

image not found or type unknown



OLAP (англ. online analytical processing, интерактивная аналитическая обработка) — технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу.

Причина использования OLAP для обработки запросов — скорость. Реляционные базы данных хранят сущности в отдельных таблицах, которые обычно хорошо нормализованы. Эта структура удобна для операционных баз данных (системы OLTP), но сложные многотабличные запросы в ней выполняются относительно медленно.

OLAP-структура, созданная из рабочих данных, называется OLAP-куб. Куб создаётся из соединения таблиц с применением схемы звезды или схемы снежинки. В центре схемы звезды находится таблица фактов, которая содержит ключевые факты, по которым делаются запросы. Множественные таблицы с измерениями присоединены к таблице фактов. Эти таблицы показывают, как могут анализироваться агрегированные[en] реляционные данные. Количество возможных агрегирований определяется количеством способов, которыми первоначальные данные могут быть иерархически отображены.

Например, все клиенты могут быть сгруппированы по городам или регионам страны (Запад, Восток, Север и так далее), таким образом, 50 городов, восемь регионов и две страны составят три уровня иерархии с 60-ю членами. Также клиенты могут быть объединены по отношению к продукции; если существуют 250 продуктов по 20 категориям, три группы продукции и три производственных подразделения, то количество агрегатов составит 16 560. При добавлении измерений в схему количество возможных вариантов быстро достигает десятков миллионов и более.

OLAP-куб содержит базовые данные и информацию об измерениях (агрегаты). Куб потенциально содержит всю информацию, которая может потребоваться для ответов на любые запросы. При огромном количестве агрегатов зачастую полный расчёт происходит только для некоторых измерений, для остальных же производится «по требованию».

Существуют три типа OLAP

многомерная OLAP (Multidimensional OLAP — MOLAP);

реляционная OLAP (Relational OLAP — ROLAP);

гибридная OLAP (Hybrid OLAP — HOLAP).

MOLAP — классическая форма OLAP, так что её часто называют просто OLAP. Она использует суммирующую базу данных и создаёт требуемую многомерную схему данных с сохранением как базовых данных, так и агрегатов.

ROLAP работает напрямую с реляционной базой данных, факты и таблицы с измерениями хранятся в реляционных таблицах, и для хранения агрегатов создаются дополнительные реляционные таблицы.

HOLAP использует реляционные таблицы для хранения базовых данных и многомерные таблицы для агрегатов.

Особым случаем ROLAP является «ROLAP реального времени» (Real-time ROLAP — R-ROLAP). В отличие от ROLAP в R-ROLAP для хранения агрегатов не создаются дополнительные реляционные таблицы, а агрегаты рассчитываются в момент запроса. При этом многомерный запрос к OLAP-системе автоматически преобразуется в SQL-запрос к реляционным данным.

Каждый тип хранения имеет определённые преимущества, хотя есть разногласия в их оценке у разных производителей. MOLAP лучше всего подходит для небольших наборов данных, он быстро рассчитывает агрегаты и возвращает ответы, но при этом генерируются огромные объёмы данных. ROLAP считается более масштабируемым решением, притом более экономичным к пространству хранения, но с ограничениями по возможностям аналитической обработки. HOLAP находится посреди этих двух подходов, он достаточно хорошо масштабируется, и позволяет преодолеть ряд ограничений. Архитектура R-ROLAP позволяет производить многомерный анализ OLTP-данных в режиме реального времени.

Сложность в применении OLAP состоит в создании запросов, выборе базовых данных и разработке схемы, в результате чего большинство продуктов OLAP поставляются вместе с огромным количеством предварительно настроенных запросов. Другая проблема — в базовых данных, они должны быть полными и непротиворечивыми.

OLAP – удобный инструмент анализа

Централизация и удобное структурирование – это далеко не все, что нужно аналитику. Ему ведь еще требуется инструмент для просмотра, визуализации информации. Традиционные отчеты, даже построенные на основе единого хранилища, лишены одного – гибкости. Их нельзя “покрутить”, “развернуть” или “свернуть”, чтобы получить желаемое представление данных. Конечно, можно вызвать программиста (если он захочет придти), и он (если не занят) сделает новый отчет достаточно быстро – скажем, в течение часа (пишу и сам не верю – так быстро в жизни не бывает; давайте дадим ему часа три). Получается, что аналитик может проверить за день не более двух идей. А ему (если он хороший аналитик) таких идей может приходить в голову по нескольку в час. И чем больше “срезов” и “разрезов” данных аналитик видит, тем больше у него идей, которые, в свою очередь, для проверки требуют все новых и новых “срезов”. Вот бы ему такой инструмент, который позволил бы разворачивать и сворачивать данные просто и удобно! В качестве такого инструмента и выступает OLAP.

Хотя OLAP и не представляет собой необходимый атрибут хранилища данных, он все чаще и чаще применяется для анализа накопленных в этом хранилище сведений.

Оперативные данные собираются из различных источников, очищаются, интегрируются и складываются в реляционное хранилище. При этом они уже доступны для анализа при помощи различных средств построения отчетов. Затем данные (полностью или частично) подготавливаются для OLAP-анализа. Они могут быть загружены в специальную БД OLAP или оставлены в реляционном хранилище. Важнейшим его элементом являются метаданные, т. е. информация о структуре, размещении и трансформации данных. Благодаря им обеспечивается эффективное взаимодействие различных компонентов хранилища.

Подытоживая, можно определить OLAP как совокупность средств многомерного анализа данных, накопленных в хранилище. Теоретически средства OLAP можно применять и непосредственно к оперативным данным или их точным копиям (чтобы не мешать оперативным пользователям). Но мы тем самым рискуем наступить на уже описанные выше грабли, т. е. начать анализировать оперативные данные, которые напрямую для анализа непригодны.