



Создание систем поддержки принятия решений на основе искусственного интеллекта



Выполнил:

Козін В.В.

Група

ІТ-81

Перевірив:

Лавров Є. А.

Зарождение концепции хранилища данных

В той или иной степени Системы Поддержки Принятия Решений (СППР) присутствуют в любой информационной системе (ИС). Поэтому, осознанно или нет, к задаче создания системы поддержки принятия решений организации приступают сразу после приобретения вычислительной техники и установки программного обеспечения. По мере развития бизнеса, упорядочения структуры организации и налаживания межкурпоративных связей, проблема разработки и внедрения СППР становится особенно актуальной. Одним из подходов к созданию таких систем стало использование хранилищ данных.

СППР первого типа получили название Информационных Систем Руководства (Executive Information Systems, ИСР). По сути, они представляют собой конечные наборы отчетов, построенные на основании данных из транзакционной информационной системы предприятия или OLTP-системы, в идеале адекватно отражающей в режиме реального времени все аспекты производственного цикла предприятия. Для ИСР характерны следующие основные черты:

- ⦿ отчеты, как правило, базируются на стандартных для организации запросах; число последних относительно невелико;
- ⦿ ИСР представляет отчеты в максимально удобном виде, включающем, наряду с таблицами, деловую графику, мультимедийные возможности и т. п.;
- ⦿ как правило, ИСР ориентированы на конкретный вертикальный рынок, например финансы, маркетинг, управление ресурсами.

СППР второго типа предполагают достаточно глубокую проработку данных, специально преобразованных так, чтобы их было удобно использовать в ходе процесса принятия решений. Неотъемлемым компонентом СППР этого уровня являются правила принятия решений, которые на основе агрегированных данных подсказывают менеджерскому составу выводы и придают системе черты искусственного интеллекта. Такого рода системы создаются только в том случае, если структура бизнеса уже достаточно определена и имеются основания для обобщения и анализа не только данных, но и процессов их обработки. Если ИСУ есть не что иное как развитие системы оперативного управления производственными процессами, то СППР в современном понимании - это механизм развития бизнеса, который включает в себя некоторую часть управляющей информационной системы, обширную систему внешних связей предприятия, а также технологические и маркетинговые процессы развития производства.

Технология разработки и внедрения Хранилища Данных

- этапы проекта;
- выбор модели данных Хранилища;
- выбор структуры Хранилища Данных;
- витрины данных;
- хранилища метаданных (Репозитарий);
- загрузка Хранилища;
- анализ данных: OLAP;

Основное назначение модели предприятия - определение и формализация данных, действительно необходимых в процессе принятия решения. Известно два подхода к бизнес - анализу.

В ходе анализа бизнес - событий необходимо также сформировать схему взаимодействия между транзакционной и аналитической системами на предприятии. Помимо того, что транзакционная система зачастую является важнейшим источником данных для хранилища, желательно задействовать один и тот же пользовательский интерфейс в ИСР и СППР. Подходы к совместному использованию этих систем определяются именно на данной фазе выполнения проекта.

Итак, по результатам анализа бизнес-процессов и структур данных предприятия отбирается действительно значимая для бизнеса информация с учетом неопределенности будущих запросов. Следующий шаг связан с пониманием того, в каком виде и на каких аппаратных и программных платформах размещать структуру данных СППР на основе ХД.

В самом простом варианте для Хранилищ Данных используется та модель данных, которая лежит в основе транзакционной системы. Если, как это часто бывает, транзакционная система функционирует на реляционной СУБД (Oracle, Informix, Sybase и т. п.), самой сложной задачей становится выполнение запросов ad-hoc, поскольку невозможно заранее оптимизировать структуру БД так, чтобы все запросы работали эффективно.

OLAP-системы построены на двух базовых принципах:

- ⦿ все данные, необходимые для принятия решений, предварительно агрегированы на всех соответствующих уровнях и организованы так, чтобы обеспечить максимально быстрый доступ к ним;
- ⦿ язык манипулирования данными основан на использовании бизнес - понятий.

При определении программно-технологической архитектуры Хранилища следует иметь в виду, что система принятия решения, на какие бы визуальные средства представления она ни опиралась, должна предоставить пользователю возможность детализации информации.

Несколько лет назад для Хранилищ Данных было предложено использовать схемы данных, получившие названия "звезда" и "снежинка". Суть технологии проектирования этих схем заключается в выделении из общего объема информации собственно анализируемых данных (или фактов) и вспомогательных данных (называемых измерениями). Необходимо, однако, отдавать себе отчет в том, что это приводит к дублированию данных в Хранилище, снижению гибкости структуры и увеличению времени загрузки. Все это - плата за эффективный и удобный доступ к данным, необходимый в СППР.

Поскольку в Хранилищах Данных, наряду с детальными, должны храниться и агрегированные данные, в случае "снежинки" или "звезды" появляются таблицы агрегированных фактов (агрегатов). Подобно обычным фактам, агрегаты могут иметь измерения. Кроме того, они должны быть связаны с детальными фактами для обеспечения возможной детализации. По некоторым оценкам, при определении оптимального количества агрегатов следует придерживаться принципа 80:20 - 80% ускорения достигается за счет использования 20% кандидатов на агрегаты.

Идея Витрины Данных (Data Mart) возникла несколько лет назад, когда стало очевидно, что разработка корпоративного хранилища - долгий и дорогостоящий процесс. Это обусловлено как организационными, так и техническими причинами:

- информационная структура реальной компании, как правило, очень сложна, и руководство зачастую плохо понимает суть происходящих в компании бизнес-процессов;
- технология принятия решений ориентирована на существующие технические возможности и с трудом поддается изменениям;
- может возникнуть необходимость в частичном изменении организационной структуры компании;
- требуются значительные инвестиции до того, как проект начнет окупаться;
- как правило, требуется значительная модификация существующей технической базы;
- освоение новых технологий и программных продуктов специалистами компании может потребовать много времени;
- на этапе разработки бывает трудно наладить взаимодействие между разработчиками и будущими пользователями Хранилища.

Принципиальное отличие Системы Поддержки Принятия Решений на основе Хранилищ Данных от интегрированной системы управления предприятием состоит в обязательном наличии в СППР метаданных. В общем случае метаданные помещаются в централизованно управляемый Репозиторий.

Широко известны Репозитории, входящие в состав популярных CASE-средств (Power Designer (Sybase), Designer 2000 (Oracle), Silverrun (CSA Research)), систем разработки приложений (Developer 2000 (Oracle), Power Builder (Sybase)), администрирования и поддержки информационных систем (Platinum, MSP). Все они, однако, решают частные задачи, работая с ограниченным набором метаданных, и предназначены, в основном, для облегчения труда профессионалов - проектировщиков, разработчиков и администраторов информационных систем. Репозиторий метаданных СППР на основе ХД предназначен не только для профессионалов, но и для пользователей, которым он служит в качестве поддержки при формировании бизнес - запросов.

Разработка системы управления метаданными сходна с разработкой распределенной транзакционной системы. При ее создании необходимо решать следующие задачи:

- ⦿ анализ процессов возникновения, изменения и использования метаданных;
- ⦿ проектирование структуры хранения метаданных (например, в составе реляционной базы данных);
- ⦿ организация прав доступа к метаданным;
- ⦿ блокировка и разрешение конфликтов при совместном использовании метаданных (что очень часто возникает при изменении общих бизнес - понятий в рамках структурного подразделения);
- ⦿ разделение метаданных между Витринами Данных;
- ⦿ согласование метаданных ХД с Репозиториями CASE-средств, применяемых при проектировании и разработке Хранилищ;
- ⦿ реализации пользовательского интерфейса с Репозитарием.

При описании технологии заполнения Хранилища будем различать три взаимосвязанные задачи: Сбор Данных (Data Acquisition), Очистка Данных (Data Cleansing) и Агрегирование Данных (Data Consolidation).

Под Сбором Данных будем понимать процесс, который состоит в организации передачи данных из внешних источников в Хранилище. Лишь некоторые аспекты этого процесса полностью или частично автоматизированы в имеющихся продуктах. Прежде всего, это относится к интерфейсам с существующими БД. Как правило, здесь имеется несколько возможностей.

Под очисткой данных обычно понимается процесс модификации данных по ходу заполнения Хранилища: исключение нежелательных дубликатов, восстановление пропущенных данных, приведение данных к единому формату, удаление нежелательных символов (например, управляющих) и унификация типов данных, проверка на целостность.

При заполнении Хранилища агрегированными данными мы должны обеспечить выборку данных из транзакционной базы данных и других источников в соответствии с метаданными, поскольку агрегирование происходит в терминах бизнес - понятий.

Интеллектуальный анализ данных

Интеллектуальный анализ данных (ИАД) обычно определяют как метод поддержки принятия решений, основанный на анализе зависимостей между данными. В рамках такой общей формулировки обычный анализ отчетов, построенных по базе данных, также может рассматриваться как разновидность ИАД. Чтобы перейти к рассмотрению более продвинутых технологий ИАД, посмотрим, как можно автоматизировать поиск зависимостей между данными.

Существует два подхода. В первом случае пользователь сам выдвигает гипотезы относительно зависимостей между данными. Фактически традиционные технологии анализа развивали именно этот подход. Действительно, гипотеза приводила к построению отчета, анализ отчета к выдвижению новой гипотезы и т. д. Это справедливо и в том случае, когда пользователь применяет такие развитые средства, как OLAP, поскольку процесс поиска по-прежнему полностью контролируется человеком. Во многих системах ИАД в этом процессе автоматизирована проверка достоверности гипотез, что позволяет оценить вероятность тех или иных зависимостей в базе данных.

Второй подход основывается на том, что зависимости между данными ищутся автоматически. Количество продуктов, выполняющих автоматический поиск зависимостей, говорит о растущем интересе производителей и потребителей к системам именно такого типа. Сообщается о резком росте прибылей клиентов за счет верно найденной, заранее неизвестной зависимости. Упоминается пример сети британских универсамов, где ИАД применялся при анализе убытков от хищений товаров в торговых залах.

Процессы ИАД подразделяются на три большие группы: поиск зависимостей (discovery), прогнозирование (predictive modelling) и анализ аномалий (forensic analysis).

Необходимо также упомянуть об интеграции ИАД в информационные системы. Многие методы ИАД возникли из задач экспертного анализа, поэтому входными данными для них традиционно служат "плоские" файлы данных. При использовании ИАД в СППР часто приходится сначала извлекать данные из Хранилища, преобразовывать их в файлы нужных форматов и только потом переходить собственно к интеллектуальному анализу. Затем результаты анализа требуется сформулировать в терминах бизнес - понятий.

- Создание СППР на основе хранилищ данных - сложный, но обозримый процесс, требующий знания бизнеса, программно-технического инструментария и опыта выполнения крупных проектов. Вместе с тем внедрение подобных систем может дать преимущества в бизнесе, которые будут тем ощутимее, чем раньше организация начнет создание СППР. По прогнозам консалтинговой фирмы Gartner Group, к 2010 году примерно 90-95% компаний будут использовать хранилища данных.
- Значимость информационных систем подобного уровня признается и представителями большинства российских компаний. Однако в силу ряда причин, инициативные или заказные работы ведутся зачастую достаточно бессистемно, в основном в двух направлениях:
- закупка и тестирование разнообразных продуктов, применяемых при создании СППР и ХД (к сожалению, большинство из них плохо сопрягаются друг с другом, из-за чего создается ложное впечатление "неподъемности" проблемы);
- решение частного вопроса о повышении производительности отчетных систем путем локального перепроектирования структуры хранения или перехода на более современные и сложные программные средства.