9. Hundt C., Schlarb M. and Schmidt B., «SAUCE: A web application for interactive teaching and learning of parallel programming», Journal of Parallel and Distributed Computing, vol. 105, pp. 163–173, July, 2017, doi: 10.1016/j.jpdc.2016.12.028.

10. Robinson P. E. and Carroll J., «An Online Learning Platform for Teaching, Learning, and Assessment of Programming», p. 10, 2017.

11. Bertrand S., Marzat J., Besnerais G. L., Manzanera A., Maniu C. S., and. Makarov M, «Integrating Experimental Data Sets and Simulation Codes for Students into a MOOC on Aerial Robotics», IFAC-PapersOnLine, vol. 52, ed. 9, pp. 50–55, 2019, doi: 10.1016/j.ifacol.2019.08.123.

12. Manzoor H., Naik A., Shaffer C. A., North C., and Edwards S. H., «Auto-Grading Jupyter Notebooks», в Proceedings of the 51st ACM Technical Symposium on Computer Science Education, Portland OR USA, february, 2020, pp. 1139–1144. doi: 10.1145/3328778.3366947.

13. Syaifudin Y. W. and others «Web application implementation of Android programming learning assistance system and its evaluations», IOP Conf. Ser.: Mater. Sci. Eng., v. 1073, ed. 1, p. 012060, february, 2021, doi: 10.1088/1757-899X/1073/1/012060.

14. Barlow M., Cazalas I., Robinson C., and Cazalas J., «MOCSIDE: an Open-source and Scalable Online IDE and Auto-Grader for Introductory Programming Courses», p. 10.

15. Akahane Y., Kitaya H., and. Inoue U, «Design and evaluation of automated scoring Java programming assignments», 2015, IEEE/ACIS 16 th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Takamatsu, June, 2015, pp. 1–6. doi: 10.1109/SNPD.2015.7176255.

16. Maicus E., Peveler M., Patterson S., and Cutler B., «Autograding Distributed Algorithms in Networked Containers», в Proceedings of the 50th ACM Technical Symposium on Computer Science Education, Minneapolis MN USA, february, 2019, pp. 133–138. doi: 10.1145/3287324.3287505.

DOI 10.53364/24138614_2022_25_2_107
УДК 004.738:378

Хажиахмет Т.Н., магистрант
Научный руководитель: **Дузбаев Н.Т.,** проректор по цифровизации и инновациям PhD, асс. профессор
Международный университет информационных технологий, Алматы, РК.

[1]E-mail: tima17.1995@gmail.com
[2]E-mail: n.duzbayev@iitu.edu.kz

РАЗРАБОТКА ETL-СИСТЕМЫ ДЛЯ ЗАГРУЗКИ ДАННЫХ В ХРАНИЛИЩЕ ДАННЫХ

ДЕРЕКТЕР ҚОЙМАСЫНА МӘЛІМЕТТЕРДІ ЖҮКТЕУ ҮШІН ETL ЖҮЙЕСІН ҚҰРУ

DEVELOPMENT OF AN ETL SYSTEM FOR UPLOADING DATA TO A DATA WAREHOUSE

Аннотация. В данной статье описана основная идея разработки ETL-системы для загрузки данных в Хранилище Данных. Представлены основные задачи разработки, а также описан процесс реализации ETL-системы.

Ключевые слова: разработка ETL-системы, Хранилище Данных, BI, СУБД, API, ODI, HTTP Basic Authentication, Target, SAP BO, DMZ.

Аңдатпа. Бұл мақалада мәліметтерді деректер қоймасына жүктеу үшін ETL жүйесін дамытудың негізгі идеясы сипатталған. Дамудың негізгі міндеттері ұсынылған, сонымен қатар ETL жүйесін іске асыру процесі сипатталған.

Түйін сөздер: ETL жүйесін дамыту, деректерді сақтау, BI, ДҚБЖ, API, ODI, HTTP, Target, SAP BO, DS негізгі аутентификациясы.

Abstract. This article describes the main idea of the development of ETL-systems for downloading data in the archive. The main tasks of the development were presented, as well as the process of implementing the ETL system.

Key words: development of ETL systems, Library of data, BI, Subd, API, ODI, HTTP Basic Authentication, Target, SAP BO, DS.

Введение. Хранилище Данных - это специализированная информационная база данных, разработанная и предназначенная для подготовки отчетности и бизнес-анализа с целью поддержки принятия решений в любой компании. Строится на базе систем управления базами данных и систем поддержки принятия решений.

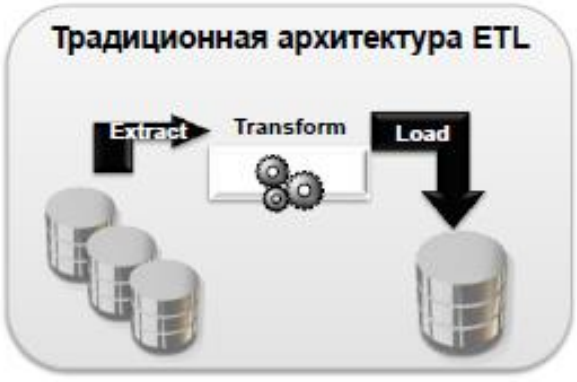

Рис. 1 - ETL (Extract, Transform, Load)

ETL - это один их основных процессов в управлении хранилищами данных, который включает в себя следующие задачи (Рис. 1):

- извлечение данных из внешних источников (таблицы баз данных, файлы);
- преобразование и очистка данных согласно потребностям;
-загрузка обработанной информации в хранилище данных.

Можно выделить следующие характеристики ETL-системы:

- Лучший доступ к данным компании;
- Возможность создавать отчеты и показатели, которые могут управлять стратегией;

В своей статье «Обзор технологий хранения данных и OLAP» Чаудуриан Дайал объяснил, что хранилище данных - это отдельная база данных, которую аналитики могут запрашивать по своему усмотрению, не влияя на работу онлайн-обработки транзакций (OLTP). [3]

Далее будет представлена основная идея ETL-системы, а также процесс разработки.

Основная идея ETL-системы

ETL позволяет предприятиям объединять данные из нескольких баз данных и других источников в единое хранилище с данными, которые были должным образом отформатированы и квалифицированы для подготовки к анализу. Этот единый репозиторий данных обеспечивает упрощенный доступ для анализа и дополнительной обработки.

Процесс разработки

При разработке ETL-системы нужно учитывать функциональные требования, которые реализуют логику системы. Они раскрывают задачи, которые нужно реализовать в разработке ETL-системы.

1) Извлечение. На первом этапе данные извлекаются из исходной системы в область временного хранения данных - STAGING AREA (Рис. 2), которая предназначена для временного хранения данных, извлеченных из систем-источников. Данная область является промежуточным слоем между источником и хранилищем данных.

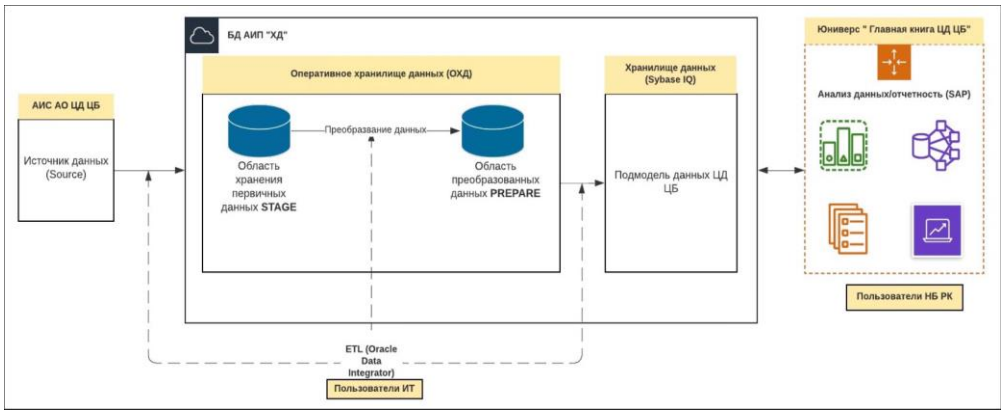

Рис. 2 - Извлечение данных в область временного хранения данных STAGE

2) Преобразование. Функция преобразования преобразует извлеченные данные в подходящий формат для анализа и хранения. Этот процесс включает преобразование извлеченных данных из их старой структуры в более денормализованный формат. Этот шаг зависит от конечной базы данных.

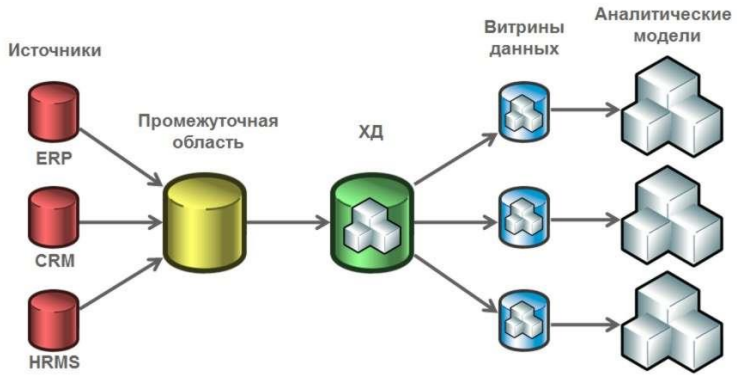

Рис. 3 - Преобразование данных, создание аналитических витрин (Data Mart)

Загрузка. Функция загрузки выполняет процесс записи преобразованных данных в базу данных. Это может занять несколько шагов, так как каждый этап может дополнять данные по-разному. Стандартная установка состоит в том, чтобы иметь необработанные, промежуточные и рабочие базы данных. Как правило, настраивается первоначальная загрузка всей информации с последующей периодической загрузкой добавочных измененных данных. [2].

ETL-инструменты

В качестве ETL-средства могут использоваться разные продукты. Один из них **Oracle Data Integrator**. Oracle Data Integrator (ODI) - это интеграционная платформа

корпоративного уровня, которая обеспечивает извлечение, преобразование и загрузку данных из разнообразных источников: баз данных, файлов и других источников.

- **АИП «ХД»** (Target). Хранилище данных (DWH).
- **Анализ данных** (SAP BO). Используя данные из АИП «ХД», производится настройка областей анализа, отчетности и витрин данных. В последствии пользователи вполне самостоятельно могут строить необходимую отчетность и проводить многомерный анализ. В качестве инструментов анализа используется **SAP Business Objects (SAP BO)**.
- **DMZ** (Demilitarized Zone - демилитаризованная зона) - технология обеспечения безопасности внутренней сети при предоставлении доступа внешних пользователей к определенным ресурсам внутренней сети (таким как почтовые, WWW-, FTP-серверы и др.).
- **API** (Application Programming Interface) - это программный посредник, который позволяет двум приложениям взаимодействовать друг с другом.
- **HTTP Basic Authentication** предоставляет механизм аутентификации. Это простая схема аутентификации, встроенная в протокол HTTP. Клиент отправляет HTTP-запросы, которые содержат слово Basic, и строку username:password.

Заключение. Современные корпорации требуют простого и быстрого доступа к данным. Это привело к растущему спросу на преобразование данных в самообслуживаемые системы. ETL играют жизненно важную роль в этой системе. Они обеспечивают аналитикам и специалистам по данным доступ к данным из нескольких прикладных систем. Это имеет огромное значение и позволяет компаниям получать новые идеи. В статье описана причина реализации ETL-системы и представлены ключевые моменты разработки. Результатом разработки данной системы является подготовка отчетов и бизнес-анализа из преобразованных и агрегированных данных хранилища данных с целью поддержки принятия решений в компании.

Список использованной литературы

1. Дэвид Тейлор, [Электронный ресурс] URL: https://www.guru99.com/etl-extract-load-process.html (дата обращения: 12.02.2022).

2. IBM Cloud Education, [Электронный ресурс] URL: https://www.ibm.com/cloud/learn/etl (Дата публикации: 28.04.2020).

3. Чжао, Ширли (2017-10-20). "Что такое ETL?". Качество данных Experian. (дата обращения: 12.12.2018).

4. Тревор Потт (4 июня 2018 года). "Извлекать, преобразовывать, загружать? Скорее, чрезвычайно трудно заряжать, амирит?". www.theregister.co.uk. (Дата публикации: 12.12.2018).

5. "ETL не мертв. Это по-прежнему имеет решающее значение для успеха бизнеса". Информация об интеграции данных. 8 июня 2020 года. (дата обращения: 14.07.2020).

References

1. David Taylor, [Electronic resource] URL: https://www.guru99.com/etl-extract-load-process.html (retrieved: 12.02.2022).

2. IBM Cloud Education, [Electronic resource] URL: https://www.ibm.com/cloud/learn/etl (retrieved: 28.04.2020).

3. Zhao, Shirley (2017-10-20). "What is ETL? (Extract, Transform, Load) | Experian". Experian Data Quality. (retrieved: 12.12.2018).

4. Trevor Pott (4 Jun 2018). "Extract, transform, load? More like extremely tough to load, amirite?". www.theregister.co.uk. (retrieved: 12.12.2018).

5. "ETL is Not Dead. It is Still Crucial for Business Success". Data Integration Info. 8 June 2020. (retrieved: 14.07.2020).