

# ОБ ОБЩЕЙ АРХИТЕКТУРЕ СИСТЕМЫ ХРАНЕНИЯ И ОБРАБОТКИ ДАННЫХ ПРАКТИКО-ОРИЕНТИРОВАННОЙ НАПРАВЛЕННОСТИ

Рудикова Л. В., Жавнерко Е. В.

Кафедра современных технологий программирования, Гродненский государственный университет имени Янки Купалы

Гродно, Республика Беларусь

E-mail: {rudikowa, eugene.zhavnerko}@gmail.com

В статье приводится общая архитектура системы накопления и анализа данных практико-ориентированной направленности, основанная на технологии складирования данных. Отличительной чертой при разработке системы является подход с использованием расширяемого хранилища данных на основе универсальной платформы хранимых данных, предоставляющей доступ для хранения и последующего анализа данных различной структуры и различных предметных областей. Приводятся общая архитектура универсальной системы хранения и обработки данных указанной направленности, выделяются структурные составляющие.

## ВВЕДЕНИЕ

Развитие подходов и концепции к построению систем, связанных с накоплением данных в хранилище и последующим использованием алгоритмов Data Mining [1-3] является особо перспективным, т.к. белорусский сектор соответствующих ИТ-разработок находится еще на стадии формирования. Рассматриваемая тематика, связанная с разработкой общей концепции построения систем накопления и анализа данных [4], включает следующие направления исследований. Прежде всего, структурирование предметной области, связанной с объектами художественной ценности, построение модели для хранилища данных, выбор и разработка методологии и алгоритмов обработки информации. Кроме того, системой предусматривается также сбор и накопление информации, касающейся персональных данных людей, прежде всего, известных личностей, а также, направление, характеризующее некоторые аспекты предметной области демографических и миграционных данных [5, 6].

## I. ОБ АРХИТЕКТУРЕ СИСТЕМЫ ХРАНЕНИЯ И ОБРАБОТКИ ДАННЫХ ПРАКТИКО-ОРИЕНТИРОВАННОЙ НАПРАВЛЕННОСТИ

На рисунке 1 представлена разработанная архитектура для предлагаемой системы хранения и обработки данных практико-ориентированной направленности.

Основными компонентами архитектуры предлагаемой системы являются: оперативные источники данных; ETL-процесс; хранилище данных; подсистема анализа; пользователи. Приведем кратко характеристику основных компонентов системы.

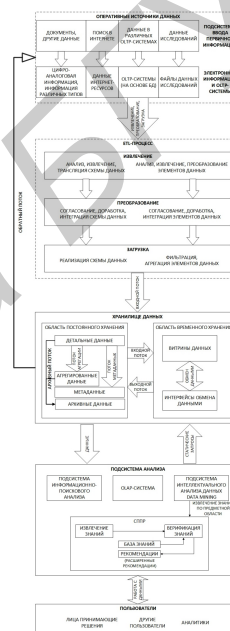


Рис. 1 – Общая архитектура системы хранения и обработки данных практико-ориентированной направленности

Оперативные источники данных включают различные документы и данные, которые обрабатываются OLTP-системами. ETL-процесс представляет собой процесс извлечения (получения) информации из OLTP-систем (баз данных), затем ее дальнейшее преобразование к формату хранимых данных в хранилище, и непосредственной загрузки данных в хранилище данных. Архитектура области хранения данных может проектироваться в зависимости от требований представления и использования итоговой информации. В рассматриваемом случае архитектура проектируется с учетом следующих областей. Область временного хранения данных – используется для временного хранения данных, которые извлекаются из транзакционных систем; таким образом, эта область – промежуточный

слой между OLTP-системами и хранилищем данных, которая состоит из витрин данных и интерфейсов обмена данными с другими системами и приложениями. Область постоянного хранения данных состоит из детальных данных; агрегированных данных; метаданных и архивных данных. Подсистема анализа также важная часть предлагаемого архитектурного решения, основными требованиями к которой являются расширяемость и модульность структуры. В подсистему информационно-поискового анализа необходимо включить информационный поиск, визуальный поиск и семантический поиск. постановка четких запросов к базе данных и получение соответствующей информации – это главные задачи, решаемые подсистемой информационного и визуального поиска.

Использование технологии OLAP в предлагаемой системе – это соответствующий подбор программных решений класса Business Intelligence для быстрого создания и настройки многомерных кубов с целью получения итоговых агрегированных запросов по конкретному направлению, соответствующих направлению предметной области. Подсистема интеллектуального анализа данных Data Mining предполагает возможность с использованием соответствующих методов обработки данных обнаружения неизвестных ранее нетривиальных и практически полезных знания, которые можно интерпретировать соответствующим образом и на этой основе принимать соответствующие решения по направлениям деятельности, указанным выше. Основное назначение системы поддержки принятия решений (СППР) состоит в получении рекомендаций на основе изучения имеющихся исторических и текущих данных о состоянии требуемого объекта исследования и сравнение этой информации с информацией, которая хранится в базе данных системы. Основные функции СППР – это извлечение знаний, их верификация и формирование рекомендаций требуемого образца. В СППР обязательно должна присутствовать возможность обновления знаний с учетом текущего момента времени (актуализация знаний). К системе могут иметь доступ различные группы пользователей, которым доступен и различный функционал системы. Отметим, что, кроме администраторов системы, можно указать также следующие группы пользователей: лица, принимающие решения, аналитики и другие пользователи. Лица, принимающие решения (ЛПР) – это группа пользователей, от решения которых зависит анализ объективной составляющей ситуации или явления; выявление предпочтений ЛПР, генерация возможных решений, оценка возможных альтернатив, анализ последствий принимаемых решений, выбор лучшего варианта и т.д. Аналитики – пользователи, которым необходимо сформулировать и проверить конкретные гипотезы, а также создавать требуемые запросы раз-

личного плана и подготовить отчеты по конкретным направлениям предметной области. Другим пользователям, как правило, доступна общая информация, связанная с универсальной системой и для них ограничены возможности по обработке имеющихся данных предлагаемой системы.

Однако, стоит помнить, что, в зависимости от целей использования системы, всегда можно определить конкретного пользователя и предоставить ему определенные права на доступ и обработку информации, а также – к отдельным модулям и функционалу системы.

## II. ЗАКЛЮЧЕНИЕ

Разработка общей концепции и реализация Интернет-системы складирования и обработки данных практико-ориентированной направленности, связанных с различными видами деятельности людей, может быть рассмотрена в аспекте создания некоторого федерального хранилища данных, что, несомненно, является актуальной темой исследования. Система такого рода позволит собрать достаточно обширные сведения по различным предметным областям, а также получать необходимые аналитические сводки, проводить обработку данных и применять соответствующие методы и алгоритмы Data Mining.

*Результаты работы получены в процессе выполнения ГПНИ «Разработка методологии и средств построения универсальных систем хранения, обработки и анализа структурированных данных большого объема практико-ориентированной направленности».*

## III. СПИСОК ЛИТЕРАТУРЫ

1. Devlin, B.A. An Architecture for a Business and Information System / B.A. Devlin, P.T. Murphy. – IBM Systems Journal, 1988. – Vol 17, No 1. – P. 60–80.
2. Inmon, W.H. Building the Data Warehouse / W.H. Inmon // Third Edition. – John Wiley and Sons, Inc. New York, 2002. – 428 p.
3. Kimbell, R. The Data Warehouse Toolkit: The Complete Guide to Dimensional Data Warehouses / R. Kimbell, M. Ross // Second Edition. – J. Wiley and Sons, 2002. – 447 p.
4. Примеры реализации хранилищ данных предприятия [Электронный ресурс]. – Режим доступа: [http://www.internet-technologies.ru/articles/article\\_994.html](http://www.internet-technologies.ru/articles/article_994.html). – Дата доступа: 26.12.2016.
5. Belyi, A. Global multi-layer network of human mobility // Alexander Belyi, Iva Bojic, Stanislav Sobolevsky, Izabela Sitko, Bartosz Hawelka, Lada Rudikova, Alexander Kurbatski, Carlo Ratti / International Journal of Geographical Information Science. – Mode of access: [<http://www.tandfonline.com/doi/full/10.1080/13658816.2017.1301455>]. – Date of access: [14.04.2017].
6. Белый, А.Б. Данные сервиса Flickr и структура сообществ стран // А.Б. Белый, Л.В. Рудикова, С.Л. Соболевский, А.Н. Курбацкий / Международный конгресс по информатике: информационные системы и технологии = International Congress on Computer Sciens : Information Systems and Technologies : материалы Междунар. науч. конгресса, Минск, 24 окт.–27 нояб. 2016 г. / БГУ; редкол.: С.В. Абламейко (отв. ред.) [и др.]. – Минск, 2016. – С. 851–855.