

image not found or type unknown



Объем данных в организациях настолько возрос, что привел к увеличению массива знаний, который выходит за рамки экономической ценности и практической применимости. Это дало толчок к развитию информационных технологий, появлению интеллектуальных технологий анализа деловых данных, аналитических систем и систем интеллектуальной поддержки принятия решений на их базе. Новые информационные технологии позволили найти нетривиальные подходы к автоматизации управленческого труда и отказаться от старых методов управления.

Технологии интеллектуального анализа данных обеспечивают формирование аналитических данных посредством выполнения операции очищения данных локальных баз предприятия, применения статистических методов и других сложных алгоритмов. Появлению аналитических систем способствовало осознание руководящим звеном предприятий факта, что в базах данных содержится не только информация, но и знания (скрытые закономерности). Последние позволяют охарактеризовать процесс управления предприятием и дать интеллектуальную информацию для более обоснованного принятия решений.

Можно выделить следующие технологии интеллектуального анализа данных:

- ○ оперативный анализ данных посредством OLAP-систем;
- поиск и интеллектуальный выбор данных Data Mining;
- деловые интеллектуальные технологии BIS;
- интеллектуальный анализ текстовой информации.

Аналитические системы OLAP (On-Line Analytical Processing) предназначены для анализа больших объемов информации в интерактивном режиме для создания интеллектуального капитала (аналитических данных), позволяющего руководителю принять обоснованное решение. Они обеспечивают:

- ○ агрегирование и детализацию данных по запросу;
- выдачу данных в терминах предметной области;
- анализ деловой информации по множеству измерений (например, поставщик, его местоположение, поставляемый товар, цены, сроки поставки и т.д.);

- многопроходный анализ информации, который позволяет выявить не всегда очевидные тенденции в исследуемой предметной области;
- произвольные срезы данных по наименованию, выбираемых из разных внутренних и внешних источников (например, по наименованию товара, сроках его изготовления);
- выполнение аналитических операций с использованием статистических и других методов;
- согласование данных во времени для использования в прогнозах, трендах, сравнениях (например, согласование курса рубля).

Аналитические системы позволяют использовать данные новым образом. Вместо поиска отдельных фактов они позволяют получать результаты не через экспериментирование, теоретизирование или моделирование, а посредством информационных операций (установление корреляций, тенденций, других статистических методов). Появилась еще одна форма информационного процесса – **наблюдение за текущей информацией**.

Концепция технологии OLAP была сформулирована Эдгаром Коддом в 1993 году. Она стала ключевым компонентом организации данных в информационных хранилищах и их применении. Эта технология основана на построении многомерных наборов данных – OLAP-кубов. Целью использования технологий OLAP является анализ данных и представление этого анализа в виде, удобном для восприятия управленческим персоналом и принятия на их основе решений.

Основные требования, предъявляемые к приложениям для многомерного анализа:

- ○ предоставление пользователю результатов анализа за приемлемое время (не более 5 сек.);
- осуществление логического и статистического анализа, его сохранение и отображение в доступном для пользователя виде;
- многопользовательский доступ к данным;
- многомерное представление данных;
- возможность обращаться к внутренним и внешним базам.

Аналитические данные содержат факты и агрегатные данные.

Факт – это число, значение. Над фактами производятся различные операции: суммирование, группировка, вычисление средних, максимальных, минимальных значений для получения агрегатных данных.

Агрегатное данное – суммарное, среднее, минимальное, максимальное и другое значение, полученное посредством статистических операций над фактами. Операции над фактами выполняются вдоль определенных измерений.

Под **измерением** понимается один из ключей данных, в разрезе которого можно выполнять разные операции: получать, фильтровать, группировать и отражать информацию о фактах. Примеры измерений: страна, клиент, товар, поставщик.

Измерения могут иметь иерархическую структуру. Например, в стране может быть несколько городов, в городе – несколько клиентов, их могут обслуживать различные поставщики из тех же или других городов и стран.

Для отображения иерархии измерений используются различные модели иерархий. Модели иерархий служат основой построения многомерных баз данных и метаданных в информационных хранилищах.

Разработано несколько способов хранения аналитических данных. Наибольший эффект достигается при использовании многомерных кубов. Рассмотрим на примерах понятие многомерного куба.

Посредством запроса «Какова суммарная стоимость заказов, сделанных клиентами <страны>» можно построить одномерный набор агрегатных значений. Здесь измерение – страна, агрегат – суммарная стоимость заказов рис 8.5.

Страна Суммарная стоимость заказов(млн .руб)

Россия 500

Франция 300

Англия 400

Рис 8.5

Посредством запроса «Какова суммарная стоимость заказов, сделанных клиентами <страны> и доставленных <компанией>» можно построить двумерный набор данных следующего вида (рис 8.6)

Страна Поставщик суммарная стоимость заказов (млн. руб.)

Россия П1 300

П2 200

Франция ПЗ 100

П4 200

Англия П5 400

Рис.8.6

Измерениями являются страна и поставщик, агрегатом – суммарная стоимость заказов.

Посредством запроса «Какова суммарная стоимость заказов, сделанных клиентами <страны> в <году> и доставленных <компанией>» получаем трехмерный куб (рис. 8.7). Этот куб позволяет получать «исторические данные» за любой период.

The diagram shows a 3D cube with three dimensions: 2002 г., 2001 г., and 2000 г. Below the cube is a table with three columns: Страна, Поставщик, and Суммарная стоимость заказов. The table has three rows, representing the data for each year.

Страна	Поставщик	Суммарная стоимость заказов

Рис. 8.7. Пример трехмерного куба

Измерениями являются год, страна, поставщик, агрегатом – суммарная стоимость заказов.

Если учесть, что в каждой стране может существовать несколько клиентов, то добавляется четвертое измерение. Некоторым аналитикам может потребоваться до 20 измерений.

Многомерный куб сам по себе для анализа не пригоден. Для использования из многомерного куба извлекают обычные двумерные таблицы. Эта операция называется «**разрезанием**» куба. При этом куб можно разрезать по разным измерениям, получая разные срезы (сечения) – разные таблицы.

Агрегатные данные берутся из нижних уровней иерархии, а затем суммируются для получения значений более высоких уровней. Для того чтобы ускорить процесс перехода, просуммированные значения для разных уровней хранятся в кубе. Таким образом, то, что пользователю кажется кубом, грубо говоря, состоит из множества

более примитивных кубов. Например, если данные содержали информацию о дневной выручке по конкретному товару в одном магазине, то при формировании куба OLAP-приложение считает итоговые суммы для разных уровней иерархий (недель, месяцев, одного магазина, магазинов города, страны). Это повышает скорость доступа к аналитическим данным, но увеличивает объем памяти для хранения.

Как исходные, так и агрегатные данные могут храниться либо в реляционных, либо в многомерных базах данных MDD (Multidimensional Data).

Многомерный анализ данных может быть произведен посредством клиентских приложений и серверных OLAP-систем.

Клиентские приложения, содержащие OLAP-средства, позволяют вычислять агрегатные данные. Агрегатные данные размещаются в кэш внутри адресного пространства такого OLAP-средства. Кэш – быстродействующий буфер большой емкости, работающий по специальному алгоритму. При этом, если исходные данные находятся в реляционной базе, вычисления производятся OLAP-средствами клиентского приложения. Если исходные данные размещаются на сервере баз данных, то OLAP-средства приложений посылают SQL-запросы на сервер баз данных и получают агрегатные данные, вычисленные сервером.

Примерами клиентских приложений, содержащими OLAP-средства, являются приложения статистической обработки данных SEWSS (Statistic Enterprise – Wide SPS System) фирмы StatSoft и MS Excel 2000. **Excel** позволяет создать и сохранить небольшой локальный многомерный OLAP-куб и отобразить его двух- или трехмерные сечения (разреза). MS Excel 2000 позволяет делать запросы к OLAP-серверам.

Многие средства проектирования позволяют создавать простейшие OLAP-средства. Например, Borland Delphi и Borland C++ Builder.

Отметим, что клиентские приложения применяются при малом числе измерений (не более шести) и небольшом разнообразии значений этих измерений.

Серверные OLAP-системы развили идею сохранения кэш с агрегатными данными.

В них сохранение и изменение агрегатных данных, поддержка содержащего их хранилища осуществляется отдельным приложением (процессом), называемым OLAP-сервером. Клиентские приложения делают запросы к OLAP-серверу и

получают требуемые агрегатные данные. Серверные OLAP-системы рассчитаны на любое количество измерений.

Применение OLAP-серверов сокращает трафик сети, время обслуживания запросов, сокращает требования к ресурсам клиентских приложений.

В масштабе предприятия обычно используются OLAP-серверы типа Oracle Express Server, MS SQL Server 2000 Analysis Services и др.

Серверные OLAP-системы на базе информационных хранилищ поддерживают все способы хранения данных.

Аналитическая система обеспечивает выдачу агрегатных данных по запросам клиентов. Сложность аналитических систем вызвана реализацией сложных интеллектуальных запросов. **Интеллектуальные запросы** осуществляют поиск по условию или алгоритму вычисления ответа. Например, выбрать для выпуска изделия, приносящие максимальную прибыль. Само условие может доопределяться в ходе формирования ответа, что усложняет алгоритм формирования ответа. Данные для формирования ответа могут находиться в разных внутренних и внешних базах.

Существующий язык запросов SQL расширяется возможностью построения интеллектуальных запросов. Пример такого запроса – сравнить данные о продажах в конкретные месяцы, но разные годы. Для таких запросов используются непроцедурные языки обращения к многомерным базам данных. Примером такого языка запросов является язык MDX (Multidimensional Expressions). Он позволяет формировать запрос и описывать алгоритм вычислений. Язык SQL используется для извлечения данных из локальных баз. Язык MDX служит для извлечения данных из многомерных баз и информационных хранилищ.

Аналитические данные используются в системах поддержки принятия решений.

Самые современные аналитические системы основываются на информационных хранилищах и обеспечивают весь спектр аналитической обработки. Доступ к информационным хранилищам реализован посредством транзакций и по интеллектуальным запросам. По интеллектуальным запросам OLAP-системы информационное хранилище выдает аналитические данные. По запросам, объединенным в транзакции других систем, информационное хранилище обеспечивает их обработку, выдачу ответов и отчетов, но не обеспечивает функцию анализа данных. Именно этим и отличаются **OLTP-системы** (On-Line

Transaction Processing) от OLAP-систем.

Примером OLAP-систем является Brio Query Enterprise корпорации Brio Technology. OLAP-средства включают в свои системы отечественные фирмы 1С, Парус и др.

Технологии Data Mining (добыча данных) разработаны для поиска и выявления в данных скрытых связей и взаимозависимостей с целью предоставления их руководителю в процессе принятия решений. Для этого используются статистические методы корреляции, оптимизации и методы, позволяющие находить эти зависимости и синтезировать дедуктивную (обобщающую) информацию. Технологии **Data Mining** обеспечивают:

- поиск зависимых данных (реализацию интеллектуальных запросов);
- выявление устойчивых бизнес-групп (выявление групп объектов, близких по заданным критериям);
- ранжирование важности измерений при классификации объектов для проведения анализа (страна, город, район, поставщик);
- прогнозирование бизнес-показателей (например, ожидаемые продажи, спрос);
- оценка влияния принимаемых решений на достижение успеха предприятия;
- поиск аномалий и т.д.

Технологии Data Mining позволяют **наблюдать за текущей информацией** с целью поиска отклонений, тенденций без вникания в смысл самих данных. Их используют, например, для оценки поведения покупателей, чтобы внести изменения в рекламную тактику, для корректировки выпуска продукции, изменения ценовой политики и т.д.

Интеллектуальные деловые технологии BIS (Business Intelligence Services) преобразуют информацию из внутренних и внешних баз в интеллектуальный капитал (аналитические данные) предприятия. Главной задачей систем интеллектуального выбора данных является поиск функциональных и логических закономерностей в накопленных данных для подсказки обоснованных управленческих решений. Они основаны на применении технологий информационного хранилища и алгоритмов автоматизации деловых процессов (Workflow). Аналитические данные предоставляются руководству всех уровней и работникам аналитических служб организации по запросам в удобном виде.

Для **интеллектуального анализа текстовой информации** разработаны **структурные аналитические технологии (SAT)**. Они ориентированы на углубленную обработку неструктурированной информации. Реализуют уникальную

способность человека интерпретировать (толковать) содержание текстовой информации и устанавливать связи между фрагментами текста. САТ реализованы на базе гипертекстовой технологии, лингвистических процессоров, семантических сетей. Гипертекстовая технология используется для построения смысловых связей между фрагментами текста, лингвистические процессоры и семантические сети – для анализа смысловых фрагментов.

Структурные аналитические технологии предназначены для решения разнообразных задач аналитического характера на основе структуризации предварительно отобранной текстовой информации. Являются инструментом создания аналитических докладов, отчетов, статей, заметок для использования в информационно-аналитических службах организаций, отраслей, государственного управления, СМИ и т.д.