

## ПРОГРАММНАЯ ПОДДЕРЖКА ОБРАБОТКИ И ХРАНЕНИЯ ИНФОРМАЦИИ ПРИ АВТОМАТИЗАЦИИ ПРОИЗВОДСТВЕННЫХ ПРОЦЕССОВ ПРОМЫШЛЕННОГО ПРЕДПРИЯТИЯ

Белорусский государственный университет информатики и радиоэлектроники  
г.Минск, Республика Беларусь

Горегляд В. В.

Дробот С. В., канд.техн.наук, доцент

Использование хранилища данных и системы бизнес-анализа позволяет скрыть от конечных пользователей весь механизм обработки и хранения данных. Это является толчком к более плотной аналитической работе с информацией, что означает повышение степени эффективности использования корпоративных информационных активов. В статье приводится краткое описание жизненного цикла данных и его процессов, исходя из которых разрабатывалась архитектура системы обработки и хранения. Спроектированная архитектура дает представление об алгоритме получения данных от первичных источников до конечного пользователя.

Поддержка современных производственных систем требует сложных процессов обработки и анализа данных. При выполнении этих процессов данные проходят сложный жизненный цикл. Жизненный цикл данных (рисунок 1) может быть описан в виде четырех этапов и подсистем, их реализующих:

1. Данные генерируются как результат обслуживания пользователей системы. Функционирование этого этапа обеспечивается *OLTP* – системами.

2. Данные очищаются, модифицируются и загружаются в хранилище данных с помощью *ETL* – процессов извлечения, преобразования и загрузки. Хранение данных обеспечивается подсистемой хранилища данных (*Data Warehouse*).

3. Данные преобразуются в формат, удобный для обеспечения эффективного доступа и анализа. Поддержку эффективного доступа к данным обеспечивают *OLAP* – системы и многомерная модель данных.

4. Данные анализируются в плане эффективности функционирования предприятия и принятия управленческих решений. Для поддержки анализа данных и составления отчетов существует множество приложений [1].

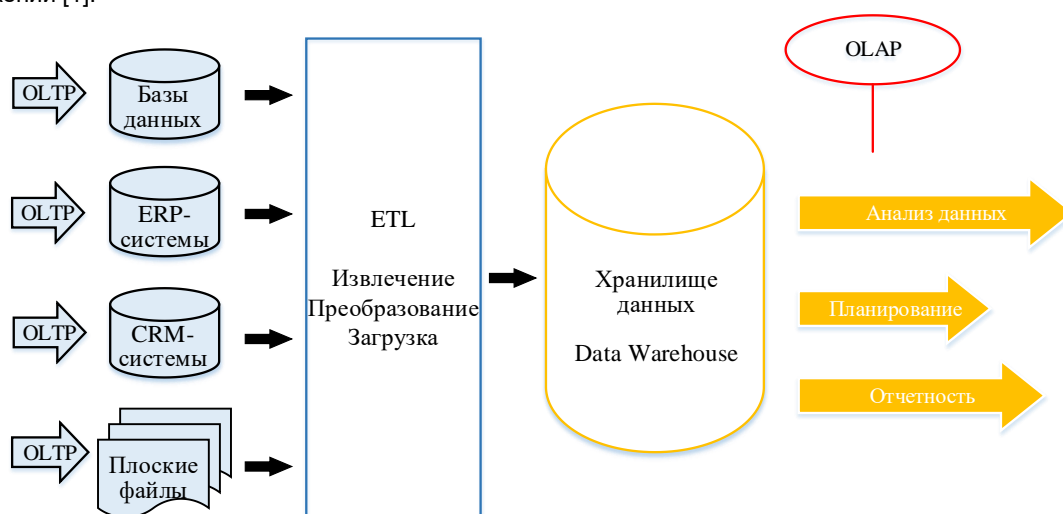


Рисунок 1 – Жизненный цикл данных

Целью работы является формирование единого информационного ресурса для построения отчетности на основе единой модели корпоративных данных. При этом должно поддерживаться высокое качество обрабатываемых данных – их полнота, достоверность, надежность и непротиворечивость, а также высокая скорость подготовки аналитической отчетности на всех уровнях. Хранилище данных должно обеспечивать решение этих задач.

Процесс разработки хранилища данных весьма трудоемок. Можно отметить основные задачи, которые требуется решить в процессе разработки хранилища данных:

- выбор оптимальной структуры хранения данных, обеспечивающей высокую скорость выполнения запросов и минимизацию объема оперативной памяти;
- первоначальное наполнение и дальнейшее пополнение хранилища;
- обеспечение единой методики работы с разнородными данными и создание удобного интерфейса пользователя [2, 3].

Разработанная архитектура системы для хранения и обработки данных, представлена на рисунке 2.

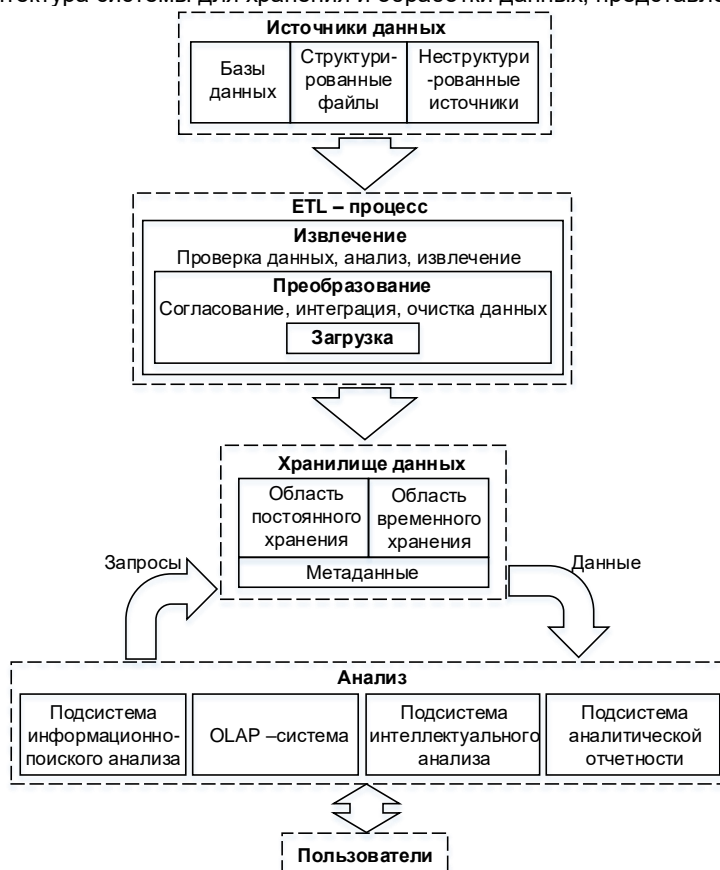


Рисунок 2 – Архитектура системы обработки и хранения

Первичные источники включают в себя всевозможные документы и данные, которые обрабатываются OLTP – системами.

ETL-процесс – это процесс получения информации из OLTP. Как правило, ETL разбивают на три этапа: извлечение, преобразование, загрузка. В нашем случае процессу извлечения отдается большая часть действий, т.к. извлечение данных занимает много времени, оно является общим для выполнения всех процессов, поэтому выполняется параллельно с преобразованием и загрузкой. В то время, как данные извлекаются, происходит преобразование полученной информации и загрузка уже преобразованной без ожидания завершения предыдущих этапов.

Область хранилища данных целесообразно проектировать с учетом областей временного и постоянного хранения. Метаданные обычно определяются как «данные о данных», которые предназначены для выражения семантики информации, улучшения ее поиска и выборки, понимания и использования. Таким образом, метаданные – это информация, которая делает данные полезными.

Область анализа включает в себя четыре подсистемы. В подсистему информационно-поискового анализа входят информационный и визуальный поиск, постановка четких запросов к базе данных и получение соответствующей информации. OLAP-система – это возможность использования многомерных витрин данных на базе Oracle OLAP Option и Oracle Business Intelligence. Подсистема интеллектуального анализа (Data Mining) применяется для выявления в данных, содержащихся в хранилище, скрытых закономерностей, зависимостей и взаимосвязей, полезных при принятии решений на различных уровнях управления. Система аналитической отчетности реализована средствами Oracle Discoverer.

К системе могут иметь доступ различные группы пользователей, которым предоставляются определенные права для работы в задачах в соответствии с предусмотренными для них ролями.

Использование этой структуры позволяет решать более широкий спектр производственных и аналитических задач. Объединение блока ETL повышает быстродействие системы, благодаря параллельному выполнению процессов извлечения, преобразования и загрузки, а также делает более гибкой работу с данными.

#### Список использованных источников:

[1] Бергер, А.Б. Microsoft SQL Server 2005 Analysis Services. OLAP и многомерный анализ данных / Бергер А.Б., Горбач И.В. и др. – СПб.: БХВ-Петербург, 2007. – 928 с.

[2] Научная библиотека по физике и новым технологиям. [Электронный ресурс]: <http://bourabai.kz/>

[3] Жуковский, О.И. Информационные технологии и анализ данных: учебное пособие. / О.И. Жуковский. – Томск: Эль Контент, 2014. – 128с.

[4] Ralph Kimball. The data warehouse ETL toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data / Ralph Kimball, Joe Caserta // John Wiley & Sons. – 2004. – P. 30-36.