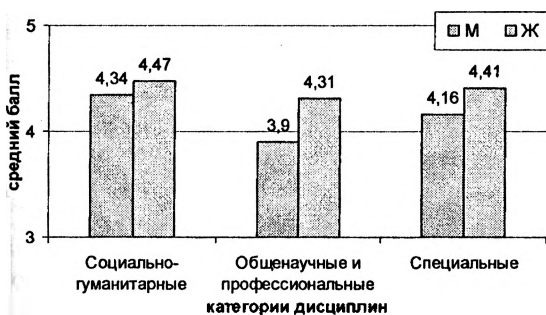


Рисунок 2. Зависимость успеваемости от пола.

По результатам анализа можно сделать вывод о том, что разрыв в успеваемости больше по общенаучным и профессиональным дисциплинам. При этом успеваемость представителей женского пола выше по всем дисциплинам. Также определено, что дисперсия отличается существенно и больше в случае представителей мужского пола. Это свидетельствует о большей неравномерности успеваемости.