



УНИВЕРСИТЕТ ИТМО

# Технологии хранения больших данных

Лекция №7. Озера данных, витрины данных

Ирина Алексеевна Радченко  
[iradche@gmail.com](mailto:iradche@gmail.com)



# Введение

## Обзор темы

- Что такое озера данных и витрины данных
- Значение и роль в экосистеме больших данных

## Цели лекции

- Понимание концепций озер данных и витрин данных
- Изучение преимуществ и недостатков каждого подхода

# Озера данных (Data Lakes)

## Определение

- Хранилище данных в сыром виде, без структурирования

## Особенности

- Поддержка различных типов данных: структурированные, неструктурированные, полуструктурированные
- Гибкость в хранении и обработке данных

## Примеры использования

- Аналитика, машинное обучение, архивирование



# Архитектура озера данных

## Компоненты

- Источники данных
- Хранилище данных
- Средства обработки и анализа

## Процесс

- Сбор данных, хранение в сыром виде, обработка по необходимости



# Преимущества озер данных

## Преимущества

- Гибкость в работе с данными
- Масштабируемость
- Низкие затраты на хранение

## Недостатки

- Сложность в управлении и организации данных
- Риски безопасности и качества данных



# Витрины данных (Data Marts)

## Определение

- Оптимизированное хранилище данных для конкретных бизнес-процессов или подразделений

## Особенности

- Структурированные и агрегированные данные
- Быстрый доступ и обработка запросов

## Примеры использования

- Бизнес-аналитика, отчеты, поддержка решений



# Архитектура витрины данных

## Компоненты

- Источники данных
- Средства ETL (Extract, Transform, Load)
- Хранилище данных
- Средства анализа и визуализации

## Процесс

- Извлечение данных, трансформация, загрузка в витрину данных



# Преимущества витрин данных

## Преимущества

- Оптимизация для бизнес-анализа
- Улучшенная производительность запросов
- Поддержка конкретных бизнес-потребностей

## Недостатки

- Ограниченность по объему и типам данных
- Необходимость ETL-процессов





# Сравнение озер данных и витрин данных

## Основные различия

- Структура данных: сырые данные vs. структурированные данные
- Цель использования: гибкость vs. оптимизация для анализа
- Управление данными: сложность vs. упрощенное управление

## Критерии выбора

- Тип данных, объем данных, потребности в анализе

# Интеграция озер данных и витрин данных

## Подходы к интеграции

- Использование озера данных в качестве источника для витрин данных
- Совместное использование для различных задач анализа

## Преимущества интеграции

- Комбинирование гибкости и производительности
- Повышение эффективности аналитических процессов



# Пример интеграции

## Процесс интеграции

- Сбор данных в озеро данных
- Трансформация и загрузка данных в витрины данных для анализа

## Пример сценария

- Аналитика клиентских данных: сбор из разных источников, обработка и анализ в витринах данных



# Практические примеры и кейсы

Реальные примеры использования озер данных и витрин данных

- Аналитика маркетинговых данных, машинное обучение, отчетность в реальном времени



# Инструменты и технологии

## Инструменты для озер данных

- Apache Hadoop, Amazon S3, Microsoft Azure Data Lake

## Инструменты для витрин данных

- Microsoft SQL Server, Amazon Redshift, Google BigQuery



# Лучшие практики и рекомендации

## Управление данными

- Организация и управление данными в озерах данных
- Оптимизация ETL-процессов для витрин данных

## Безопасность и качество данных

- Обеспечение безопасности данных
- Поддержание качества данных



# Заключение

## Подведение итогов

- Важность выбора подходящего подхода к хранению данных
- Роль озер данных и витрин данных в современной аналитике

## Дополнительные материалы и литература

# Вопросы и обсуждение

Вопросы от студентов

Обсуждение примеров





# Благодарность за внимание

## Контактная информация

- Email: [iradche@gmail.com](mailto:iradche@gmail.com)
- Телеграм: [@dadaistka](https://t.me/dadaistka)

## Следующая лекция

- Тема следующей лекции:  
«Популярные NoSQL хранилища и СУБД: Scylla/Cassandra, HDFS, HBase, Redis, Memcached, ElasticSearch/Solr, MongoDB, ClickHouse»

# Спасибо за внимание!

[www.ifmo.ru](http://www.ifmo.ru)

IT'sMO *re than a*  
UNIVERSITY