

БЫСТРОДЕЙСТВУЮЩИЕ МНОГОМЕРНЫЕ ХРАНИЛИЩА ДАННЫХ

OLAP (англ. online analytical processing, интерактивная аналитическая обработка) — технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу. Реализации технологии OLAP являются компонентами программных решений класса Business Intelligence. Причина использования OLAP для обработки запросов — скорость. Реляционные базы данных хранят сущности в отдельных таблицах, которые обычно хорошо нормализованы. Эта структура удобна для операционных баз данных (системы OLTP), но сложные многотабличные запросы в ней выполняются относительно медленно. OLAP-структура, созданная из рабочих данных, называется OLAP-куб.

ВВЕДЕНИЕ

Требования к информационным системам и технологиям при управлении рисками на предприятии включают в себя автоматизированный доступ, консолидацию данных, анализ, прогнозирование и оценку рисков, предоставление полученных результатов. Учитывая многообразие и большой объем данных для анализа, информационные системы и основанные на них системы поддержки, принятия решений являются в настоящее время важнейшим условием эффективного управления рисками в масштабе всей компании.

I. Принципы организации хранилища данных

Корпоративное хранилище данных (ХД, DW - Data Warehouse) позволяет создать единый аналитический центр для сбора, хранения, отсеивания и предварительной обработки данных для оценки рисков предприятия. Использование хранилища данных, как аналитического центра системы управления рисками предприятия, является естественным развитием процессов автоматизации технологий корпоративного управления. Принципы организации хранилища:

- Проблемно-предметная ориентация. Данные объединяются в категории и хранятся в соответствии с областями, которые они описывают, а не с приложениями, которые они используют.
- Интегрированность. Данные объединены так, чтобы они удовлетворяли всем требованиям предприятия в целом, а не единственной функции бизнеса.
- Некорректируемость. Данные в хранилище данных не создаются: то есть поступают из внешних источников, не корректируются и не удаляются.
- Зависимость от времени. Данные в хранилище точны и корректны только в том случае, когда они привязаны к некоторому промежутку или моменту времени.

По своей сути Хранилище данных представляет собой центр, в который собирается вся необходимая информация из различных подразделений предприятия. Прежде чем попасть в хранилище данные должны быть соответствующим образом обработаны.

Хранилища данных являются структурированными: они содержат базовые данные, которые образуют единый источник для обработки данных во всех системах поддержки принятия решений. Если возникают разногласия во мнениях, с помощью хранилища данных можно выполнить согласование информации. А элементарные данные, присутствующие в хранилище, могут быть представлены в различной форме, отвечая не только известным требованиям, но еще и неизвестным.

II. Типы хранилищ данных

Хранилища данных исключительно велики, поскольку в них содержатся исторические и детализированные данные. По причине размера объемов данных, находящихся в хранилище, данные подразделяются на два класса: активно и неактивно используемые данные. Хранилища данных содержат интегрированные данные. Они интегрированы на множестве уровней: на уровне ключа, атрибута, на описательном, структурном уровне и так далее. Общие данные и общая обработка данных консолидированы и являются единообразными для всех данных, которые подобны или схожи в хранилище данных.

Куб потенциально содержит всю информацию, которая может потребоваться для ответов на запросы пользователей. При большом количестве агрегатов зачастую полный расчёт происходит только для некоторых измерений, для остальных же производится «по требованию». Существуют три типа OLAP:

- многомерная OLAP (Multidimensional OLAP — MOLAP);
- реляционная OLAP (Relational OLAP — ROLAP);

- гибридная OLAP (Hybrid OLAP — HОLAP).

MOOLAP — классическая форма OLAP, так что её часто называют просто OLAP. Она использует суммирующую БД, специальный вариант процессора пространственных БД и создаёт требуемую пространственную схему данных с сохранением как базовых данных, так и агрегатов. ROLAP работает напрямую с реляционным хранилищем, факты и таблицы с измерениями хранятся в реляционных таблицах, и для хранения агрегатов создаются дополнительные реляционные таблицы. HОLAP использует реляционные таблицы для хранения базовых данных и многомерные таблицы для агрегатов.

Особым случаем ROLAP является «ROLAP реального времени» (Real-time ROLAP — R-ROLAP). В отличие от ROLAP в R-ROLAP для хранения агрегатов не создаются дополнительные реляционные таблицы, а агрегаты рассчитываются в момент запроса. При этом многомерный запрос к OLAP-системе автоматически преобразуется в SQL-запрос к реляционным данным. Каждый тип хранения имеет определённые преимущества, хотя есть разногласия в их оценке у разных производителей. MOOLAP лучше всего подходит для небольших наборов данных, он быстро рассчитывает агрегаты и возвращает ответы, но при этом генерируются огромные объёмы данных. ROLAP оценивается как более масштабируемое решение, использующее к тому же наименьшее возможное пространство. При этом скорость обработки значительно снижается. HОLAP находится посреди этих двух подходов, он достаточно хорошо масштабируется и быстро обрабатывается. Архитектура R-ROLAP позволяет производить многомерный анализ OLTP-данных в режиме реального времени.

Сложность в применении OLAP состоит в создании запросов, выборе базовых данных и разработке схемы, в результате чего большинство современных продуктов OLAP поставляются вместе с предварительно настроенными запросами. Другая проблема — в базовых данных. Они должны быть полными и непротиворечивыми.

С точки зрения пользователя, все варианты выглядят похожими по возможностям. Наибольшее применение OLAP находит в продуктах для финансового планирования, хранилищах данных, решениях класса Business Intelligence.

Маркова Анастасия Александровна, аспирант кафедры систем управления БГУИР, 17anamar@yandex.ru.

Научный руководитель: Кузнецов Александр Петрович, доктор технических наук, профессор кафедры систем управления БГУИР, kuznar@bsuir.by.

III. Выводы

Логическая структура данных хранилища данных существенно отличается от структуры данных источников данных. Для разработки эффективного процесса преобразования необходима хорошо проработанная модель корпоративных данных и модель технологии принятия решений. Данные для пользователя удобно представлять в многомерных базах данных.

Хранилища данных содержат интегрированные данные. Они интегрированы на множестве уровней: на уровне ключа, атрибута, на описательном, структурном уровне и так далее. Общие данные и общая обработка данных консолидированы и являются единообразными для всех данных, которые подобны или схожи в хранилище данных.

Кроме извлечения данных из базы, для принятия решений важен процесс извлечения знаний, в соответствии с информационными потребностями пользователя. С точки зрения пользователя в процессе извлечения знаний из базы данных должны решаться следующие преобразования: данные → информация → знания → полученные решения.

Внедрение хранилища данных в работу предприятий позволяет увеличить скорость принятия обоснованных решений с учетом автоматизированной обработки информации, эффективно использовать не только финансовые, но и человеческие и материальные ресурсы компании.

1. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных OLAP и DataMining, 2004. гл. 4, 5, 7.
2. Ralf Kimball, Margy Ross. The Data Warehouse. 2-Edition. N.Y.: John Wiley, 2002.
3. Ильяшенко, О. Ю. Роль BI-систем в совершенствовании процессов обработки и анализа бизнес информации / И. В. Ильин // Наука и бизнес: пути развития – 2017. – с. 124–131.
4. Андерсон, К. Аналитическая культура. От сбора данных до бизнес-результатов. / М.: Манн, Иванов и Фербер // – 2017. – с. 336.
5. Chen, H. C., Chaing, R. H. L., Storey, V. C. Business intelligence and analytics: From big data to big impact. // MIS Quarterly – 2012. – p. 1165–1188.
6. Trieu, V. H. Getting value from Business Intelligence systems: A review and research agenda. Decision Support Systems – 2017. – p. 111–124.