



УНИВЕРСИТЕТ ИТМО

Технологии хранения больших данных

Лекция №8. Популярные NoSQL хранилища и СУБД:
Scylla/Cassandra, HDFS, HBase, Redis, Memcached,
ElasticSearch/Solr, MongoDB, ClickHouse

Ирина Алексеевна Радченко
iradche@gmail.com



Введение

Обзор темы

- Значение NoSQL-хранилищ и СУБД в современных системах
- Введение в различные типы NoSQL баз данных

Цели лекции

- Понимание ключевых характеристик популярных NoSQL-систем
- Изучение примеров использования и возможностей каждой системы

Scylla/Cassandra

Общие характеристики

- Колонко-ориентированные базы данных
- Высокая масштабируемость и производительность

ScyllaDB

- Совместима с Cassandra, но с большей производительностью

Примеры использования

- Реализация распределенных систем, аналитика в реальном времени



Пример запроса в БД Cassandra

```
SELECT name, age FROM users WHERE user_id = '12345';
```



HDFS (Hadoop Distributed File System)

Общие характеристики

- Дистрибутивная файловая система для больших данных
- Высокая отказоустойчивость и масштабируемость

Использование

- Хранение и обработка больших объемов данных

Примеры использования

- Аналитика данных, машинное обучение

HBase

Общие характеристики

- Колонко-ориентированная база данных поверх HDFS
- Поддержка случайного чтения и записи больших данных

Примеры использования

- Аналитические приложения, хранение больших таблиц

Пример запроса в БД HBase

```
Get get = new Get(Bytes.toBytes("row1"));  
Result result = table.get(get);
```

Redis

Общие характеристики

- База данных типа ключ-значение
- Высокая производительность и поддержка различных структур данных

Примеры использования

- Кэширование, очереди сообщений, сеансовое хранение



Пример команды в Redis

- SET user:1000 "John Doe"
- GET user:1000

```
SELECT name, COUNT(*) FROM users GROUP BY name;
```



Memcached

Общие характеристики

- Система кэширования типа ключ-значение
- Высокая производительность и простота использования

Примеры использования

- Кэширование веб-страниц, уменьшение нагрузки на базу данных



Пример команды в Memcached

```
set user:1000 0 900 "John Doe"  
get user:1000
```



ElasticSearch/Solr

Общие характеристики

- Полнотекстовый поиск и аналитика
- Распределенные поисковые системы

Примеры использования

- Поиск по сайтам, аналитика логов, мониторинг данных

Пример запроса в Elasticsearch

```
GET /users/_search
{
  "query": {
    "match": {
      "name": "John"
    }
  }
}
```



MongoDB

Общие характеристики

- Документо-ориентированная база данных
- Высокая гибкость и масштабируемость

Примеры использования

- Веб-приложения, хранение JSON-документов

Пример запроса в MongoDB

```
db.users.find({ "name": "John Doe" })
```



ClickHouse

Общие характеристики

- Колонко-ориентированная СУБД для аналитики
- Высокая производительность при обработке больших объемов данных

Примеры использования

- Веб-аналитика, бизнес-отчеты, мониторинг



Пример запроса в ClickHouse

```
SELECT name, COUNT(*) FROM users GROUP BY name;
```



Сравнение NoSQL СУБД

Ключевые характеристики

- Типы данных, производительность, масштабируемость

Критерии выбора

- Потребности в хранении данных, требования к производительности, специфика использования



Практические примеры и кейсы

Реальные примеры использования различных NoSQL СУБД

- Аналитика, кэширование, полнотекстовый поиск



Инструменты и технологии

Инструменты для работы с NoSQL базами данных

- Apache Kafka, Hadoop, Spark

Интеграция и использование

- Инструменты ETL, визуализация данных



Лучшие практики и рекомендации

Оптимизация производительности

- Настройка индексов, кэширование

Безопасность данных

- Шифрование, контроль доступа



Заключение

Подведение итогов

- Важность выбора подходящей NoSQL СУБД для конкретной задачи
- Роль NoSQL систем в современной аналитике и хранении данных

Дополнительные материалы и литература

Вопросы и обсуждение

Вопросы от студентов

Обсуждение примеров

Благодарность за внимание

Контактная информация

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

Спасибо за внимание!

www.ifmo.ru

IT'sMO *re than a*
UNIVERSITY