

image not found or type unknown



## **Введение.**

### 1. ОСНОВА КОНЦЕПЦИИ OLAP

OLAP — представляет собой инструмент для анализа больших объёмов данных. Взаимодействуя с OLAP-системой, пользователь сможет осуществлять гибкий просмотр информации, получать произвольные срезы данных и выполнять аналитические операции детализации, свёртки, сквозного распределения, сравнения во времени. Вся работа с OLAP-системой происходит в терминах предметной области.

OLAP-системы являются часть более общего понятия Business Intelligence, которое включает в себя помимо традиционного OLAP-сервиса средства организации совместного использования документов, возникающих в процессе работы пользователей хранилища. Технология Business Intelligence обеспечивает электронный обмен отчётными документами, разграничение прав пользователей, доступ к аналитической информации из Internet/Intranet.

В основе концепции OLAP лежит принцип многомерного представления данных. По измерениям в многомерной модели выделяют факторы, влияющие на деятельность предприятия (например, время, продукты, отделения компании и т.п.) и получают гиперкуб, который затем наполняется показателями деятельности предприятия (цены, продажи, план, прибыли, убытки и т.п.). Наполнение это может вестись как реальными данными оперативных систем, так и прогнозируемыми на основе исторических данных, то есть данных, накопленных за определённый период времени.

## **Основная часть.**

### 2. ОБЩИЕ ТРЕБОВАНИЯ К OLAP-СИСТЕМАМ И СПОСОБЫ ХРАНЕНИЯ ДАННЫХ В НИХ

В 1993 году Е.Ф. Коддом — создателем концепции реляционных СУБД и, по совместительству, OLAP — были сформулированы критерии OLAP. Они заключаются в недостатках реляционной модели и, в первую очередь, указывают на невозможность «объединять, просматривать и анализировать данные с точки зрения множественности измерений, то есть самым понятным для корпоративных

аналитиков способом». Общие требования к системам OLAP, расширяют функциональность реляционных СУБД и включают многомерный анализ как одну из своих характеристик.

Кодд определил 12 правил, которым должен удовлетворять программный продукт класса OLAP:

Многомерное концептуальное представление данных (Multi-Dimensional Conceptual View). Концептуальное представление модели данных в продукте OLAP должно быть многомерным по своей природе, то есть позволять аналитикам выполнять интуитивные операции «анализа вдоль и поперёк» («slice and dice»), вращения (rotate) и размещения (pivot) направлений консолидации.

Прозрачность (Transparency). Пользователь не должен знать о том, какие конкретные средства используются для хранения и обработки данных, как данные организованы и откуда берутся.

Доступность (Accessibility). Аналитик должен иметь возможность выполнять анализ в рамках общей концептуальной схемы, но при этом данные могут оставаться под управлением оставшихся от старого наследства СУБД, будучи при этом привязанными к общей аналитической модели. То есть, инструмент OLAP должен накладывать свою логическую схему на физические массивы данных, выполняя все преобразования, требующиеся для обеспечения единого, согласованного и целостного взгляда пользователя на информацию.

Устойчивая производительность (Consistent Reporting Performance). С увеличением числа измерений и размеров базы данных аналитики не должны столкнуться с каким бы то ни было уменьшением производительности. Устойчивая производительность необходима для поддержания простоты использования и свободы от усложнений, которые требуются для доведения OLAP до конечного пользователя.

Клиент – серверная архитектура (Client-Server Architecture). Главная идея работы в среде клиент – сервер — это то, что серверный компонент инструмента OLAP должен быть достаточно интеллектуальным и обладать способностью стоять общую концептуальную схему на основе обобщения и консолидации различных логических и физических схем корпоративных баз данных для обеспечения эффекта прозрачности.

Равноправие измерений (Generic Dimensionality). Все измерения данных должны быть равноправными. Дополнительные характеристики могут быть предоставлены отдельным измерениям. Но поскольку все они симметричны, данная дополнительная функциональность может быть предоставлена любому измерению. Базовая структура данных, формулы и форматы отчётов не должны опираться на какое-то одно измерение.

Динамическая обработка разреженных матриц (Dynamic Sparse Matrix Handling). Инструмент OLAP должен обеспечивать оптимальную обработку разреженных матриц. Скорость доступа должна сохраняться вне зависимости от расположения ячеек данных и быть постоянной величиной для моделей, имеющих разное число измерений и различную разреженность данных.

Поддержка многопользовательского режима (Multi-User Support). Зачастую несколько аналитиков имеют необходимость работать одновременно с одной аналитической моделью или создавать различные модели на основе одних корпоративных данных. Инструмент OLAP должен предоставлять им конкурентный доступ, обеспечивать целостность и защиту данных.

Интуитивное манипулирование данными (Intuitive Data Manipulation). Детализация данных в колонках и строках, агрегация и другие манипуляции, свойственные структуре иерархии, должны выполняться в максимально удобном, естественном и комфортном пользовательском интерфейсе.

Гибкий механизм генерации отчётов (Flexible Reporting). Должны поддерживаться различные способы визуализации данных, то есть отчёты должны представляться в любой возможности ориентации.

Неограниченное количество измерений и уровней агрегации (ed Dimensions and Aggregation Levels). Настоятельно рекомендуется допущение в каждом серьёзном OLAP инструменте как минимум пятнадцати измерений в аналитической модели. Более того, каждое из этих измерений должно допускать практически неограниченное количество определённых пользователем уровней агрегации.

Набор этих требований, послуживших фактическим определением OLAP, следует рассматривать как рекомендательный, а конкретные продукты оценивать по степени приближения к идеально полному соответствию всем требованиям.

Позднее все эти требования были переработаны в так называемый тест FASMI, который также определяет требования к продуктам OLAP. FASMI — это

аббревиатура от названия каждого пункта теста:

**Fast (Быстрый).** Приложение OLAP должно обеспечивать минимальное время доступа к аналитическим данным — в среднем порядка 5 секунд;

**Analysis (Анализ).** Приложение OLAP должно давать пользователю возможность осуществлять числовой и статистический анализ;

**Shared (Разделяемый доступ).** Приложение OLAP должно предоставлять возможность работы с информацией многим пользователям одновременно;

**Multidimensional (Многомерность).** Приложение должно обеспечивать многомерное концептуальное представление данных, включая полную поддержку для иерархий.

**Information (Информация).** Приложение OLAP должно давать пользователю возможность получать нужную информацию, в каком бы электронном хранилище данных она не находилась.

Данные могут храниться либо в реляционных, либо в многомерных структурах. Поэтому в настоящее время применяются три способа хранения данных:

**MOLAP (Multidimensional OLAP)** — исходные и агрегатные данные хранятся в многомерной базе данных. Хранение данных в многомерных структурах позволяет манипулировать данными как многомерным массивом, благодаря чему скорость вычисления агрегатных значений одинакова для любого из измерений. Однако в этом случае многомерная база данных оказывается избыточной, так как многомерные данные полностью содержат исходные реляционные данные.

**ROLAP (Relational OLAP)** — исходные данные остаются в той же реляционной базе данных, где они изначально и находились. Агрегатные же данные помещают в специально созданные для их хранения служебные таблицы в той же базе данных.

**HOLAP (Hybrid OLAP)** — исходные данные остаются в той же реляционной базе данных, где они изначально находились, а агрегатные данные хранятся в многомерной базе данных.

Некоторые OLAP-средства поддерживают хранение данных только в реляционных структурах, некоторые — только в многомерных. Однако большинство современных серверных OLAP-средств поддерживают все три способа хранения данных. Выбор способа хранения зависит от объема и структуры исходных данных, требований к скорости выполнения запросов и частоты обновления OLAP-кубов.

Отметим также, что подавляющее большинство современных OLAP-средств не хранит «пустых» значений (примером «пустого» значения может быть отсутствие продаж сезонного товара вне сезона).

Одним из популярных средств разработки OLAP-систем является семейство программных продуктов Oracle Express OLAP компании Oracle. Программное обеспечение Oracle Express предоставляет широкие возможности для создания аналитических систем на основе сервера многомерных баз данных — Oracle Express Server. В состав инструментальных средств Oracle Express входят средства создания и администрирования многомерных баз данных — Express Administrator, средство визуального создания “облегченных” клиентских приложений и презентаций — Express Analyzer, профессиональная инструментальная среда объектно-ориентированной разработки OLAP-приложений — Express Objects, позволяющая создавать сложные интегрированные клиентские приложения, и другие средства, связанные с публикацией данных в Интернете.

## **Заключение.**

В последние годы аналитическая обработка данных привлекает все большее внимание в мире. Например, аналитические модули появились в составе всех основных западных и российских финансово-производственных приложений — ведь в условиях рыночной экономики качество информационной поддержки деятельности руководителей и аналитиков является одним из факторов достижения успеха предприятия. OLAP и является той технологией, которая превращает “сырые” данные OLTP в информацию и знание для конечных пользователей.

В заключение, можно определить OLAP как совокупность средств анализа данных, накопленных в хранилище, а также его основные особенности:

Выделение из большого объёма исторических данных содержательной информации (знаний) с использованием средств обработки информации на основе методов искусственного интеллекта

Использование мощной вычислительной техники и специального хранилища данных, которое накапливает информацию из различных источников за большой период времени, а также обеспечение оперативного доступа к данным.