

Применение технологии аналитической обработки информации в реальном времени для анализа финансовых расчетов предприятия

Валенда С.В.; Чваркова И.Л.

Кафедра Интеллектуальных систем, Факультет Радиофизики и компьютерных технологий
Белорусский Государственный Университет
г. Минск, Республика Беларусь
e-mail: siarhei.valenda@gmail.com, iryna.chvarkova@gmail.com

Аннотация – В данной статье рассматривается задача накопления и анализа больших объемов данных в масштабах производственных предприятий. Предложен вариант поэтапного решения существующей проблемы анализа информации, позволяющий исследовать основные процессы финансовых расчетов предприятия и просматривать их в динамике. Приведен пример использования модели аналитической обработки информации для анализа деятельности производственной компании.

Ключевые слова: размерностные хранилища данных, многомерные кубы данных, аналитическая обработка данных в реальном времени (OLAP).

I. ВВЕДЕНИЕ

Обработка информации, полученной в результате финансовых расчетов крупного предприятия или производственной компании, всегда является сложной задачей, для решения которой создаются специализированные отделы и привлекаются высококвалифицированные специалисты. Главной целью этого процесса является создание максимально полных и наглядных отчетов, отражающих динамику данных, основных тенденций, на основе которых будет проводиться планирование дальнейшей деятельности организации.

Частую применяемые технические средства оказываются слишком сложными либо вообще отсутствуют, что утяжеляет процесс проведения успешного анализа информации. Одним из решений этой проблемы является использование технологии аналитической обработки информации в реальном времени OLAP (Online Analytical Processing). Основой идеей данного подхода является представление множества документов, отчетных форм, таблиц баз данных и прочей необходимой для анализа информации в виде реляционного хранилища данных. В дальнейшем это хранилище используется для интеграции всех данных организации, а также для разделения повседневно используемых данных и данных, которые необходимы только для анализа.

Следующий этап включает в себя непосредственно создание многомерной базы данных, позволяющей получать разнообразные представления накопленных данных. Использование представлений данных по конкретным показателям позволяет производить более успешное и полное планирование деятельности предприятия.

II. СРАВНЕНИЕ ПОДХОДОВ ОБРАБОТКИ ИНФОРМАЦИИ

Классический реляционный OLTP (Online Transaction Processing) подход для осуществления анализа данных подразумевает под собой написание сложных запросов к реляционной базе данных и создание статистики на основе полученной выборки. Данный процесс плохо применим для задач анализа по следующим причинам: большое число таблиц в

реляционной базе данных, сложный характер самих связей, сложность написания необходимых запросов, большое количество избыточных данных, которые невозможно применить в анализе.

Агрегирование анализируемых данных в рамках специализированного хранилища частично устраняет эти проблемы. Следующим шагом в интеграции технологии аналитической обработки информации становится создание многомерной базы данных. Использование заранее агрегированных данных хранилища позволяет создавать многомерные базы данных любой сложности, в зависимости от структуры самого хранилища, просматривать интересующие данные в проекции на бесконечное число измерений и временную ось, добиваться многократного уменьшения времени выполнения необходимых выборок. Созданные на основе данных хранилища вычисляемые атрибуты и иерархии придают анализу дополнительную гибкость.

III. МОДЕЛЬ СИСТЕМЫ МОНИТОРИНГА ОПЕРАЦИОННОЙ ДЕЯТЕЛЬНОСТИ ПРЕДПРИЯТИЯ

Обобщенная модель аналитической системы может быть сведена к представлению в виде 4-ех основных уровней (рис. 1): уровень данных (база данных предприятия, плоские файлы, таблицы, документы), уровень хранилища, уровень многомерного куба, уровень представлений (выходных данных анализа, отчетов).

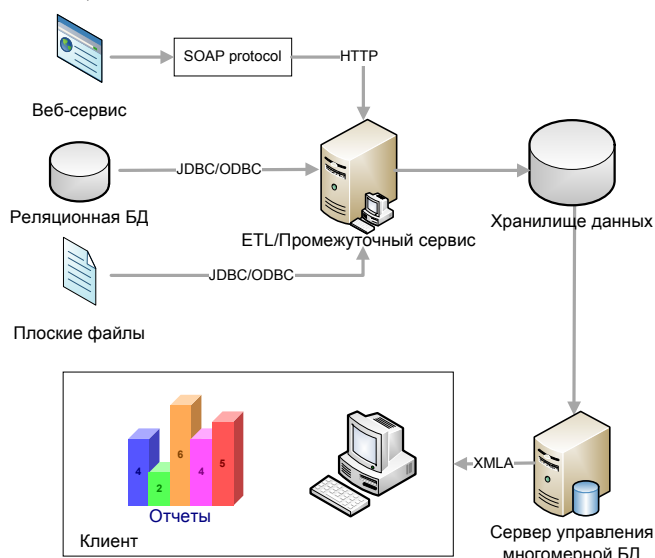


Рис. 1. Модель системы для анализа операционной деятельности

Информация, находящаяся в базе данных предприятия денормализуется и проходит стадию "очистки", прежде чем попасть в хранилище. Сама структура хранилища проектируется с учетом представления основных данных в виде двух ключевых элементов: фактов (исследуемых процессов) и

измерений. При этом измерения являются характеристиками фактов и могут быть представлены в виде иерархически разделенных полей. При помощи специально разработанных программных продуктов производится построение многомерной базы данных. Наиболее распространенным подходом в создании многомерных кубов является реляционный подход (Relational OLAP), который накладывает наибольшие требования к структуре самого хранилища данных. Особенности данного метода являются: данные берутся непосредственно из реляционного хранилища, запрос к многомерной базе данных трансформируется в SQL запрос к хранилищу, дополнительное ускорение работы системы достигается за счет кэширования данных.

Одним из наиболее распространенных программных продуктов, использующих реляционный метод аналитической обработки информации в реальном времени является свободно распространяемый сервер Mondrian ROLAP. Открытый интерфейс программирования приложений, а также дополнительные библиотеки позволяют интегрировать сервер практически в любое приложение.

Последний уровень аналитической системы может быть представлен в виде графиков, сводных таблиц Microsoft Excel, готовых отчетов, либо быть частью системы поддержки принятия решений.

IV. МОДЕЛЬ ИНФОРМАЦИОННОЙ СИСТЕМЫ АНАЛИЗА ФИНАНСОВЫХ РАСЧЕТОВ ПРОИЗВОДСТВЕННОГО ПРЕДПРИЯТИЯ

В ходе выполнения данного исследования была создана модель хранилища данных финансовых расчетов производственного предприятия, которая служила основой для проведения анализа исследуемой системы. В центре модели данных лежали некоторые основные финансовые процессы, информация о которых была собрана из различных источников (файлы, таблицы, базы данных):

- Наименование банка, который принял платеж. Данное поле содержит текстовое значение, которое также должно быть проанализировано в качестве процесса.

- Сумма платежа. Данное поле вводится для того, чтобы определить продукты, которые приносят наибольшую прибыль

- Размер текущей скидки. Данные о процессе могут быть использованы при построении бизнес стратегий.

Измерения для описания анализируемых процессов были представлены связанными посредством ключевых полей стандартными таблицами хранилища данных: товары, заказчик, заказы, отделы, время и др.

Для построения разнообразных проекций были введены дополнительные таблицы измерений: таблица, содержащая информацию о курсах валют и основных биржевых индексах за весь отчетный период. Ключевой таблицей любого анализа, примененной в модели, является таблица времени, имеющая иерархическую структуру полей, представляющая одну из основных

проекции при исследовании динамических показателей и процессов.

Основываясь только на информации, накопленной в таблицах, файлах и базах данных, и заполняемой сотрудниками различных подразделений компании, можно получить необходимые отчетные данные, которые в дальнейшем могут служить базисом для построения системы поддержки принятия решений. Сумма платежей по отдельным банкам может быть спроектирована на временную ось, рассмотрена применительно к определенной группе товаров и географическому региону. При этом сама выборка формируется за минимальное время.

V. ЗАКЛЮЧЕНИЕ

Разработанная модель позволяет достичь ряда преимуществ по сравнению с традиционными методами анализа больших объемов данных уровня предприятия:

- Предприятие получает средство, позволяющее проводить мониторинг финансовых расчетов как всего предприятия в целом, так и отдельных его подразделений, проводить подробный анализ, строить стратегии для будущего развития.

- Низкая стоимость внедрения, благодаря использованию открытого программного обеспечения и стандартов.

- Интеграция с MS Excel позволяет сократить время на обучение сотрудников, так как большинство сотрудников уже знакомы с интерфейсом этого программного продукта.

- Множество средств для формирования отчетов, обладающих понятным интерфейсом, позволяет уменьшить количество рутинной работы и получать необходимые отчеты в максимально короткие сроки.

- Система позволяет проследить динамику изменения основных показателей работы организации в целом и оперативно оценить показатели эффективности расходования средств предприятия.

- Автоматизация процессов загрузки новых данных в хранилище позволяет работать с обновленными данными.

- Автоматическая рассылка готовых отчетов делает возможным ежедневно оповещать руководителей компании об изменениях ключевых показателей, связанных с основными процессами в деятельности предприятия.

[1] Kimpball, R. The Data Warehouse Toolkit / R.Kimpball. – Wiley India Pvt. Ltd., 2009. – 460 p.

[2] Wrembel, R. Data warehouses and OLAP: concepts, architectures, and solutions / R. Wrembel, Chr. Koncilia. – Idea Group Inc., 2007. – 332p.

[3] Ponniah, P. Data warehousing fundamentals for IT professionals / P. Ponniah. – John Wiley&Sons Inc., 2010. – 571 p.

[4] Холод, И. И. Методы и модели анализа данных: OLAP и Data Mining / И. И. Холод, М. С. Куприянов, В. В. Степаненко. – Санкт-Петербург: БХВ-Петербург, 2004. – 336 с.

[5] Pedersen, T. B. Multidimensional Database Technology / T. B. Pedersen, C. S. Jensen // IEEE Computer, 2001. – V. 34, №12. – P. 40-46.