

УДК 004.03

АНАЛИЗ ОТЕЧЕСТВЕННЫХ РЕШЕНИЙ ПО УПРАВЛЕНИЮ ДАННЫМИ ДЛЯ ОПТИМИЗАЦИИ ЦЕНТРАЛИЗОВАННОЙ МОДЕЛИ ХРАНЕНИЯ МЕТАДАННЫХ В УСЛОВИЯХ ИМПОРТОЗАМЕЩЕНИЯ

Цынченко Я.М. (ИТМО)

Научный руководитель – доцент, факультет технологического менеджмента и инноваций Горгадзе А.А (ИТМО)

Введение. Неумение правильно организовать данные, которые получает, накапливает и использует организация, приводит к убыткам и упущенным возможностям. Информация – ценный актив. Чтобы управлять основополагающим активом важно понимать, какую структуру он имеет и как с ним обращаться. Данные, для управления другими данными принято называть метаданными. Функция управления метаданными является одной из основополагающей в организации управления данными. Корпорации внутри нашего государство активно открывают проекты импортозамещения иностранных продуктов, так как из-за санкций многие иностранные вендоры отказываются оказывать техническую поддержку. Важным является изучение потребностей корпораций во введении метаданных, а также покрытие этих потребностей отечественными решениями на рынке.

Основная часть. В современных корпоративных средах эффективное управление информацией становится ключевым фактором успеха. Функция метаданных должна обеспечивать согласованность данных и делать информацию доступной для сотрудников из разных областей деятельности. Для достижения этих целей многие корпорации предпочитают использовать продукты вендоров с централизованной системой хранения метаданных. Такой подход обеспечивает единый и структурированный доступ к информации, что упрощает ее анализ и использование [2].

Корпорации выдвигают расширенные требования к техническим метаданным, так как причиной является разнообразие источников данных в различных областях бизнеса. Что создает потребность в специализированных решениях, способных обрабатывать и анализировать данные из различных источников [3]. В анализе рассматривались возможности сканирования технических метаданных в отечественных решениях таких компаний, как Arenadata, Ростелеком, DIS-group. В рамках анализа было выявлено, что отечественные продукты вендоров на текущий момент не полностью соответствуют требованиям корпораций по сканированию технических метаданных. Для покрытия функционального дефицита в работе была предложена модель централизованной структуры хранения метаданных с добавлением блока собственной разработки для сканирования метаданных с использованием открытого программного обеспечения Apache Airflow [1]. Это решение имеет ряд преимуществ, включая импортонезависимость и гибкость настройки под конкретные потребности организации

Выводы. В результате анализа предложений отечественных вендоров был выявлен функциональный дефицит в области сканирования технических метаданных. Для решения данной проблемы предложена модель централизованной структуры хранения метаданных с использованием открытого программного обеспечения Apache Airflow.

Список использованных источников:

1. Apache Airflow: Use Cases, Architecture, and Best Practices. – URL: <https://www.run.ai/guides/machine-learning-operations/apache-airflow> (дата обращения: 13.02.2024). – Текст: электронный.

2. DAMA-DMBOK: Свод знаний по управлению данными. Второе издание / Data International [пер. с англ. Г. Агафонова]. – Москва: Издательство «ОЛИМП-БИЗНЕС», 2020. – 828 с.
3. Warden, P. Big Data Glossary: a guide to the new generation of dta tools. Big Data Glossary / P. Warden. – Sebastopol, CA: O'Reilly Media, 2011. – 44 с.