

DATA MESH: ПРИНЦИПЫ ПОСТРОЕНИЯ ДЕЦЕНТРАЛИЗОВАННОЙ АРХИТЕКТУРЫ ДАННЫХ

Павлович Н.В.

Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь

Научный руководитель: Тонкович И.Н. – канд.хим.наук, доцент, доцент кафедры ПИКС

Аннотация. В статье рассматриваются основные идеи, принципы и компоненты проектирования и разработки децентрализованной архитектуры данных, описываются лучшие практики проектирования децентрализованной архитектур.

Ключевые слова: *Data Mesh*, домены данных, децентрализованная архитектура, Data Warehouse.

Введение. В современном мире данные играют все более важную роль в бизнесе. Компании используют данные для принятия решений, улучшения продуктов и услуг, анализа рынка и конкурентов и многих других задач. Однако управление данными становится все сложнее с ростом объема и сложности данных. Традиционные подходы к управлению данными, такие как централизованная архитектура, часто не могут обеспечить эффективное использование данных. Такие архитектуры требуют высокой степени централизации данных и управления, что может затруднить быстрое развертывание и гибкость в использовании данных. В этой связи методология *Data Mesh* становится все более популярной в качестве решения для эффективного управления данными.

Основная часть. *Data Mesh* – это методология управления данными, которая предполагает разделение данных на отдельные домены, каждый из которых отвечает за свои данные и может предоставлять их другим доменам. Эта методология основана на четырех основных принципах:

- 1 Домены данных. Данные разделены на отдельные домены, что обеспечивает своего рода независимость от системы в целом.
- 2 Децентрализованная архитектура. Каждый домен имеет свою собственную архитектуру и может использовать разные технологии для управления данными.
- 3 Владение данными. Каждый домен имеет владельца данных, который отвечает за качество данных и защиту конфиденциальности.
- 4 Самообслуживание. Каждый домен может самостоятельно принимать решения о том, как использовать свои данные, без необходимости обращения к централизованному управлению данными [1].

Основная идея *Data Mesh* заключается в том, что управление данными должно быть распределенным, децентрализованным и ориентированным на команды. *Data Mesh* – это относительно новый подход к управлению данными, который призван решить проблемы, связанные с централизованными и монолитными системами управления данными. Сравнивая *Data Mesh* с другими подходами к управлению данными, можно выделить аспекты, которые призвана убрать новая методология. В централизованной модели управления данными все данные управляются централизованной командой, которая контролирует доступ и обработку данных. Это ограничивает возможности инноваций и развития внутри компании, так как все зависит от единственной команды, которая может затормозить процесс.

Модель управления данными с использованием *Data Warehouse* предполагает, что все данные собираются в хранилище данных, где они агрегируются и используются для анализа. Однако, при большом объеме данных, такой подход может быть затруднительным, так как обработка и анализ данных занимает много времени.

В модели управления данными с использованием *Data Lake* данные хранятся в неструктурированном формате и могут быть использованы для анализа и обработки в даль-

нейшем. Однако, недостаток этой модели в том, что она может быть очень сложной и трудной в управлении, особенно при большом объеме данных.

Архитектура *Data Mesh* состоит из нескольких компонентов, каждый из которых отвечает за определенные задачи:

1 Доменные данные (*Domain Data*) – это данные, связанные с конкретной функциональной областью или бизнес-доменом. Каждый домен имеет свою собственную команду, которая отвечает за сбор, хранение и управление данными в этом домене.

2 Доменный *API* – это интерфейс, который позволяет получать доступ к данным конкретного домена. *API* предоставляется каждой командой домена и используется для взаимодействия с другими командами и подразделениями.

3 Платформа данных (*Data Platform*) – это инфраструктура, которая поддерживает работу с данными в компании. Она включает в себя инструменты для сбора, хранения, обработки и анализа данных, а также для управления и мониторинга доступа к данным.

4 Команда по качеству данных (*Data Quality Team*) – это команда, которая отвечает за контроль качества данных в компании. Она разрабатывает стандарты и методологии для оценки качества данных, а также контролирует их соответствие этим стандартам.

5 Каталог данных (*Data Catalog*) – это реестр, который содержит информацию о всех данных в компании, их источниках, метаданных, структурах и т.д. Каталог данных используется для управления доступом к данным и для облегчения поиска необходимых данных.

Архитектура *Data Mesh* создает экосистему данных, которая позволяет распределенным командам управлять своими данными и обмениваться ими с другими командами в компании.

Использование подхода *Data Mesh* позволяет добиться:

1 Распределенная модель управления данными обеспечивает большую гибкость в работе с данными, так как каждая команда или бизнес-юнит может работать со своими данными, не зависимо от других команд.

2 *Data Mesh* помогает расширить масштаб возможностей работы с данными, что может привести к более быстрому и качественному принятию решений на основе данных.

3 Методология *Data Mesh* поддерживает культуру данных в организации, где каждый сотрудник может лучше понимать, как работать с данными и как они могут использоваться для улучшения бизнес-процессов [2].

Для того, чтобы успешно внедрить *Data Mesh*, следует придерживаться следующих лучших практик:

1 Разделение на области. Эта практика требует определения области данных в организации и разделения их на автономные части (*Data Domain*). Каждая область данных должна иметь свойственные ей правила, стандарты, технологии и процессы.

2 Управление метаданными. Определяющее значение имеет систему управления метаданными, которая будет хранить и управлять метаданными областей данных, а также связывать их между собой.

3 Реализация *API*. Необходимо создать *API*-интерфейсы, чтобы обеспечить доступ к данным из различных областей данных.

4 Системы управления доступом. Потребуется реализовать системы управления доступом, которые будут обеспечивать безопасный доступ к данным в соответствии с правилами и стандартами, установленными для каждой области данных.

5 Постоянное улучшение. Внедрение *Data Mesh* – это непрерывный процесс. Постоянно улучшайте свои процессы, технологии и правила, чтобы сохранять конкурентные преимущества на рынке [3].

Теперь рассмотрим, что делает *Data Mesh* уникальным и почему он может быть более эффективным для некоторых компаний.

Data Mesh предоставляет децентрализованный подход к управлению данными, что позволяет каждой команде внутри организации управлять своими данными. Это облегчает

процесс разработки и инноваций, так как каждая команда может работать в своей собственной области без необходимости ждать разрешения от централизованной команды. *Data Mesh* предполагает использование общих стандартов и инструментов для работы с данными, что упрощает процесс обмена данными между командами внутри организации.

Data Mesh предоставляет возможность использовать разные технологии и инструменты для управления данными в зависимости от потребностей каждой команды. Это позволяет использовать наиболее подходящие инструменты для каждой конкретной задачи, что повышает эффективность работы. *Data Mesh* предполагает использование метаданных для описания данных и их контекста. Это упрощает поиск и понимание данных, а также позволяет лучше понимать их ценность и использование. *Data Mesh* ставит акцент на культурные изменения в организации, так как для успешной реализации этого подхода необходимо сотрудничество и коммуникация между командами [4].

Однако, не стоит отрицать, что подход *Data Mesh* имеет ряд недостатков. Распределенная модель управления данными может привести к сложностям в согласовании и сопоставлении данных, когда команды с разными подходами и схемами работы с данными сотрудничают между собой. *Data Mesh* требует от организации значительных усилий в разработке инфраструктуры, чтобы создать подходящую архитектуру данных и обеспечить безопасность и конфиденциальность данных. *Data Mesh* может потребовать от сотрудников компании дополнительных знаний и навыков для работы с данными, что может быть сложно для людей, не имеющих опыта работы с данными.

Заключение. *Data Mesh* может быть более эффективным для компаний, которые работают в большом масштабе и имеют сложные системы управления данными. Этот подход позволяет улучшить эффективность и производительность, упростить управление данными и улучшить коммуникацию внутри организации. Кроме того, можно ожидать, что *Data Mesh* будет становиться все более популярным среди компаний, которые ищут новые способы управления данными и увеличения эффективности своих бизнес-процессов.

Список литературы

1. Концепция *Data Mesh*. Принципы, идеи, применение на практике [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/post/685626/> – Дата доступа: 10.02.2023.
2. *Data Mesh architecture* [Электронный ресурс]. – Режим доступа: <https://www.datamesh-architecture.com/> – Дата доступа: 11.02.2023.
3. *Data Platforms. Data Mesh* [Электронный ресурс]. – Режим доступа: <https://www.montecarlodata.com/blog-what-is-a-data-mesh-and-how-not-to-mesh-it-up/> – Дата доступа: 11.02.2023.
4. *Data Mesh* [Электронный ресурс]. – Режим доступа: <https://www.bigdataschool.ru/blog/data-fabric-vs-data-mesh.html> – Дата доступа: 13.02.2023.

UDC 004.623 – 004.652

DATA MESH: PRINCIPLES FOR BUILDING A DECENTRALIZED DATA ARCHITECTURE

Paulovich N.V.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Tonkovich I.N. – PhD, associate professor, associate professor of the Department of ICSD

Annotation. The article discusses the main ideas, principles and components of designing and developing a decentralized data architecture, describes the best practices for designing decentralized architectures.

Keywords: Data Mesh, data domains, decentralized architecture, Data Warehouse.