
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 004.6

DOI 10.5281/zenodo.11296083

ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ПРОЦЕССОВ ИНТЕГРАЦИИ ДАННЫХ ПОСРЕДСТВОМ ETL-ИНСТРУМЕНТОВ

INFORMATION SUPPORT OF PROCESSES DATA INTEGRATION THROUGH ETL-TOOLS

ГАНЕЕВ ТИМУР АЙБУЛАТОВИЧ,

ФГБОУ ВО «Российский государственный социальный университет».

КУЗНЕЦОВ АНДРЕЙ СЕРГЕЕВИЧ,

заместитель руководителя по научной деятельности,

кандидат технических наук, доцент,

ФГБОУ ВО «Российский государственный социальный университет».

GANEV TIMUR AIBULATOVICH,

Russian State Social University.

KUZNETSOV ANDREI SERGEEVICH,

Deputy Head for Scientific Activities,

Candidate of Technical Sciences, Associate Professor,

Russian State Social University.

В статье описываются процессы интеграции модели ETL с использованием методологии функционального моделирования систем IDEF0. Подробно рассматривается создание обобщенной и детализированной функциональной моделей процессов интеграции данных посредством ETL-инструментов. Проведено вербальное моделирование предметной области. На основе структурного анализа и системного подхода выделены составляющие функции – подпроцессы интеграции на уровне данных. Создано информационное функциональное описание процессов интеграции посредством ETL-инструментов. Приведен набор визуальных информационных моделей, формализующих и детализирующих процессы интеграции данных с применением модели ETL.

The article describes the processes of integrating an ETL-model using the IDEF0 system functional modeling methodology. The creation of generalized and detailed functional models of data integration processes using ETL-tools is discussed in detail. Verbal modeling of the subject area was carried out. Based on structural analysis and a systems approach, the constituent functions are identified - integration subprocesses at the data level. An informational functional description of integration processes using ETL-tools has been created. A set of visual information models is presented that formalize and detail the processes of data integration using the ETL-model.

Ключевые слова: интеграция ETL, IDEF0-методология, информационное обеспечение, модель интеграции.

Key words: *integration of ETL, IDEF methodology, information support, integration model.*

Введение.

В условиях растущего объема информации организации сталкиваются с проблемой объединения данных из разнородных источников для обеспечения целостности, доступности и анализа информации, что непосредственно влияет на принятие обоснованных управленческих решений. Традиционно, решением этой задачи являются ETL-процессы, предоставляющие структурированный подход к интеграции данных.

ETL – это аббревиатура, которая означает извлечение (Extract), трансформация (Transform) и загрузка данных (Load). В этом процессе данные берутся из первой системы, преобразуются в формат, подходящий к протоколам второй системы, и выгружаются в базу данных второй системы [1].

Основная часть.

1. Извлечение данных.

Процесс начинается с интеграции данных. Этот этап аналогичен действию "Extract" в рамках процедуры ETL, однако, если смотреть на процесс изнутри системы, на этом этапе идет загрузка данных, а не их извлечение. Материалы, поступающие в рамках данного процесса, представляют собой необработанный массив – он предстоит к анализу, без какой-либо гарантии качества. Одной из основных проверок на данном этапе является подтверждение количества элементов. Если число элементов оказывается меньше ожидаемого, это свидетельствует о неполадках в процессе сбора данных [2].

Следующий шаг – это валидация данных. В границах действия "Extract" данные подвергаются тщательной проверке, что и называется валидацией: загруженная информация проходит через фильтры проверки и отсева. Система оценивает объемность информации, её правильность и отсутствие ошибок, речь идет о "испорченных" данных. По завершении валидации система выдает сводку об ошибках. Обнаруженные несоответствия подлежат корректировке.

Иногда применяется процедура обогащения данных – это когда на основании уже существующих сведений система формирует новые. Это актуально, когда в распоряжении системы имеются иные внутренние ресурсы для обработки. В отдельных случаях такие данные могут быть сгенерированы при помощи алгоритмов вычислений.

Этот этап подразумевает извлечение данных из первичных систем, которые могут быть очень разнообразными; такие, как базы данных, CRM-системы, файлы Ms Excel и многое другое. Полученные данные часто неструктурированные и несовместимые между собой, поэтому они подвергаются процессу преобразования.

2. Трансформация данных.

Маппинг – ключевая ступень процесса "Transform", на которой информация переформатируется для соответствия новым стандартам. После валидации информация обычно принимает форму таблицы, к которой добавляются дополнительные строки и столбцы, что может выполняться по различным алгоритмам в зависимости от того, какие инструменты и настройки ETL используются, включая специальные скрипты и формулы системы [3].

Агрегация – это еще один сегмент трансформации, где данные оптимизируются под нужды детализации и представления. Благодаря процессу агрегации информация компилируется в новую таблицу, где все элементы упорядочены в соответствии с требованиями нового места хранения.

На этапе трансформации данные очищаются от ошибок, неполных записей, дубликатов. Они также приводятся к однородному формату, чтобы данные были согласованными и точными.

3. Загрузка данных.

После трансформации данные готовы к последнему этапу – загрузке в конечную систему. Это может быть хранилище данных, где данные будут храниться в надежном и легко доступном виде для аналитики и отчётности.

Завершающий этап – "Load", на котором отфильтрованные и преобразованные данные покидают системы ETL и направляются в новое место хранения. Здесь используются разработанные специально для этих целей средства, включая разнообразные коннекторы и интерфейсные элементы систем хранения данных [4].

4. Функциональная модель процесса интеграции данных.

Модель функциональных процессов, выраженная через нотацию IDEF0, является воплощением упорядоченного и системного подхода к описанию интегрирования информационных технологий. Методология IDEF0 служит инструментом для детальной визуализации функций комплексного взаимодействия систем, охватывая аспекты данных, которые система получает (входы) и выдает (выходы), механизмы и управление [5].

Эта модель помогает глубже понять процесс интеграции систем, выделять ключевые этапы процесса и выявлять основные компоненты и взаимосвязи между ними. Такой подход помогает более эффективно управлять процессом интеграции, а также создать основу для анализа и оптимизации процесса.

Ниже представлена функциональная модель для интеграции ETL.

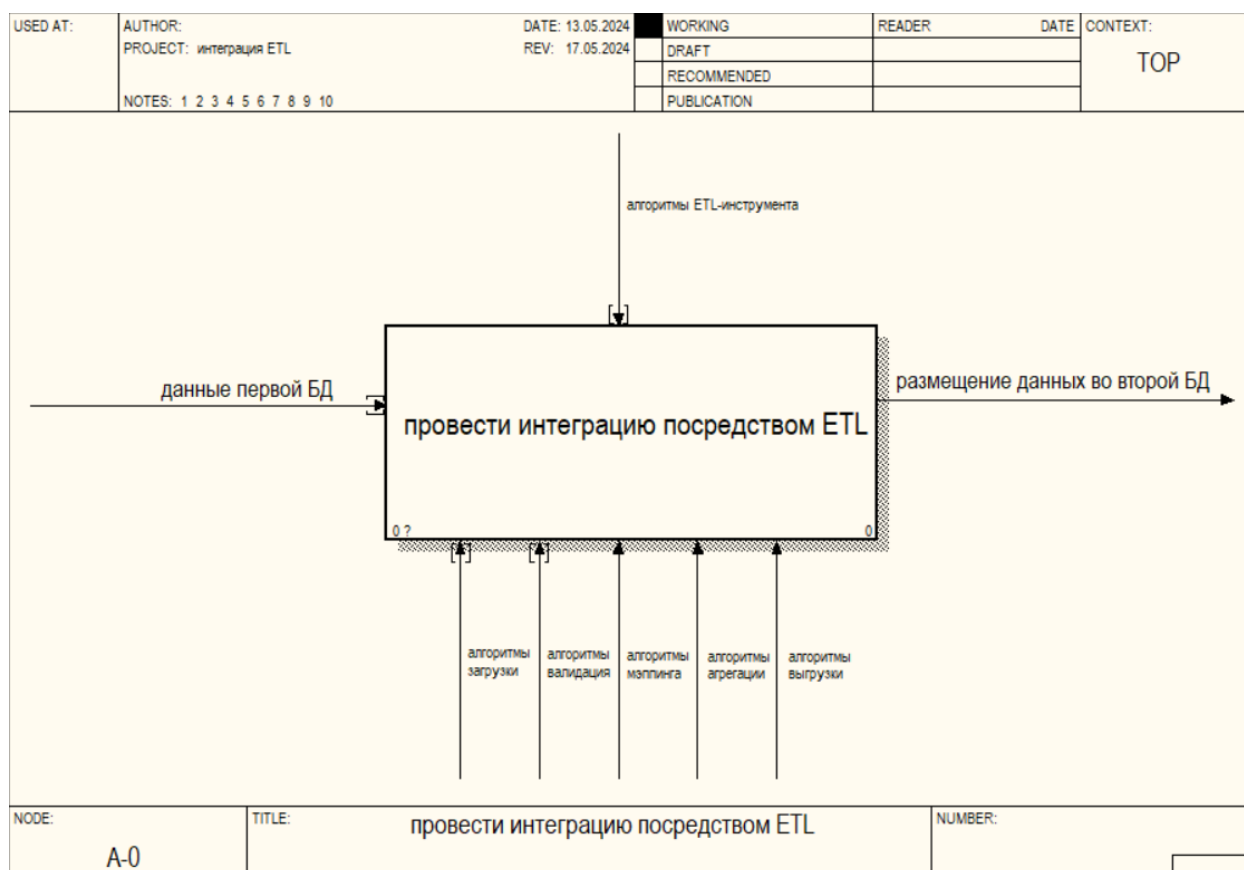


Рис. 1. Обобщенная функциональная модель ETL-процесса

Обобщенная функциональная модель состоит из процесса проведения интеграции посредством ETL, механизмы осуществления процесса состоят из алгоритмов загрузки,

валидации, маппинга, агрегации, выгрузки. Управление происходит с помощью различных алгоритмов ETL-инструмента. Вход представляет собой данные, извлеченные из первой базы данных. Выход – размещение данных во второй БД [6].

Рассмотрим детализированную функциональную модель (рис. 2).

В отличие от обобщенной функциональной модели, детализированная функциональная модель состоит из трех процессов [7]. Первый процесс – извлечение данных, механизмы: алгоритмы загрузки и валидации. Второй процесс – преобразование данных, механизмы: алгоритмы маппинга и агрегации. Третий этап – проведение загрузки данных во вторую базу данных, механизм: алгоритм загрузки. Детализированная функциональная модель помогает подробнее описать процесс интеграции с помощью ETL-инструментов.

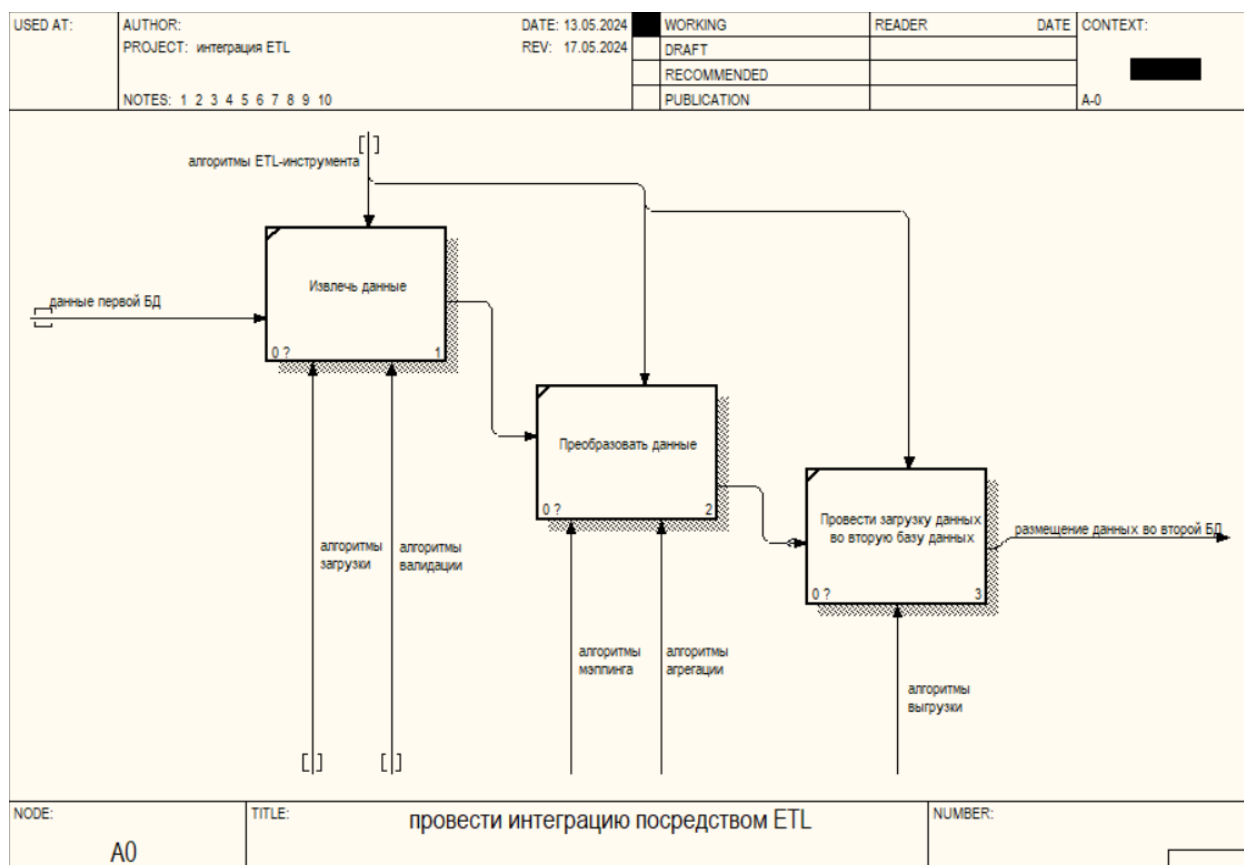


Рис. 2. Детализированная функциональная модель ETL

Заключение.

Интеграция данных через ETL – это комплексный процесс, требующий детального понимания как исходных данных, так и целей их использования. Правильно спланированный и выполненный ETL-процесс является ключевым фактором успеха в управлении данными и последующем анализе.

Работа придерживается нотации IDEF0. IDEF0 – методика, предназначенная для воплощения в модельном изображении схемы и операций системы, в дополнение к визуализации потоков данных и материальных элементов, которые обеспечивают взаимосвязь между операциями.

Использование стандартов и методологий, таких как IDEF0, играет важную роль при анализе и разработке информационных систем, поскольку помогает организовать процессы,

выявлять уязвимые места и оптимизировать работу над проектом. Эти модели могут служить основой для последующей разработки, внедрения и управления проектами интеграции систем, что способствует повышению эффективности и успешности проектов в области информационных технологий.

СПИСОК ЛИТЕРАТУРЫ

1. Егорян В.В., Калугин А.В. Виды интеграции программных приложений // Столыпинский вестник. 2022. № 9. 9 с.
2. Громов Н.Д., Платошин А.И., Панов А.В. Сравнительный анализ средств и платформ для автоматизации ETL процессов в современных хранилищах данных // Международный журнал гуманитарных и естественных наук. 2023. № 11-4 (86). С. 46-48.
3. Использование ETL-процессов для анализа данных. URL: <https://habr.com/ru/companies/ot-us/articles/731224> (дата обращения: 16.05.2024).
4. Абдираимов Н.А., Алимжанова Л.М. Использование ETL-процесса для создания хранилища данных в банках второго уровня // Интеллектуальные технологии на транспорте. 2023. № S1 (35-1). С. 10-13.
5. Кольева Н.С., Кожахметова Р.Н. Проектирование информационного сопровождения совершенствования компетенций // НИР/S&R. 2021. № 1 (5). С. 34-37.
6. Кузнецов А.С. Информационное обеспечение процессов управления качеством продукции из многокомпонентных эластомерных композитов // Известия Кабардино-Балкарского научного центра РАН. 2024. № 1 (117). С. 21-28.
7. Информационно-алгоритмическое обеспечение имитационной системы управления экспериментальными исследованиями процессов смешения и структурирования эластомерных композитов для дополнительного обучения инженеров-технологов / А.С. Кузнецов, В.Ф. Корнюшко, Е.Г. Шмакова, С.В. Пивнева, В.В. Халюкин // Современные информационные технологии и ИТ-образование. 2019. Т. 15. № 4. С. 837-845.

© Ганеев Т.А., Кузнецов А.С., 2024.