

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ СКЛАДИРОВАНИЯ ДАННЫХ ДЛЯ ПОСТРОЕНИЯ АРХИТЕКТУРЫ СИСТЕМЫ СБОРА И АНАЛИЗА ДАННЫХ ПРОИЗВЕДЕНИЙ ИСТОРИЧЕСКОЙ ЦЕННОСТИ

Рудикова Л. В., Бандысик С. Ю.

Кафедра современных технологий программирования, Гродненский государственный университет имени Янки Купалы

Кафедра программного обеспечения информационных технологий, Белорусский государственный университет информатики и радиоэлектроники

Гродно, Минск, Республика Беларусь

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

Представлен подход к построению архитектуры для системы сбора и анализа данных о произведениях исторической ценности с использованием технологий складирования данных. Предлагается общая архитектура для построения систем хранения, обработки и анализа данных, выделены основные структурные элементы системы. Лежащая в основе уровневая архитектура, позволяет построить расширяемую систему с независимыми между собой слоями. Предполагается, что представленная система будет способствовать генерации дополнительной информации о предметах исторической ценности.

ВВЕДЕНИЕ

Развитие методов записи и удешевление стоимости хранения данных способствует возрастанию объема накапливаемой информации из различных источников, которая практически никак не обработана. При более тщательном рассмотрении предметных областей, для которых используются технологии складирования данных, можно выделить средства и ресурсы, используемые для анализа накопленной информации [1].

В результате работы над объектами живописи и предметами исторических ценности [2],

многие лаборатории накапливает огромное количество данных, которые представляют собой потенциальные возможности по добыче новой, аналитической информации, используя ее можно выявить какие-либо тенденции, а также получить новую информацию, используя алгоритмы Data Mining [3-5]. Для этой цели необходима централизованная система хранения и анализа, которая будет способствовать генерации дополнительной информации о предметах исторической ценности, а также предоставлять пользователям доступ к уже накопленным данным.

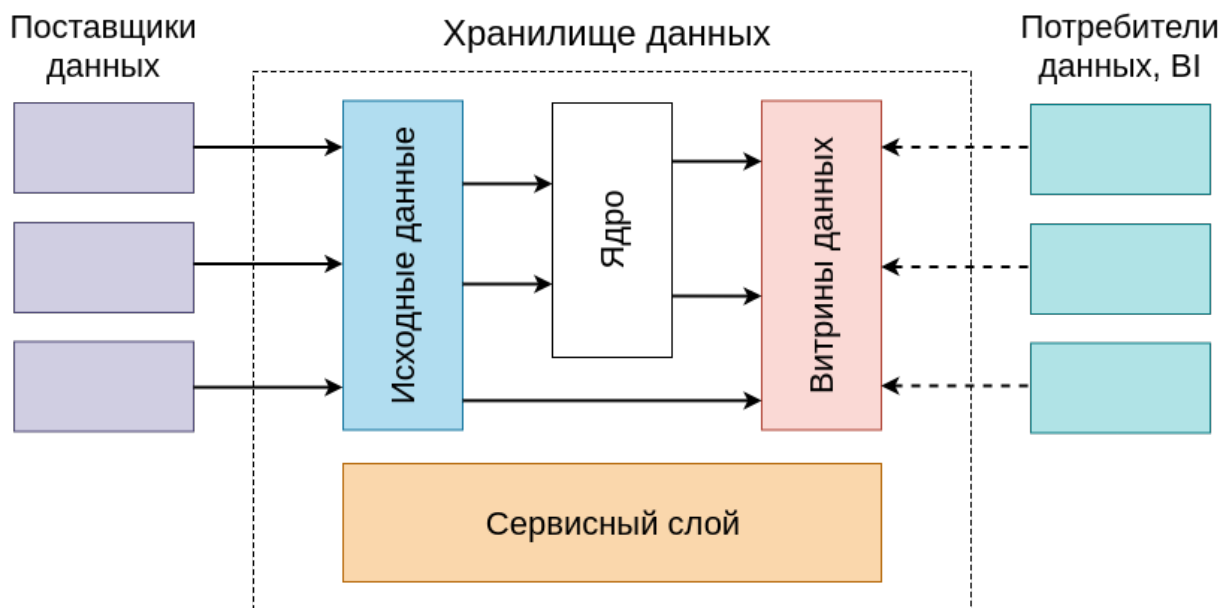


Рис. 1 – Схема архитектуры системы

ОБЩАЯ СХЕМА АРХИТЕКТУРА СИСТЕМЫ ДЛЯ СБОРА И АНАЛИЗА ДАННЫХ ПРОИЗВЕДЕНИЙ ИСТОРИЧЕСКОЙ ЦЕННОСТИ

Реализация архитектуры для систем складирования данных является непростой задачей. Для борьбы со сложностью системы необходимо использовать уровневую архитектуру, где слои реализованы независимо и абстрагированы от других. Рассматривая их по отдельности, можно более качественно выделить зоны ответственности для каждого.

Схематично можно представить следующую схему для построения системы для сбора и анализа данных произведений исторической ценности (см. рис. 1).

На схеме предложены основные слои. Слой исходных данных; Слой ядра; Слой витрин данных. Также дополнительно выделяется Сервисный слой, который отвечает за управление загрузкой в систему данных, а также за контроль качества.

Изложим некоторые характеристики основных компонентов системы.

Слой исходных данных, который отвечает за загрузку исходной информации без изменений из систем-поставщиков. Назначение слоя: позволяет данным на последующих слоях не зависеть от устройства поставщиков данных.

Слой ядра, является основным компонентом системы, который придает хранилищу данных упорядоченность, так как его главная роль – это объединение данных от нескольких поставщиков, преобразование к единым структурам. Как раз при работе на этом слое проводится важная работа по качественному обеспечению данных. К тому же здесь происходят общие преобразования, которые могут оказаться ресурсозатратными. Назначение слоя ядра: обеспечение целостности данных, а также их качества. Независимость от структуры, своими потребителями и данных поставщиков.

Слой витрин данных. Оперировать данными от надежного поставщика, используемого внутри системы, который расположен на слое ядра. Также, в некоторых случаях, возможно получение данных из слоя исходных данных. Данный способ используется в задачах, где необходим быстрый доступ к данным. Там где скорость доступа важнее качества данных. Назначение слоя: подготавливает данные, трансформируя их в удобные для анализа структуры.

Представленное логическое распределение по слоям физически может варьироваться. Для достижения быстродействия, могут быть использованы различные платформы, которые хранят, либо преобразовывают данные на слоях. Также для поддержки информационного контроля выполняющихся потоков данных, а также процессов получения информации от поставщиков, данные, представленные в таблицах для уровней

архитектуры помечаются специальными meta-атрибутами.

Сервисный слой осуществляет вспомогательные операции для всех представленных уровней, структурирует обрабатываемые потоки данных. Структуры хранения представлены областью метаданных и областью для работы с качеством данных.

ЗАКЛЮЧЕНИЕ

Приведена общая схема архитектуры для построения систем хранения, обработки и анализа данных. Также были выделены и даны краткие характеристики основных структурных элементов. Таким образом, изложенная архитектурная концепция может быть взята за основу для реализации различного вида систем, предусматривающих использование технологий хранения данных. Так как подобного рода системы позволяют собирать достаточно объемные данные из разных предметных областей, применять необходимые алгоритмы Data Mining, а также генерировать необходимые аналитические отчеты и обрабатывать данные. В частности, данную архитектуру планируется использовать для создания системы сбора и анализа данных произведений исторической ценности.

Таким образом, предлагаемое архитектурное решение представляет собой расширяемую и гибкую модульную систему в основе которой лежит сбор информации из различных источников для последующего анализа, включая генерацию новых, неочевидных зависимостей. Представленная архитектура системы предусматривает возможность расширения и гибкой модификации. На ее базе можно построить систему для использования в качестве вспомогательного инструмента для построения отчетов и проведения различного рода исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Рудикова, Л. В. Об общей архитектуре универсальной системы хранения и обработки данных практико-ориентированной направленности / Л. В. Рудикова // Системный анализ и прикладная информатика : международный научно-технический журнал.- 2017.- № 2.- С.12-19
2. Рудикова, Л. В. О разработке системы для поддержки экспертизы объектов художественной ценности / Л. В. Рудикова // Технологии информатизации и управления. Вып. 3: в 2 кн.. Кн. 2 : сборник научных статей/под. ред. А. М. Кадан, Е. А. Свицкого.- Минск: РИВШ, 2017.- С.107-117
3. 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.
4. Inmon, W. H. Building the Data Warehouse / W. H. Inmon // Third Edition. – John Wiley and Sons, Inc. New York, 2002. – 428 p.
5. Kimbell, R. The Data Warehouse Toolkit: The Complete Guide to Dimensional Data Warehouses / R. Kimbell, M. Ross // Second Edition. – J. Willey and Sons, 2002. – 447 p.