



Список литературы

1. Дрозд Г. Я. Трубопроводы систем жизнеобеспечения ЖКХ как объекты пристального внимания / Г. Я. Дрозд, М. Ю. Хвортова // Сборник научных трудов Донбасского государственного технического университета. – 2016. – № 4 (47). – С. 122–140.
2. Хецуриани Е. Д. Конструктивно-технологическая система обеспечения экологической безопасности водозаборных технологических комплексов / Е. Д. Хецуриани, В. Л. Бондаренко, А. И. Ылясов, Т. Е. Хецуриани // Инженерно-строительный вестник Прикаспия. – 2020. – № 1 (31). – С. 55–60.
3. Бочаров сообщил о крупнейшей за 10 лет техногенной аварии в Волгограде // RBC.ru. – Режим доступа: <https://www.rbc.ru/society/31/10/2022/635f759f9a794758ea609e9a> (дата обращения 31.08.2022), свободный. – Заглавие с экрана. – Яз. рус.
4. Последствия аварии на канализационном коллекторе в пойме реки Царицы в Волгограде // Новости Волгограда. – Режим доступа: <https://novostivolgograda.ru/news/2022-10-31/cherez-vodu-i-truby-prihoditsya-prohodyat-volgogradtsy-iz-za-chp-na-kollektore-2413092> (дата обращения 31.08.2022), свободный. – Заглавие с экрана. – Яз. рус.
5. Ковиева Е. Обновили 26 метров трубы: когда вода вернется в дома волгоградцев? // Волгоград ТРВ. – Режим доступа: <https://volgograd-trv.ru/news/obschestvo/71131-obnovili-26-metrov-truby-kogda-voda-vernetsja-v-doma-volgogradcev.html> (дата обращения 31.08.2022), свободный. – Заглавие с экрана. – Яз. рус.
6. Dmitrieva E. Digitalization of the housing and communal services: Development prospects / E. Dmitrieva // E3S Web of Conferences : 24, Moscow, 22–24 апреля 2021 г. – Moscow, 2021. – DOI 10.1051/e3sconf/202126304037.
7. Ganchenko D. N. Digitalization of housing and communal services in the context of new industrialization / D. N. Ganchenko, A. A. Bodrov // Modern Management Trends and the Digital Economy: from Regional Development to Global Economic Growth : Proceedings of the 1st International Scientific Conference, Yekaterinburg, 14–15 апреля 2019 г. – Yekaterinburg : Atlantis Press, 2019.
8. Мицеловская О. С. Цифровые решения как факт оления эффективности функционирования сферы жилищно-коммунального хозяйства / О. С. Мицеловская // Научные труды Северо-Западного института управления РАНХиГС. – 2019. – Т. 10, № 3 (40). – С. 161–164.
9. Корнева А. В. Внедрение цифровых решений в ЖКХ / А. В. Корнева // Вестник связи. – 2021. – № 6. – С. 34–39.
10. Кулагина Е. И. Тенденции решения современных проблем ЖКХ при помощи цифровой трансформации в государственном управлении / Е. И. Кулагина // Цифровые трансформации в развитии экономики и общества : материалы XVIII Международной научно-практической конференции, Липецк, 22 апреля 2021 г. : в 4 т. / под общ. ред. А. Д. Моисеева. – Воронеж: НАУКА-ЮНИПРЕСС, 2021. – С. 258–264.
11. Стратегия развития строительного комплекса и жилищно-коммунального хозяйства Российской Федерации на период до 2030 года с прогнозом до 2035 года. – Утверждена распоряжением Правительства Российской Федерации от 31 октября 2022 г. № 3268-р. – Режим доступа: <http://static.government.ru/media/files/AdmXczBBUGfGNM8tz16r7RkQcsgP3LAm.pdf>, свободный. – Заглавие с экрана. – Яз. рус.

© Нань Фэн, З. А. Агаларов, О. М. Шиккульская

Ссылка для цитирования:

Нань Фэн, Агаларов З. А., Шиккульская О. М. Системный анализ организации и проведения ремонта трубопроводных систем в рамках цифровизации ЖКХ // Инженерно-строительный вестник Прикаспия : научно-технический журнал / Астраханский государственный архитектурно-строительный университет. Астрахань : ГАОУ АО ВО «АГАСУ», 2022. № 4 (42). С. 104–108.

УДК 004.89

DOI 10.52684/2312-3702-2022-42-4-108-113

ПРИМЕНЕНИЕ ЕТЛ-ПРОЦЕССОВ ДЛЯ АВТОМАТИЗАЦИИ АНАЛИЗА ДАННЫХ ПОРОЗНИЧНЫМ ПРОДАЖАМ

М. И. Шиккульский, О. В. Медведева, В. М. Баркова, Л. А. Плешакова

Шиккульский Михаил Игорьевич, кандидат технических наук, доцент кафедры «Прикладная информатика», Астраханский государственный технический университет, г. Астрахань, Российская Федерация, тел.: +7-917-171-31-09; e-mail: shikul_m@mail.ru;

Медведева Ольга Викторовна, магистрант, Астраханский государственный архитектурно-строительный университет, г. Астрахань, Российская Федерация, тел.: +7-917-170-15-53; e-mail: lettmail@yandex.ru;

Баркова Виктория Михайловна, студент, Астраханский государственный технический университет, г. Астрахань, Российская Федерация, тел.: +7-967-823-47-30; e-mail: barkova-vika@mail.ru

Плешакова Людмила Александровна, кандидат технических наук, доцент кафедры «Прикладная информатика», Астраханский государственный технический университет, г. Астрахань, Российская Федерация, тел.: +79371214834; e-mail: lpleshakova@rambler.ru

Недостаток анализа данных может привести к убыткам торгового предприятия. В статье рассматриваются основные этапы интеллектуального анализа данных, представленные в виде модели в нотации IDEF0. Отмечено, что наименее изученными являются задачи ETL-процесса (извлечение, преобразование, загрузка информации в хранилище данных аналитической платформы). Для описания ETL-процесса спроектирована и описана структура OLAP-хранилища, разработан алгоритм преобразования данных для предприятия розничной торговли. Особенность алгоритма связана с загрузкой в хранилище данных информации из множества Excel-файлов со схожей структурой за разные периоды времени, необходимостью преобразования их к плоской структуре и формированию вычисляемых полей, отсутствующих в исходных данных. Идеи предложенного подхода могут быть использованы для решения аналогичных задач в других предметных областях.

Ключевые слова: интеллектуальный анализ данных, хранилище данных, OLAP-хранилище, ETL-процесс, аналитическая платформа, DataMining, розничные продажи, проектирование информационных систем, Data Warehouse.

APPLICATION OF ETL PROCESSES FOR AUTOMATION OF RETAIL SALES DATA ANALYSIS

M. I. Shikulskiy, O. V. Medvedeva, V. M. Barkova, L. A. Pleshakova

Shikulsky Mikhail Igorevich, Candidate of Technical Sciences, Associate Professor the Department of Applied Informatics, Astrakhan State Technical University, Astrakhan, Russian Federation, phone: +7-917-171-31-09; e-mail: shikul_m@mail.ru

Medvedeva Olga Viktorovna, Master, Astrakhan State University of Architecture and Civil Engineering, Astrakhan, Russian Federation, phone: +7-917-170-15-53; e-mail: lettmail@yandex.ru;

Barkova Viktoriya Mikhaylovna, Student, Astrakhan State Technical University, Astrakhan, Russian Federation, phone: +7-967-823-47-30; e-mail: barkova-vika@mail.ru;

Pleshakova Lyudmila Aleksandrovna, Candidate of Technical Sciences, Associate Professor of the Department of Applied Informatics, Astrakhan State Technical University, Astrakhan, Russian Federation, phone: +7-937-121-48-34; e-mail: lpleshakova@rambler

A lack of data analysis can lead to losses for a trading company. The article discusses the main stages of data mining, presented in the form of a model in the IDEF0 notation. It is noted that the least developed of the considered stages are the tasks of the ETL process (extraction, transformation, loading of information into the data warehouse of the analytical platform). To describe the ETL process, the structure of the OLAP storage is designed and described, and an algorithm for data transformation for a retail company is developed. The peculiarity of the algorithm is associated with loading information from a variety of Excel files with a similar structure into the data warehouse for different periods of time, the need to convert data to a flat structure and the formation of calculated fields that are missing from the source data. The ideas of the proposed approach can be used to solve similar problems in other subject areas.

Keywords: data mining, data warehouse, OLAP storage, ETL process, analytical platform, Data Mining, retail sales, information systems design, Data Warehouse.

Экономическая ситуация в нашей стране, сложившаяся в последнее время, усложняет процесс ведения бизнеса, требует от руководства розничных магазинов постоянного мониторинга и своевременного реагирования на сформировавшиеся тенденции. Любая ошибка планирования для предприятия торговли может оказаться фатальной. В то же время своевременные решения позволят компании выделиться на фоне конкурентов и занять лидирующее положение в отрасли.

В значительной степени помочь руководству предприятия розничной торговли в выборе решений может использование методов интеллектуального анализа данных.

На текущий момент задачи интеллектуального анализа находятся на пике своей популярности. Существует множество методов и аналитических платформ, которые позволяют упростить и ускорить анализ данных для обеспечения возможности более качественного, обоснованного оперативного и стратегического планирования деятельности предприятия [1].

Основные этапы интеллектуального анализа данных авторами были описаны в виде модели в нотации IDEF0, представленной на рисунке 1.

Исходные данные для анализа обычно берутся из OLTP-системы (Online Transaction Processing). В базе данных OLTP-системы хранятся сведения об отдельных продажах, поставках товаров магазина и прочей оперативной деятельности.

Для анализа данных имеющаяся информация должна быть преобразована с помощью ETL-процессов (от англ. Extract, Transform, Load – извлечение, преобразование, загрузка) в формат, удобный для их обработки с помощью аналитической платформы [2].

На втором этапе осуществляется интеллектуальный анализ преобразованных данных с помощью различных методов DataMining, таких как нейронные сети, кластерный и регрессионный анализ, поиск ассоциативных правил и т. д. Применение перечисленных методов позволит определить группы совместно покупаемых товаров, спрогнозировать спрос на продукцию, выявить ключевые характеристики товаров и покупателей, влияющих на прибыль компании и в наибольшей степени соответствующих текущим рыночным условиям, обнаружить скрытые закономерности, не представленные явно в исходных данных [3].

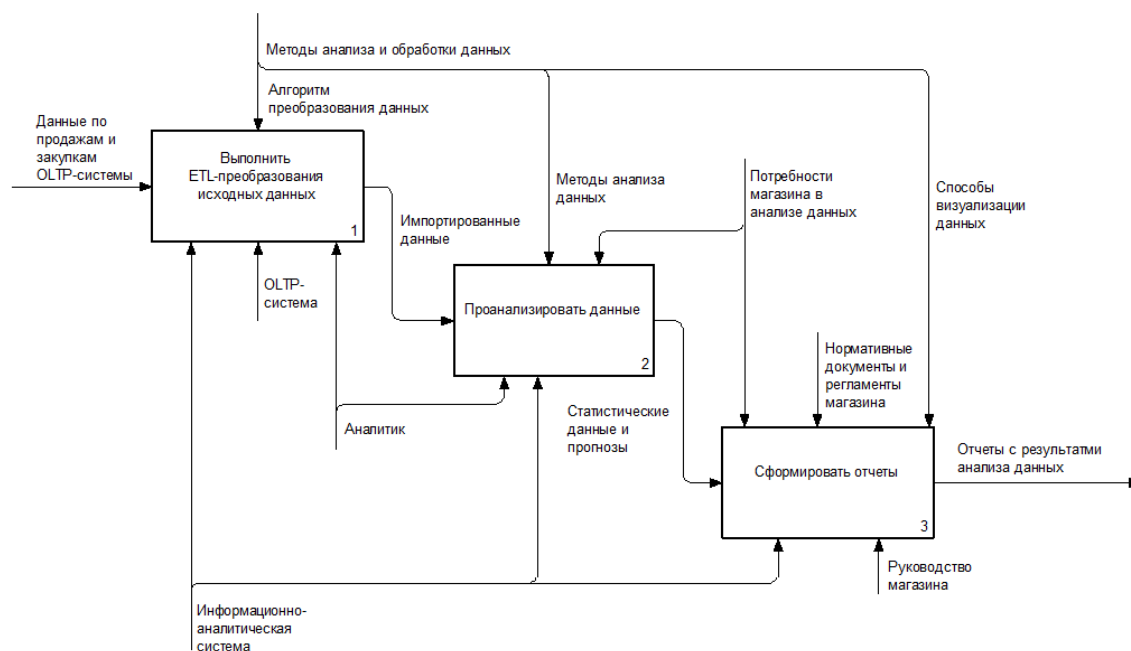


Рис. 1. Этапы интеллектуального анализа данных

Для того, чтобы руководство магазина смогло разобраться в результатах анализа и принять на их основе взвешенное решение, выполняется третий этап – визуализация данных. Визуализированные данные могут быть представлены в графическом и/или табличном виде, отфильтрованы и отсортированы по различным признакам в зависимости от целей моделирования и ключевых потребностей магазина.

Решению задач интеллектуального анализа данных для розничных магазинов посвящено множество работ [4–14]. В данных исследованиях в основном предлагаются решения для адаптации известных методов Data Mining к потребностям розничной торговли. Для визуализации данных также имеется большое разнообразие инструментов, включенных в состав предлагаемых на рынке аналитических платформ. При этом задачам ETL-процессов в рассмотренных исследованиях уделено незаслуженно мало внимания. Возможно, это упущение обусловлено тем, что на извлечение, преобразование и загрузку данных оказывают существенное влияние технические характеристики решаемых задач, связанные с особенностями среды хранения данных в OLTP-системах, возможностями обработки данных в аналитической платформе, спецификой решаемой задачи [15]. Поэтому авторы статьи углубились в решение задачи ETL-процесса для частной предметной области – анализа данных о закупках и продажах товаров одного из магазинов «Светофор» компании «Торгсервис-30».

Магазины «Светофор» представляют собой розничные склады-магазины самообслуживания, работающие в режиме жесткого дискаунтера. Цель данной торговой сети – обеспечить

население качественными товарами по самым низким ценам. Для учета продаж и закупок товаров в магазинах «Светофор» используется OLTP-система «1С:Управление торговлей 8.3».

В качестве аналитической платформы для разрабатываемой информационно-аналитической системы была выбрана программа Loginom, предоставляющая широкие возможности для анализа и визуализации данных.

С учетом специфики решаемых в информационно-аналитической системе задач была спроектирована структура OLAP-хранилища, позволяющая отслеживать статистические данные по основным измерениям, необходимым для анализа данных (рис. 2).

Основные статистические данные по решаемым задачам размещаются в центральных таблицах фактов (называемых также процессами). Такие таблицы обозначены символом . Измерения, по которым проводится анализ, содержат справочные данные и изображены на рисунке 2 с символами и в зависимости от уровня агрегации [16].

В спроектированной OLAP-структуре представлено три процесса («Продажи», «Поступления», «Товарные запасы») и шесть измерений («Торговая точка», «Тип операции», «Документ оприходования», «Номенклатура», «Категория продукта», «Контрагент»).

Таким образом, перед авторами стояла задача разработать алгоритм ETL-процесса, преобразующего Excel-отчеты системы «1С:Управление торговлей 8.3» в таблицы аналитической платформы Loginom, соответствующие структуре, представленной на рисунке 2.

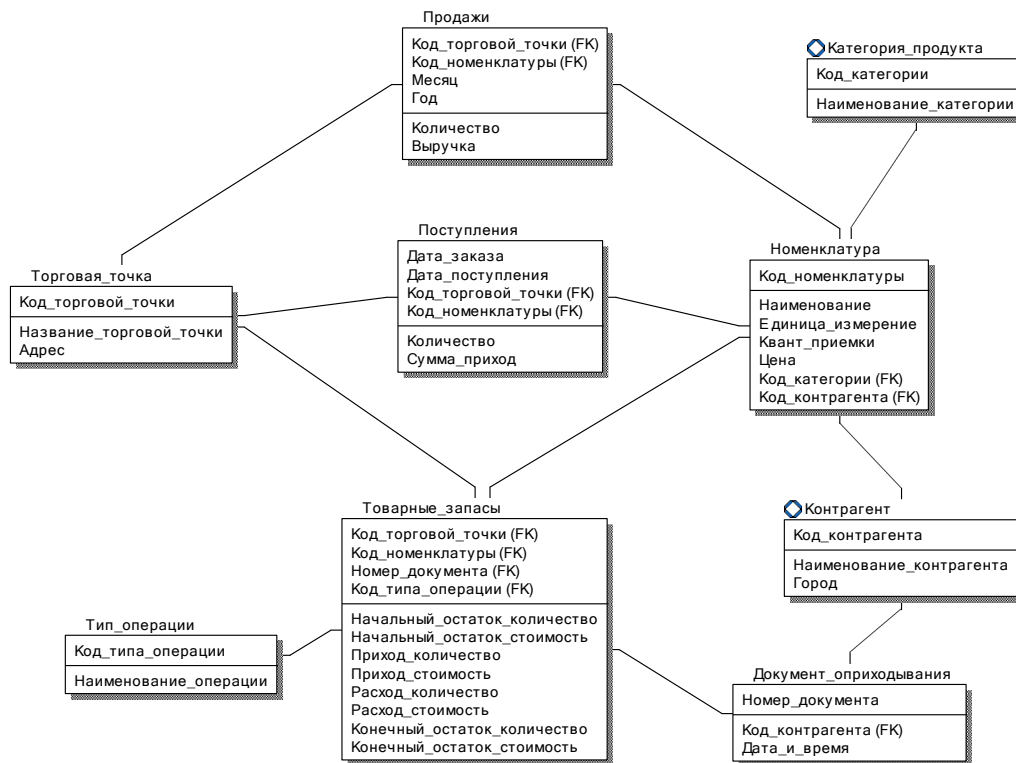


Рис. 2. Структура OLAP-хранилища для розничного магазина

Основные проблемы преобразования данных для описываемой задачи были связаны с тем, что из-за ограничения доступа к системе «1С:Управление торговлей» руководство магазина может выгружать данные по продажам, покупкам и остаткам товаров только в отдельные Excel-файлы по месяцам. Это ограничивает возможности по импорту данных.

Вторая особенность импортируемых файлов – наличие группировки в стандартных отчетах 1С. Сгруппированные данные должны быть преобразованы в плоские таблицы для возможности их последующей обработки в Loginom.

Третий аспект, усложняющий импорт, состоит в отсутствии в исходных таблицах выделенных полей с данными по отдельным атрибутам, необходимым для анализа (таким, как дата сделки, категории товаров и т. д.).

Все эти сложности требуют дополнительного преобразования данных. В связи с этим был разработан алгоритм и сценарий для трансформации данных, представленный в виде диаграммы деятельности на рисунке 3.

Перед импортом данных пользователь может задать рабочую директорию и тип импортируемого файла. При желании пользователь может оставить значения по умолчанию.

Процесс преобразования импортируемых файлов отличается в зависимости от их типа. Здесь можно выделить две группы загружаемых файлов: справочники (номенклатуры, категории товаров, контрагенты) и процессы (продажи, поступления, спрос).

Данные по справочникам представлены в едином файле, поэтому для их загрузки в информационно-аналитическую систему должны быть выполнены следующие действия:

- преобразование сгруппированной Excel-таблицы в плоскую таблицу с помощью специально разработанных для этих целей VBA-макросов (рис. 4);
- импорт содержимого справочников в Loginom;
- при необходимости вычисление показателей справочников, не представленных в исходных данных, таких как идентификаторы справочников.

Вторая группа импортируемых данных по процессам представлена в виде совокупности файлов – по одному файлу на каждый месяц, поэтому перед импортом файлов формируется их список из рабочей директории с помощью Java-скрипта. Далее в цикле подготавливается каждый файл рабочей директории с помощью групповой обработки.

В теле цикла выполняются следующие преобразования:

- сгруппированные данные Excel-файла преобразуются в плоскую таблицу (с помощью VBA-макросов по аналогии с преобразованием для справочников);
- трансформированные данные файла импортируются в Loginom;
- при импорте добавляются колонки с номером итерации цикла и названием файла для последующей идентификации импортированного набора данных;
- на основании имени импортированного файла определяется период загружаемых данных

(месяц и год), данные параметры также добавляются в отдельные колонки выходной таблицы; при необходимости с помощью калькулятора вычисляются дополнительные атрибуты

(значения внешних ключей для связывания таблиц, даты и т. д.).

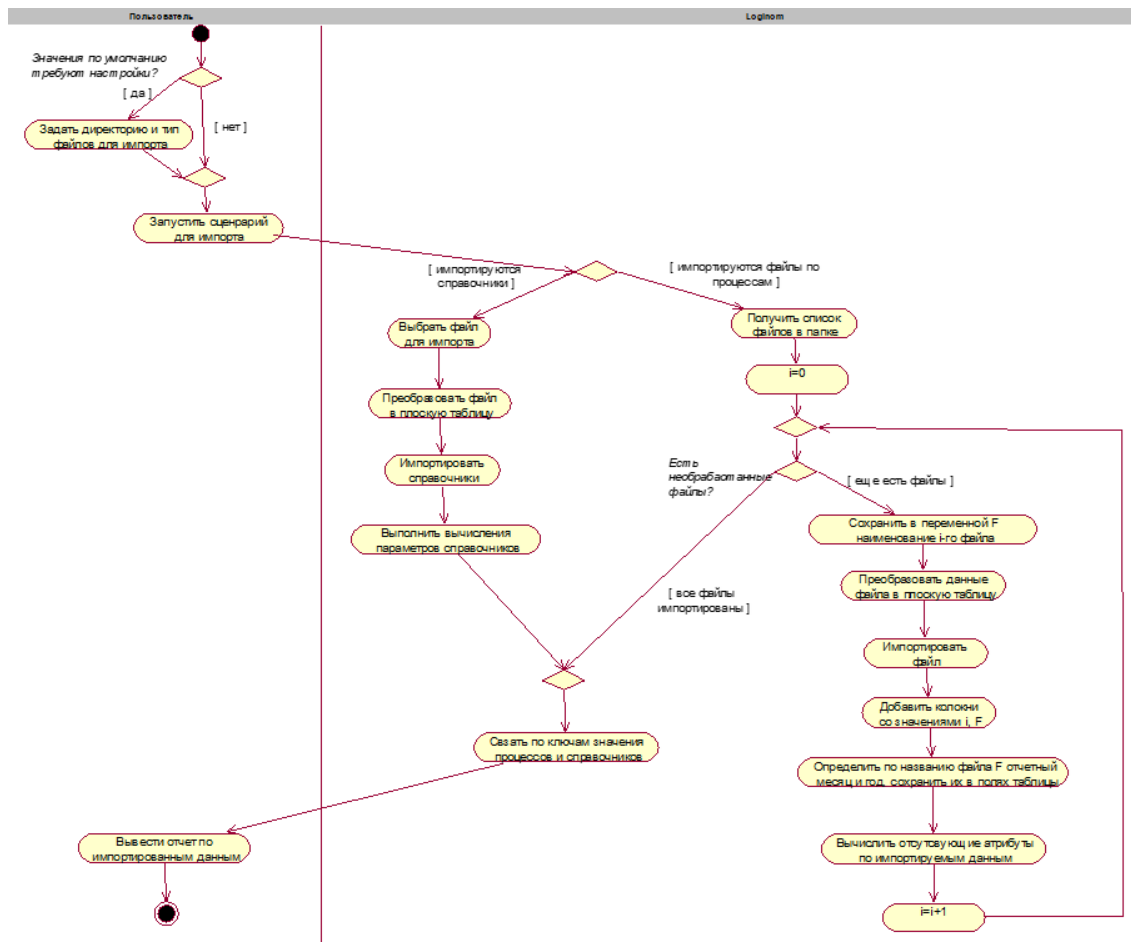


Рис. 3. Алгоритм ETL-процесса

1	2	3
7	Поставщик	
8	Дата поступления	Дата поступления
9	Номенклатура	
11	АВИКОМП ТД ООО	
12	Мешки мусорные, рулон 240л/15 шт., черный 18 мм	30.08.2022
14	АВР ГРУПП ООО	
15	Игровой набор продуктов в сетке 20 предметов. Размер 20x20см	
16	Настольная игра Хоккей/Футбол в коробе 58x33x7см	
17	Игровой набор посуды 23 предмета. Размер 18x18 см	24.08.2022
18	Доска знаний мольберт 730x500x70 арт. ДЗМ1	24.08.2022
19	Машинка детская "Самосвал". Размер 29x16 см	24.08.2022
21	Агро-Юг ООО	

1	2	3	4
1	Поставщик	Номенклатура	Дата пост Количест
2	АВР ГРУПП ООО	Игровой набор продуктов в сет	345
3	АВР ГРУПП ООО	Настольная игра Хоккей/Футбол	345
4	АВР ГРУПП ООО	Игровой набор посуды 24.08.2022	135
5	АВР ГРУПП ООО	Доска знаний мольберт 24.08.2022	48
6	АВР ГРУПП ООО	Машинка детская "Са 24.08.2022	162
7	Агро-Юг ООО	Филе пангасиуса 220 24.08.2022	507,39

До преобразования

После преобразования

Рис. 4. Пример преобразования данных в плоскую таблицу

Все импортированные данные по процессу сливаются в единую таблицу. Справочники и процессы также связываются по внешним ключам. Результаты импорта каждой таблицы может просмотреть пользователь для их оценки и корректировки сценария.

Благодаря преобразованию и импорту данных из OLTP-системы в таблицы Loginom аналитик сможет разрабатывать сценарии для

анализа данных, используя все преимущества аналитической платформы, а руководство магазина – изучать аналитические отчеты и принимать обоснованные решения, направленные на дальнейшее развитие компании.

Внедрение информационно-аналитической системы позволит улучшить ассортиментную политику торгового предприятия, повысить прибыль компании, удовлетворить запросы потребителей.

Список литературы

1. Луньков А. Д. Интеллектуальный анализ данных / А. Д. Луньков, А. В. Харламов. – Режим доступа: elibrary.sgu.ru/uch_lit/1141.pdf, свободный. – Заглавие с экрана. – Яз. рус.
2. Википедия. – Режим доступа: <https://ru.wikipedia.org/wiki/ETL/>, свободный. – Заглавие с экрана. – Яз. рус.

3. Основные функции ETL-систем // Хабрахабр. – Режим доступа: <https://habrahabr.ru/post/248231/>, свободный. – Заглавие с экрана. – Яз. рус.
4. Щепакин М. Б. Факторная модель управления развитием рынка оптовой и розничной торговли в национальной экономике / М. Б. Щепакин, М. В. Облогин, В. М. Михайлова // Экономика, предпринимательство и право. – 2020. – Т. 10, № 4. – С. 1095–1122. – DOI: 10.18334/erpp.10.4.100650.
5. Алексейчик Т. В. Анализ финансового состояния предприятия оптово-розничной торговли с использованием математических методов / Т. В. Алексейчик, А. С. Стасюк, Т. В. Богачев, О. В. Домакур // Информатизация в цифровой экономике. – 2022. – Т. 3, № 2. – DOI: 10.18334/ide.3.2.115225.
6. Дзюба А. П. Исследование инструментов управления спросом на электропотребление в рамках оптового и розничного рынков электроэнергии России / А. П. Дзюба, И. А. Соловьева // Экономика, предпринимательство и право. – 2016. – Т. 6, № 2. – С. 147–162. – DOI: 10.18334/erpp.6.2.35309.
7. Степанова Е. Н. Организация продаж в розничной торговле: современные тенденции / Е. Н. Степанова // Вестник Хабаровского государственного университета экономики и права. – 2021. – № 2 (106). – С. 47–53. – DOI 10.38161/2618-9526-2021-2-47-53. – EDN JQUYZO.
8. Агаева А. Н. Анализ тенденций развития маркетинговых исследований регионального рынка розничной торговли / А. Н. Агаева, Т. З. Васильченко // Фундаментальные исследования. – 2013. – № 11–5. – С. 970–974. – Режим доступа: <https://fundamental-research.ru/ru/article/view?id=33235>, свободный. – Заглавие с экрана. – Яз. рус.
9. Петраковский С. Г. Анализ особенностей рынка секторов розничных продаж / С. Г. Петраковский, А. Г. Петраковский, Ю. В. Гуняков, Т. В. Дрокина // Российское предпринимательство. – 2008. – № 2–1. – Режим доступа: <https://cyberleninka.ru/article/n/analiz-osobennostey-elektronnogo-rynka-sektora-rozничnyh-prodazh>, свободный. – Заглавие с экрана. – Яз. рус.
10. Моисеев Б. В. Анализ факторов, влияющих на эффективность продаж в розничной торговле, и способы управления ими / Б. В. Моисеев, А. А. Ветрова // Прикаспийский журнал: управление и высокие технологии. – 2012. – № 4. – Режим доступа: <https://cyberleninka.ru/article/n/analiz-faktorov-vliyayuschih-na-effektivnost-prodazh-v-rozничnoy-torgovle-i-sposoby-upravleniya-imi>, свободный. – Заглавие с экрана. – Яз. рус.
11. Ребенок И. И. Методы интеллектуального анализа и прогнозирования данных стационарной розничной торговой сети / И. И. Ребенок, М. П. Малыгина // Современные проблемы науки и образования. – 2014. – № 3. – Режим доступа: <https://science-education.ru/ru/article/view?id=13761>, свободный. – Заглавие с экрана. – Яз. рус.
12. Урумбаева О. Б. Концепция интеллектуального управления энергосетью / О. Б. Урумбаева, Т. А. Шалаев, О. М. Шикунская // Инженерно-строительный вестник Прикаспия. – 2020. – № 3 (33). – С. 69–74.
13. Нечипорук Д. В. Особенности технологии Data Mining / Д. В. Нечипорук // Молодой исследователь Дона. – 2017. – № 1 (4). – Режим доступа: <https://cyberleninka.ru/article/n/osobennosti-tehnologii-data-mining>, свободный. – Заглавие с экрана. – Яз. рус.
14. Баева В. Р. ETL: актуальность и применение. Преимущества и недостатки etl инструментов / В. Р. Баева, А. Ю. Дроздов // Вестник науки. – 2019. – № 5 (14). – Режим доступа: <https://cyberleninka.ru/article/n/etl-aktualnost-i-primenenie-preimuschestva-i-nedostatki-etl-instrumentov>, свободный. – Заглавие с экрана. – Яз. рус.
15. Проектирование баз данных. Распределенные базы и хранилища данных. Агрегирование. – Режим доступа: http://www.intuit.ru/studies/professional_retraining/953/courses/214/lecture/5508/, свободный. – Заглавие с экрана. – Яз. рус.

© М. И. Шикунский, О. В. Медведева, В. М. Баркова, Л. А. Плешакова

Ссылка для цитирования:

Шикунский М. И., Медведева О. В., Баркова В. М., Плешакова Л. А. Применение etl-процессов для автоматизации анализа данных по розничным продажам // Инженерно-строительный вестник Прикаспия : научно-технический журнал / Астраханский государственный архитектурно-строительный университет. Астрахань : ГАОУ АО ВО «АГАСУ», 2022. № 4 (42). С. 108–113.