

image not found or type unknown



Технологии интеллектуального анализа данных обеспечивают формирование аналитических данных посредством выполнения операции очищения данных локальных баз организации, 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 интеллектуального анализа данных;
- \* Деловые интеллектуальные технологии БИС;
- \* Интеллектуальный анализ текстовой информации.

Аналитические системы OLAP (On-Line Analytical Processing) предназначены для анализа больших объемов информации в интерактивном создании 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 параметров (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);

\* Выполнение аналитических операций с использованием статистических и других методов;

\* Согласование данных во времени для использования в прогнозах, трендах, сравнениях (например, согласование курса рубля).

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 любой информации независимо от места ее хранения и объема.

Многомерный анализ может быть реализован средствами анализа данных офисных приложений и распределенными системами OLAP-.

Исходные и аналитические данные могут храниться по-разному. Наибольший эффект достигается при использовании многомерных кубов. Рассмотрим на примерах понятие многомерного куба.

Если учесть, что в каждой стране может существовать несколько клиентов, то добавляется четвертое измерение.

Вообще под измерением понимается один из ключей данных, в разрезе которого можно получать, фильтровать, группировать и отражать информацию о фактах. Примеры измерений: страна, клиент, товар, поставщик. Факт - это число, значение. Факты можно суммировать вдоль определенного измерения. Их можно группировать, выполнять над ними другие статистические операции. Агрегатное значение - суммарное, среднее, минимальное, максимальное и другое значение, полученное посредством статистических операций.

В настоящее время применяются три способа хранения многомерных баз данных:

\* Системы оперативной аналитической обработки многомерных баз данных MOLAP (многомерные OLAP) - исходные и агрегатные данные хранятся в многомерной базе данных. Многомерные базы данных представляют собой гиперкубы или поликубы. В гиперкубах все измерения имеют одинаковую размерность, в поликубе каждое измерение имеет свою размерность. Многомерная база данных оказывается избыточной, так как она полностью содержит исходные данные реляционных баз.

\* Системы оперативной аналитической обработки реляционных баз данных

ROLAP (реляционных OLAP) - исходные данные остаются в реляционной базе, агрегатные данные размещаются в кэш той же базы.

\* Гибридные системы оперативной аналитической обработки данных HOLAP (Hybrid OLAP) - исходные данные остаются в реляционной базе, а агрегатные данные хранятся в многомерной базе данных (MDD). 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 (Статистика предприятие - Широкий СПС система) фирмы StatSoft и MS Excel 20.00. Excel позволяет создать и сохранить двух или трехмерный небольшой OLAP-куб.

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. интеллектуальный данные офисный

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. 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.

Заметим, что MS Excel 2000 позволяет делать запросы к OLAP-серверам.

Серверных 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 (многомерные выражения). 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. 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 поэтому эти

системы называются системами OLTP-(On-Line обработки транзакций) в отличие от OLAP-систем.

Примером OLAP-систем является Brio Query Enterprise корпорации Brio Technology. OLAP-средства включают в свои системы фирмы 1С, Парус и др. Технологии интеллектуального анализа данных (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 другие, позволяющие находить эти зависимости и синтезировать дедуктивную (обобщающую) информацию. Технологии 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;

\* Прогнозирование бизнес - показателей (например, ожидаемые продажи, спрос);

\* Оценка влияния принимаемых решений на достижение успеха предприятия;

\* Поиск аномалий и т.д.

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

Интеллектуальные деловые технологии БИС (Business Intelligence Services) преобразуют информацию из внутренних и внешних баз в интеллектуальный капитал (аналитические данные). Главными задачами 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 (КПП). Они ориентированы на углубленную обработку неструктурированной информации. Реализуют уникальную способность человека интерпретировать (толковать) содержание текстовой информации и устанавливать связи между фрагментами текста. КПП реализованы на базе гипертекстовой технологии, лингвистических процессоров, семантических сетей.

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.