



> Конспект > 1 урок > Data Management

> Оглавление

- > [Оглавление](#)
- > [Управление данными](#)
 - > [Data management](#)
 - > [Data Governance](#)
- > [Основные практики](#)
 - > [Data Architecture](#)
 - > [Data Modeling](#)
 - > [Data Storage](#)
 - > [Data Security](#)
 - > [Data Integration](#)
 - > [Document & content](#)
 - > [Master Data & Reference](#)
 - > [DWH & BI](#)
 - > [Meta Data](#)
 - > [Data Quality](#)
- > [Каталог и происхождение данных](#)
 - > [Data catalog](#)
 - > [Происхождение данных](#)
 - > [Инструменты для управления данными](#)

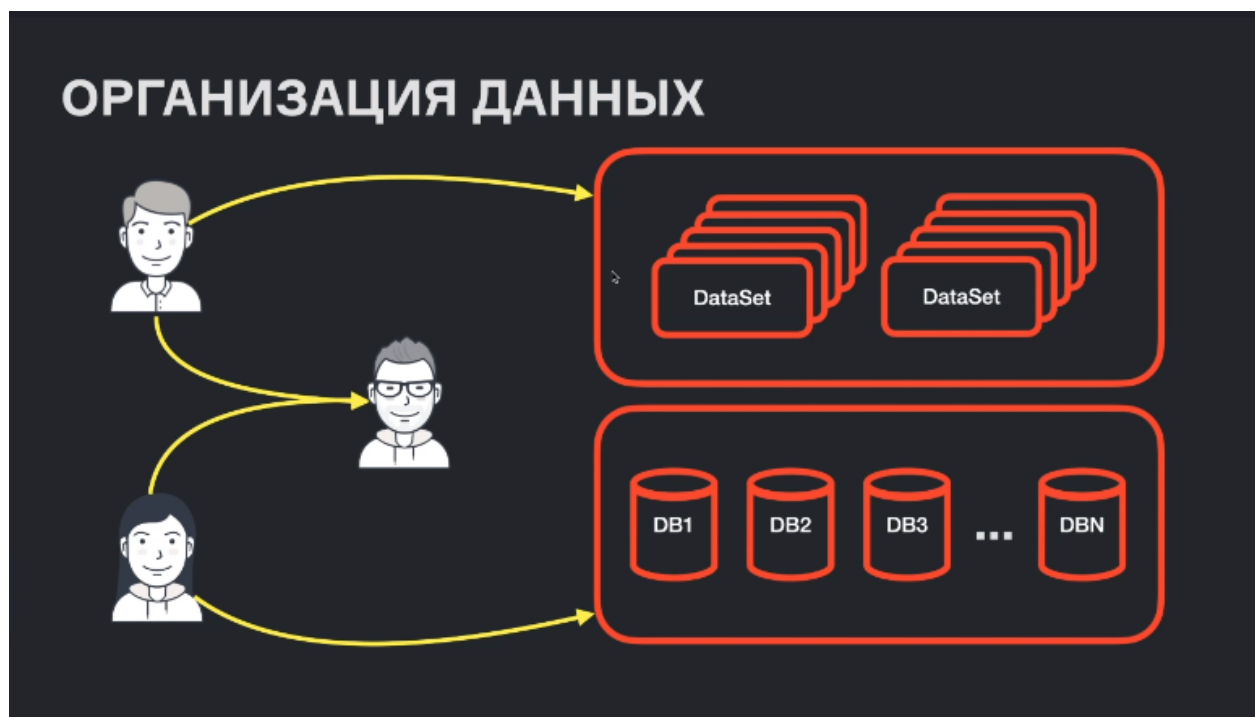
> Управление данными

> Data management

Это термин, который описывает процессы, используемые для планирования, определения, включения, создания, получения поддержки, использования, архивирования, извлечения, управления и очистки данных.

Data management задается вопросами:

- Как управлять ролями?
- Как отслеживать соблюдение политик?
- Как измерять метрики?
- Как соблюдать стандарты?
- Как идти к целям?



Процессы управления данными включает в себя:

1. Анализ данных
2. Архитектуру данных
3. Добычу данных

4. Защиту данных
5. Извлечение, преобразование и загрузку данных
6. Моделирование данных
7. Обеспечение их качества
8. Работа с хранилищами данных
9. Управление базами данных
10. Управление репозиториями данных
11. Шифрование данных

Рекомендуемая литература: DAMA [DMBok2](#) - Data Management Body of Knowledge

> **Data Governance**

Руководство данными - это совокупность процессов, ролей, политик, стандартов и показателей, которые обеспечивают эффективное и действенное использование информации, позволяя организациям достичь своих целей.

Data Governance входят:

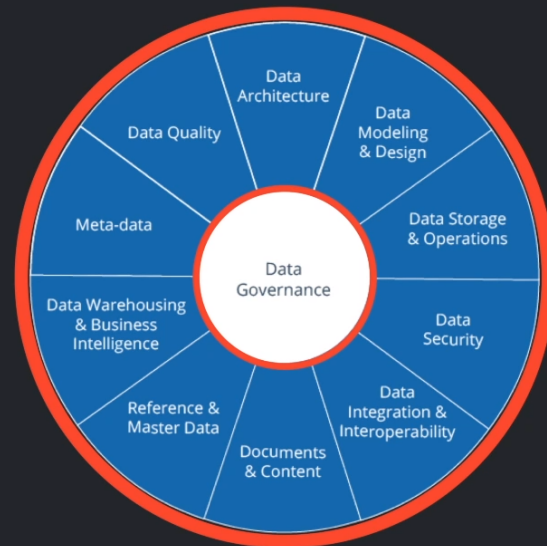
- Роли и обязанности;
- Политики и правила;
- Метрики оценки;
- Стандарты;
- Цели.

Различия между **Data Management** и **Data Governance**

DATA MANAGEMENT

В отличии от **Data Governance**, **Data Management** – отвечает за исполнительную часть в управлении данными, т.е. за претворение в жизнь всего задуманного.

Data Governance – отвечает за планирование, контроль и надзор над данными, использованием данных и связанными с ними ресурсами.



> Основные практики

> Data Architecture

Архитектура данных - описывает структуру логических и физических активов данных, ресурсов управления данными компании в соответствии с бизнес требованиями.

Для достижения результатов есть разные фреймворки:

TOGAF (The Open Group Architecture Framework)

Zachman

ASSIMPLER - архитектурный фреймворк, основанный на работе Мандара Ванарсе в Wipro в 2002 году

> Data Modeling

Моделирование данных - свод процессов, применяемых для анализа требований к данным и создания необходимых структур данных.

Различают несколько подходов:

ER model

dbt (yaml)

DDL (SQL)

> Data Storage

Хранилище данных - применяемые практики развертывания и управления хранилищем физических данных. Применяемые практики:

Backups

Storage strategy

Storage security

Storage technologies

> Data Security

Защита данных - обеспечение конфиденциальности и надлежащего доступа к PII (personal data), PHI (protected health information) и корпоративными данными. Используемые стандарты:

Обезличивание

Шифрование

GDPR

> Data Integration

Интеграция данных - обеспечение процессов по интеграции, извлечению, преобразованию и перемещению данных.

> Document & content

Управление хранением, защитой, индексированием и доступом к данным, находящихся в неструктурированном виде (источниках)

> Master Data & Reference

Это управление общими данными для уменьшения избыточности и обеспечения лучшего качества данных за счет стандартизированного определения и использования значения данных.

> **DWH & BI**

Управление аналитической обработкой данных, обеспечение доступа к данным поддержка и принятие решений для отчетности и анализа.

> **Meta Data**

Управление сбором, категоризацией, обслуживанием, интеграцией, контролем в управлении и доставки метаданных. Она позволяет понять, в каком состоянии находится ваше хранилище данных.

> **Data Quality**

Мониторинг определение и поддержание целостности данных, а так же практика по повышению качества данных. Это популярная практика в больших компаниях, позволяет избегать неправильных решений в бизнесе.

> **Каталог и происхождение данных**

> **Data catalog**

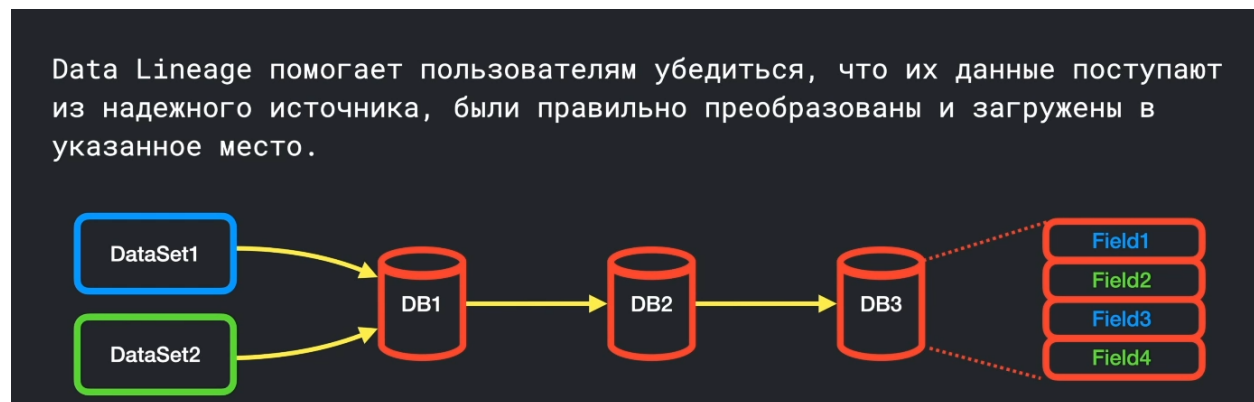
Организованная инвентаризация активов данных, основанная на метаданных для помощи в управлении данными.

Data	Storage	Type	Name
clickstream	ClickHouse	Table	user_clickstream
Item-catalog	Greenplum	Table	item_catalog
Orders	PostgreSQL	Table	user_orders
Item-images	S3	Directory	storage/users/images

Каталог данных можно отнести к практике мета данных, выглядит это обычно в виде таблицы с описанием данных и указанием к тому, где находятся данные.

> Происхождение данных

Data Lineage & Provenance - описывает происхождение и изменения данных с течением времени (как данные были преобразованы, что изменилось и почему). Это упрощает понимание, где могла или произошла ошибка.



> Инструменты для управления данными

DATA TOOLS



DataFold



GCP Data Fusion



Alation



GCP Data Catalog



Apache Atlas



Keboola



Informatica



DBT