



OLAP-системы

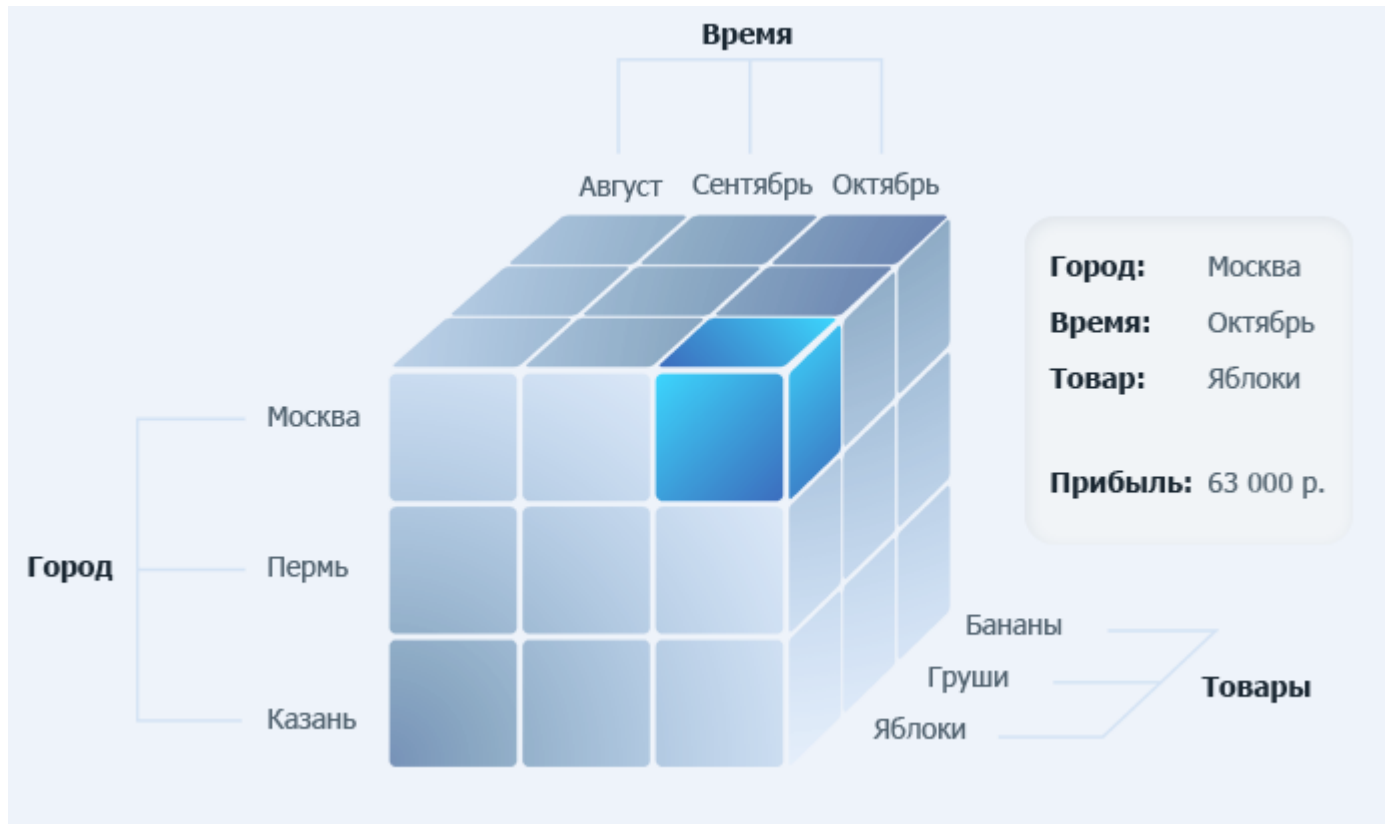
OLAP-приложения(On-Line Analytical Processing(OLAP) -оперативная аналитическая обработка данных). Это обобщенный термин, характеризующий принципы построения систем поддержки принятия решений(Decision Support System-DSS),хранилищ данных(Data Warehouse),систем интеллектуального анализа данных(Data Mining). Такие системы предназначены для нахождения зависимостей между данными (например, можно попытаться определить, как связан объем продаж товаров с характеристиками потенциальных покупателей), для проведения анализа "что если...". OLAP-приложения оперируют с большими массивами данных, уже накопленными в OLTP-приложениях, взятыми их электронных таблиц или из других источников данных.

Такие системы характеризуются следующими признаками:

- Добавление в систему новых данных происходит относительно редко крупными блоками (например, раз в квартал загружаются данные по итогам квартальных продаж из OLTP-приложения).
- Данные, добавленные в систему, обычно никогда не удаляются.
- Перед загрузкой данные проходят различные процедуры "очистки", связанные с тем, что в одну систему могут поступать данные из многих источников, имеющих различные форматы представления для одних и тех же понятий, данные могут быть некорректны, ошибочны.
- Запросы к системе являются нерегламентированными и, как правило, достаточно сложными.
- Скорость выполнения запросов важна, но не критична.

Данные OLAP-приложений обычно представлены в виде одного или нескольких гиперкубов, измерения которого представляют собой справочные данные, а в ячейках самого гиперкуба хранятся собственно данные. Например, можно построить гиперкуб, измерениями которого являются: время (в кварталах, годах), тип товара и отделения компании, а в ячейках хранятся объемы продаж. Такой гиперкуб будет содержать данных о продажах различных типов товаров по кварталам и подразделениям. Основываясь на этих данных, можно отвечать на вопросы вроде "у какого подразделения самые лучшие объемы продаж в текущем

году?", или "каковы тенденции продаж отделений Юго-Западного региона в текущем году по сравнению с предыдущим годом?" и т.п.



OLAP-системы можно разбить на три класса

- 1 класс. Наиболее сложными и дорогими из них являются основанные на патентованных технологиях серверы многомерных БД. Эти системы обеспечивают полный цикл OLAP-обработки и либо включают в себя, помимо серверного компонента, собственный интегрированный клиентский интерфейс, либо используют для анализа данных внешние программы работы с электронными таблицами. Продукты этого класса в наибольшей степени соответствуют условиям применения в рамках крупных информационных хранилищ. Для их обслуживания требуется целый штат сотрудников, занимающихся как установкой и сопровождением системы, так и формированием представлений данных для конечных пользователей. Обычно подобные пакеты довольно дороги. В качестве примеров продуктов этого класса можно привести систему Essbase корпорации Arbor Software, Express фирмы IRI (входящей теперь в состав Oracle), Lightship производства компании Pilot Software и др.
- 2 класс OLAP-систем — реляционные OLAP-системы (ROLAP). Здесь для хранения данных используются старые реляционные СУБД, а между БД и

клиентским интерфейсом организуется определяемый администратором системы слой метаданных. Через этот промежуточный слой клиентский компонент может взаимодействовать с реляционной БД как с многомерной. Подобно средствам первого класса, ROLAP-системы хорошо приспособлены для работы с крупными информационными хранилищами, требуют значительных затрат на обслуживание специалистами информационных подразделений и предусматривают работу в многопользовательском режиме. Среди продуктов этого типа — IQ/Vision корпорации IQ Software, DSS/Server и DSS/Agent фирмы MicroStrategy и DecisionSuite компании Information Advantage.

- 3 класс OLAP-систем — инструменты генерации запросов и отчетов для настольных ПК, дополненные OLAP-функциями или интегрированные с внешними средствами, выполняющими такие функции. Эти весьма развитые системы осуществляют выборку данных из исходных источников, преобразуют их и помещают в динамическую многомерную БД, функционирующую на ПК конечного пользователя. Указанный подход, позволяющий обойтись как без дорогостоящего сервера многомерной БД, так и без сложного промежуточного слоя метаданных, необходимого для ROLAP-средств, обеспечивает в то же время достаточную эффективность анализа. Эти средства для настольных ПК лучше всего подходят для работы с небольшими, просто организованными БД. Потребность в квалифицированном обслуживании для них ниже, чем для других OLAP-систем, и примерно соответствует уровню обычных сред обработки запросов. В числе основных участников этого сектора рынка — компания Brio Technology со своей системой Brio Query Enterprise, Business Objects с одноименным продуктом и Cognos с PowerPlay.

OLTP-системы

- **OLTP-системы** предназначены для ввода, структурированного хранения и обработки информации (операций, документов) в режиме реального времени.
- OLTP-системы, являясь высокоэффективным средством реализации оперативной обработки, оказались малопригодны для задач аналитической обработки.

Типичными примерами OLTP-приложений являются системы складского учета, системы заказов билетов, банковские системы, выполняющие операции по переводу денег, и т.п. Основная функция подобных систем заключается в выполнении большого количества коротких транзакций. Сами транзакции выглядят относительно просто, например, "снять сумму денег со счета А, добавить эту сумму на счет В". Проблема заключается в том, что, во-первых, транзакций очень много,

во-вторых, выполняются они одновременно (к системе может быть подключено несколько тысяч одновременно работающих пользователей), в-третьих, при возникновении ошибки, транзакция должна целиком откатиться и вернуть систему к состоянию, которое было до начала транзакции (не должно быть ситуации, когда деньги сняты со счета А, но не поступили на счет В). Практически все запросы к базе данных в OLTP-приложениях состоят из команд вставки, обновления, удаления. Таким образом, критическим для OLTP-приложений является скорость и надежность выполнения коротких операций обновления данных. Чем выше уровень нормализации данных в OLTP-приложении, тем оно, как правило, быстрее и надежнее.

Характеристики

1. Средствами традиционных OLTP-систем можно построить аналитический отчет и даже прогноз любой сложности, но заранее регламентированный. Любой шаг в сторону, любое нерегламентированное требование конечного пользователя, как правило, требует знаний о структуре данных и достаточно высокой квалификации программиста;
2. Многие необходимые для оперативных систем функциональные возможности являются избыточными для аналитических задач и в то же время могут не отражать предметной области. Для решения большинства аналитических задач требуется использование внешних специализированных инструментальных средств для анализа, прогнозирования и моделирования. Жесткая же структура баз не позволяет достичь приемлемой производительности в случае сложных выборок и сортировок и, следовательно, требует больших временных затрат для организации шлюзов.
3. В отличие от транзакционных, в аналитических системах не требуются и, соответственно, не предусматриваются развитые средства обеспечения целостности данных, их резервирования и восстановления. Это позволяет не только упростить сами средства реализации, но и снизить внутренние накладные расходы и, следовательно, повысить производительность при выборке данных.

Задачи, решаемые OLTP- и OLAP-системами

Image not found or type unknown



Сравнение OLTP и OLAP

Характеристика	OLTP	OLAP
Преобладающие операции	Ввод данных, поиск	Анализ данных
Характер запросов	Много простых транзакций	Сложные транзакции
Хранимые данные	Оперативные, детализированные	охватывающие большой период времени, агрегированные
Вид деятельности	Оперативная, тактическая	Аналитическая, стратегическая
Тип данных	Структурированные	Разнотипные

Характеристика базы данных	База данных OLTP (оперативная обработка транзакций)	База данных OLAP (хранилище данных, деловой анализ)
Содержимое	Текущие данные	Данные, накопленные за долгий период времени
Структура данных	Структура таблиц соответствует структуре транзакций	Структура таблиц понятна и удобна для написания запросов (кубы фактов — схема "звезда")
Типичный размер таблиц	Тысячи строк	Миллионы строк
Схема доступа	Предопределена для каждого типа обрабатываемых транзакций	Произвольная; зависит от того, какая именно задача стоит перед пользователем в данный момент и какие сведения нужны для ее решения
Количество строк, к которым обращается один запрос	Десятки	От тысяч до миллионов
С какими данными работает приложение	С отдельными строками	С группами строк (итоговые запросы)
Интенсивность обращений к базе данных	Большое количество бизнес - транзакций в минуту или в секунду	На выполнение запросов требуется время: минуты или даже часы
Тип доступа	Выборка, вставка и обновление	Выборка данных (почти 100 % операций)
Чем определяется производительность	Время выполнения транзакции	Время выполнения запроса

Основные выводы

В области ИТ управления существуют два взаимно дополняющих друг друга направления:

- технологии, ориентированные на оперативную (транзакционную) обработку данных — OLTP (online transaction processing) системы;
- технологии, ориентированные на анализ данных и принятие решений — OLAP (online analytical processing) системы.

Основное назначение OLAP-систем — динамический многомерный анализ исторических и текущих данных, стабильных во времени, анализ тенденций,

моделирование и прогнозирование будущего.

- 2 класс. Реляционные OLAP-системы (ROLAP). Здесь для хранения данных используются старые реляционные СУБД, а между БД и клиентским интерфейсом организуется определяемый администратором системы слой метаданных. Через этот промежуточный слой клиентский компонент может взаимодействовать с реляционной БД как с многомерной.
- 3 класс. Инструменты генерации запросов и отчетов для настольных ПК, дополненные OLAP-функциями или интегрированные с внешними средствами, выполняющими такие функции. Эти системы осуществляют выборку данных из исходных источников, преобразуют их и помещают в динамическую многомерную БД, функционирующую на ПК конечного пользователя.

OLTP-системы, являясь высокоэффективным средством реализации оперативной обработки, оказались малопригодны для задач аналитической обработки. Data Warehousing — «хранилища (склады) данных». Под этим понимается набор организационных решений, программных и аппаратных средств для обеспечения аналитиков информацией на основе данных из систем обработки транзакций нижнего уровня и других источников.