

ТЕХНОЛОГИИ РАБОТЫ С ХРАНИЛИЩАМИ ДАННЫХ

В настоящее время во многих организациях проблема использования большого объема собранных за многие годы данных ощущается все острее. В течение многих лет на предприятиях накапливаются и хранятся огромнейшие массивы информации, поэтому и возникает потребность в системах, позволяющих анализировать хранящиеся данные.

Хранилище данных (Data Warehouse) - это технология, которая объединяет структурированные данные из одного или нескольких источников, чтобы их можно было анализировать для повышения эффективности бизнес-аналитики.

Основные требования, сформулированные Ральфом Кимбалл, к хранилищам данных:

1. Поддержка высокой скорости получения данных из хранилища.
2. Поддержка внутренней непротиворечивости данных.
3. Возможность получения и сравнения так называемых срезов данных.
4. Наличие удобных утилит просмотра данных в хранилище.
5. Полнота и достоверность хранимых данных.
6. Поддержка качественного процесса пополнения данных.

Типичная структура хранилища данных существенно отличается от структуры обычной реляционной СУБД:

- Как правило, хранилища данных денормализованы, это позволяет повысить скорость выполнения запросов.
- Обычные базы данных подвержены постоянным изменениям в процессе работы пользователей, а хранилище данных относительно стабильно.
- Обычные базы данных предназначены для того, чтобы помочь пользователям выполнять повседневную работу, тогда как хранилища данных предназначены для принятия решений.

Геврасёва Ирина Петровна, Рубель Полина Александровна, Бобко Екатерина Юрьевна, Бу Акль Глория Джамал, студенты третьего курса факультета информационных технологий и управления БГУИР, irina_gevraseva@mail.ru

Научный руководитель: Трофимович Алексей Федорович, старший преподаватель кафедры информационных технологий автоматизированных систем БГУИР, trofimovich_a_f@tut.by

- Обычные базы данных чаще всего являются источником данных, попадающих в хранилище. Кроме того, хранилище может пополняться за счет внешних источников, например статистических отчетов.

Хранилища данных состоят из таблицы фактов и таблиц измерений. Таблица фактов состоит из данных, используемых для анализа информации и ключей - связей с таблицами измерений. Таблицы измерений в свою очередь содержат атрибуты событий, сохраненных в таблице фактов. Хранилища данных могут быть реализованы в виде схемы "звезда" или "снежинка".

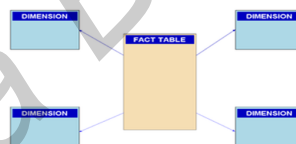


Рис. 1 – Схема хранилища данных типа звезда

Рассмотрим строение схемы "звезда". Она содержит одну таблицу фактов (fact table), несколько денормализованных таблиц измерений (dimension). Таблица фактов и таблицы измерений связаны идентифицирующими связями. Агрегированные данные хранятся совместно с исходными. Схема "снежинка" имеет подобное строение, отличительной чертой является то, что агрегированные данные могут храниться отдельно от исходных. Наиболее оптимальной и часто используемой является схема "звезда".

Список литературы

1. Building the Data Warehouse, 4th Edition / W. H. Inmon // Издательство: Wiley Publishing, Inc, 2005. – 576 с.
2. ВІPartner [Электронный ресурс] / Режим доступа: http://www.bipartner.ru/resources/dw_arch.html