

# ИССЛЕДОВАНИЕ МЕТОДОЛОГИИ ХРАНЕНИЯ ДАННЫХ ДЛЯ РЕАЛИЗАЦИИ СИСТЕМЫ МАСТЕР ДАННЫХ

## RESEARCH ON DATA STORAGE METHODOLOGY FOR THE IMPLEMENTATION OF MASTER DATA SYSTEM

**D. Lagerev**  
**E. Koptenok**  
**I. Fomin**

*Summary:* The problem of handling large volumes of data, specifically master data, is a highly relevant task in the field of modern information technology. This paper presents a comparative analysis of data storage models with the aim of selecting the best model for its application in the architecture of a master data management software suite.

*Keywords:* master data, data storage model, Data Vault, Anchor model, Inmon, data warehouse.

**Лагерев Дмитрий Григорьевич**

кандидат технических наук, доцент,

Брянский государственный технический университет

lagerevdg@yandex.ru

**Коптенок Елизавета Викторовна**

старший преподаватель,

Брянский государственный технический университет

elizabeth-evans@yandex.ru

**Фомин Илья Игоревич**

Магистрант,

Брянский государственный технический университет

pat1\_2010@mail.ru

*Аннотация.* Проблема представления больших объемов данных, в частности мастер-данных, является весьма актуальной задачей в области современных информационных технологий. В работе приводится сравнительный анализ моделей хранения данных с целью выбора наилучшей модели для ее применения в архитектуре программного комплекса управления мастер-данными.

*Ключевые слова:* мастер-данные, модель хранения данных, Data Vault, Anchor model, Inmon, хранилище данных.

### Введение

В настоящее время в Российской Федерации сложилась ситуация, когда привычное программное обеспечение для управления мастер данными зарубежных вендоров не может быть использовано во многих российских компаниях по ряду причин, а пришедшее им на смену программное обеспечение с открытым исходным кодом, и приложения, построенные на основе такого ПО, не могут пока обеспечить привычного пользователю набора функций. Таким образом, весьма актуальным вопросом является разработка системы управления мастер-данными, поддерживающей СУБД с открытым исходным кодом и обеспечивающей привычный для пользователя набор функций.

Мастер-данные («основные данные» или «нормативно-справочная информация») — это совокупность основных данных, справочной информации, то есть значения, которые могут использоваться для указания, к чему какие данные относятся. Самый простой пример применения мастер-данных — разного рода справочники или классификаторы, применяемые в деятельности организаций. К ним относится информация о клиентах, товарах, услугах, сотрудниках, сделках и многое другое [3]. Задача управления мастер-данными, включающая сбор, хранение, изменение и валидацию данных, является очень важной. Также важную роль играет скорость операций с данными, безопасность и надежность пред-

ставления данных, удобство выполнения действий для пользователя.

### Анализ программного обеспечения по управлению мастер-данными

Для управления мастер-данными можно применять следующие инструменты: MDM-системы или системы управления базами данных (СУБД).

При применении СУБД к решению поставленной задачи возникает ряд сложностей, а именно:

- потребность в специфических знаниях и навыках при работе с сервером и хранилищем данных;
- высокая трудоемкость рутинных задач по организации целостности данных;
- сложность и трудоемкость обеспечения прав доступа и версионирования;
- риск нарушения ценности данных при их модификации.

Частично эти трудности решаются за счет применения MDM-систем. Однако, данные системы имеют ряд недостатков, представленных далее.

В работе рассмотрены следующие MDM-системы: Microsoft Master Data Services, Oracle Meter Data Management, LM Soft MDM.

Наиболее популярной MDM системой для Microsoft SQL Server является решение от Microsoft [5]. Данная система впервые появилась с релизом SQL Server 2008 R2. Microsoft Master Data Services предоставляет основной функционал для управления мастер-данными на платформе Microsoft SQL Server.

К ее достоинствам можно отнести:

- возможность объединения справочников в сложные иерархические каталоги;
- гибкие политики безопасности;
- поддержка доменных записей;
- гибкие настройки бизнес-правил;
- возможность версионирования.

Недостатками данной системы является следующее:

- отсутствие в реестре российского программного обеспечения;
- не является свободно распространяемым ПО;
- система ограничена поддерживаемой СУБД: Microsoft SQL Server;
- дорогостоящая лицензия на СУБД Microsoft SQL Server;
- web версия с устаревшим non-userfriendly интерфейсом.

Система управления мастер-данными Oracle Meter Data Management разработана компанией Oracle Corporation для совместного использования с Oracle Database [6]. Данная система обладает обширной функциональностью по управлению мастер-данными.

Можно выделить следующие достоинства системы:

- возможность объединения справочников в сложные иерархические каталоги;
- возможность хранения структурированных и неструктурированных данных одновременно;
- гибкие политики безопасности;
- поддержка доменных записей;
- возможность версионирования;
- обширные интеграционные механизмы с экосистемой Oracle: Oracle ETL, шина сообщений, XML сообщения в реальном времени;
- поддержка пакетной отправки данных;
- возможность гибкой настройки бизнес-правил.

Недостатки системы представлены ниже:

- система ограничена поддерживаемой СУБД: Oracle Database;
- отсутствие в реестре российского программного обеспечения, не является свободно распространяемым ПО.

LM Soft MDM [7] — российское решение по управлению мастер-данными, функционирующее на основе программного комплекса 1С:Предприятие [8].

К его достоинствам можно отнести:

- возможность разбиения справочников на классификационные группы;
- гибкие политики безопасности;
- поддержка правописания русского языка с исправлением опечаток;
- импорт данных из различных видов источников (СУБД, web-сервисы, файлы, электронные каталоги и т.п.);
- наличие в реестре российского программного обеспечения.

Можно привести следующие недостатки LM Soft MDM:

- система ограничена возможностями программного комплекса «1С: Предприятие», так как построена на его основе;
- отсутствие принципа золотой записи;
- отсутствие иерархичности справочников;
- отсутствие механизма поддержки доменных записей.

Таким образом, анализ программного обеспечения по управлению мастер-данными позволил выделить ряд недостатков у существующего программного обеспечения. Исходя из этого, следует необходимость разработки программного обеспечения для управления мастер-данными, представляющего собой автоматизированную систему по представлению разнородной бизнес-информации в едином структурированном виде.

#### Актуальность работы и постановка задачи исследования

Основное назначение систем управления мастер-данными — это хранение больших объемов информации, агрегированной из различных источников с сохранением структуры данных, с возможностью внесения последующих изменений. От используемой модели хранения данных зависит, на сколько быстро будут выполняться операции по изменению данных и их структуры, а следовательно, эффективность управления данными. Существует множество концепций хранения данных, имеющих ряд принципиальных различий, вследствие чего возникает проблема выбора оптимальной модели хранения данных для построения на ее основе системы управления мастер-данными. В данной работе производится анализ следующих моделей хранения данных: Data Vault [1], Anchor model [2] и Inmon [2].

#### Обзор моделей хранения данных

Модель данных пригодная для построения программного комплекса управления мастер данными должна поддерживать следующие функции: очистку, слияние, сопоставление, извлечение, преобразование,

удаление, добавление, а также поддержку историчности данных и их аналитику. Рассматриваемые модели обладают данными функциями, но их эффективность в выполнении данных операций различна.

Концепция **Inmon** [2] предполагает разработку хранилища «сверху вниз». Можно выделить следующие этапы.

1. Создание нормализованной модели данных.
2. Создание витрин размерных данных.
3. Создание исчерпывающей модели для каждой первичной сущности.
4. Построение физической модели, соответствующей нормализованной.

Можно выделить следующие особенности данной модели.

1. Низкая избыточность данных.
2. Подробное представление бизнес-объектов в рамках одной сущности.
3. Гибкость в обновлении хранилища.
4. Сложность модели растет пропорционально количеству сущностей.
5. Требуется дополнительная операция ETL при построении хранилищ.
6. Поддержка хронологии предполагает отсутствие удаления и изменения данных, а только добавление новых сведений.
7. Высокая интегрированность данных.

Концепция **Anchor model** [2] — это разновидность гибких методологий моделирования данных.

Методология **Anchor Modeling** предлагает свою собственную графическую нотацию для построения моделей. По внешнему виду она близка к ER-диаграмме. Можно выделить следующие элементы модели:

1. Якорь — набор уникальных бизнес-ключей.
2. Атрибут — дополнительные сведения о якорях.
3. Связи — отношения между якорями.

Модель имеет ряд особенностей.

1. Стремление к сильно нормализованной модели.
2. Обработка изменений в хранилище данных без отмены предыдущих действий.
3. Все изменения осуществляются за счет расширений, а не модификаций.
4. Модульность, которая позволяет декомпозировать модель данных на устойчивые и хорошо управляемые компоненты, развивающиеся итеративно.
5. Высокая скорость выполнения темпоральных запросов.

Методология **Data Vault** [1] предполагает объединение третьей нормальной формы представления данных и схемы представления данных «Звезда».

Основными элементами модели являются хаб, ссылка и сателлит.

Хаб — это представление сущности в модели. Ссылки — это таблицы, применяемые для связи хабов в отношении «многие-ко-многим». Сателлиты предназначены для хранения описательных атрибутов сущностей и метаданных.

Особенности данной модели представлены далее.

1. Гибкость и расширяемость.
2. Поддержка гибких методологий разработки хранилищ данных.
3. Сложность применения методологии.
4. Избыточность данных.

Рассматриваемые методологии существенно отличаются друг от друга. Для того, чтобы сделать выбор в пользу одной из них для использования при управлении мастер-данными необходимо провести сравнительный анализ.

#### Сравнительный анализ методологий хранения данных

Для того, чтобы выбрать оптимальную модель для хранения мастер-данных, необходимо ввести ряд критериев для сравнения моделей. В качестве первого критерия выступает поддержка гибких методологий разработки хранилищ данных. Современный подход к разработке программного обеспечения очень часто использует Agile-технологии.

Однако, среди представленных моделей гибкий подход к разработке поддерживают только **Data Vault** и **Anchor model**.

Вторым критерием является адаптивность. Способность легко изменять данные и их структуру делает хранилище более гибким и удобным в применении. Данный параметр может быть оценен на основе объективных фактов о моделях. Пусть введен ряд критериев с соответствующими ему количественными значениями для последующей комплексной оценки.

0. Внесение изменений критично и ведет к разрушению модели.

1. Сложность адаптации. Требуется большое число дополнительных операций.

2. Высокая гибкость модели в силу ее расширяемости.

Тогда имеются следующие оценки моделей хранения данных: **Inmon** — 1, **Anchor model** — 1, **Data Vault** — 2.

Значимым критерием является производительность. Рассматриваются два типа операций: выборка данных и изменение данных. Для получения числовых характеристик выбрана предметная область «Медиатека», имеющая следующие сущности: «Исполнители» (приблизительно 70000 записей), «Альбомы» (240000 записей), «Треки» (2500000 записей), «Продюсеры» (1850 записей). Построены хранилища данных согласно каждой из указанных моделей, над каждой из которых многократно (1000 раз) выполнены одни и те же данные по извлечению данных (поиск всех альбомов артиста, поиск треков для указанного исполнителя, поиск исполнителей по продюсеру) и изменению данных (добавление новых записей и изменение текущей записи; стоит отметить, что для моделей Inmon и Anchor model изменение записи подразумевает добавление новой с учетом ее историчности). Рассчитано среднее время выполнения операций, результаты представлены в табл. 1.

Таблица 1.

Результаты вычисления производительности

Наименование модели хранения данных	Среднее время выполнения операций по выборке данных, секунды	Время выполнения операций по изменению данных, секунды
Inmon	0,29	0,24
Anchor model	0,21	0,31
Data Vault	0,14	0,19

Для получения количественного критерия сравнения введена обратная пропорциональная зависимость вида

$\{1/(\text{время выполнения операции 1}) + 1/(\text{время выполнения операции 2})\}$ . Для каждой модели хранения данных получены следующие значения:

Inmon — 14,37.

Anchor model — 15,36.

Data Vault — 37,59.

Стоит заметить, что модель Data Vault обладает наивысшей производительностью.

Таблица 2.

Сравнительный анализ моделей хранения данных

Наименование модели хранения данных	Поддержка гибких методологий разработки	Производительность	Адаптивность	Суммарно
Inmon	0	14,3	1	15,3
Anchor model	1	15,36	1	17,36
Data Vault	1	37,59	2	40,59

В табл. 2. представлен результат сравнительного анализа методологий хранения данных. Очевидно, что по выделенным критериям Data Vault превосходит другие методологии более чем в два раза.

### Заключение

Выбор методологии хранения данных является критично важным решением в архитектуре программного комплекса управления мастер-данными, поскольку от него зависит надежность и скорость работы программного комплекса. В результате проведенного сравнительного анализа моделей хранения данных выявлено, что наилучшим образом для решения поставленных задач подходит методология Data Vault, обладающая рядом преимуществ, таких, как адаптивность, поддержка гибких методологий разработки, простота управления данными и их структурой.

### ЛИТЕРАТУРА

1. Конаков Павел Олегович, Смоленцева Татьяна Евгеньевна Разработка архитектуры системы управления информационными ресурсами в рамках модели хранилища данных Data Vault 2.0 // Столыпинский вестник. 2022. №9.
2. Солянов Кирилл Сергеевич, Стацюк Любовь Владимировна Фреймворк проектирования логической модели хранилища данных // Инновации и инвестиции. 2020. №8.
3. Бубекова А.Г., Гантц И.С. Особенности управления мастер-данными в банковской сфере // МНИЖ. 2022. №5–1 (119).
4. Кузнецов С.В., Кознов Д.В. Управление мастер-данными в рамках итеративного подхода // Онтология проектирования. 2021. №2 (40).
5. Microsoft MDS [Электронный ресурс]. URL: <https://www.microsoft.com/en-us/download/details.aspx?id=58203> (дата обращения: 15.05.2023).
6. Oracle Meter Data Management [Электронный ресурс]. URL: <https://www.oracle.com/industries/utilities/products/meter-data-management/> (дата обращения: 15.05.2023).
7. LM Soft MDM [Электронный ресурс]. URL: <https://lmsoft.ru/products/produkty-mdm/lm-soft-mdm/> (дата обращения: 15.05.2023).
8. Официальный сайт компании 1С [Электронный ресурс]. URL: <https://1c.ru/> (дата обращения: 15.05.2023).

© Лагерева Дмитрий Григорьевич (lagerevdg@yandex.ru); Коптенко Елизавета Викторовна (elizabeth-evans@yandex.ru);

Фомин Илья Игоревич (pat1\_2010@mail.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»