

Нереляционные базы данных

Длительность курса: 128 академических часов

1 Введение в NOSQL

1 Введение в NOSQL. NOSQL vs RDBMS

объяснить разницу между NoSQL и RDBMS решениями;
сравнить NoSQL и RDBMS;
проанализировать историю NoSQL;
рассмотреть программу курса.

2 CAP теорема

объяснить глубину дилеммы на примере проблемы византийских генералов;
разбираться в алгоритме консенсуса.

Домашние задания

1 CAP теорема

Цель: В результате выполнения ДЗ вы научитесь работать с гитом.

Необходимо написать к каким системам по CAP теореме относятся перечисленные БД и почему:
mongoDB, MSSQL, Kassandra.

ДЗ сдается ссылкой на гит, где расположен миниотчет в маркдауне.

- 1 Базовые возможности mongodb**

сможет развернуть БД mongo разными способами;
писать простейшие запросы к данным.

Домашние задания

 - 1 MongoDB

Цель: В результате выполнения ДЗ вы научитесь разворачивать MongoDB, заполнять данными, создавать индексы и делать запросы.

Необходимо:

 - установить MongoDB одним из способов: VM, докер;
 - заполнить данными;
 - создать индексы и сравнить производительность.

Сдача ДЗ осуществляется в виде миниотчета.

- 2 Кластерные возможности mongodb**

шардировать в mongodb: mongos;
правильно выбрать ключ шардирования.

- 3 map-reduce framework**

строить сложные запросы;
писать свои функции.

- 4 Утилиты mongodb**

мониторить работающий кластер;
находить тонкие места;
обеспечивать безопасность.

- 5 оптимизация производительности mongodb**

построить индексы;
оптимизировать CRUD.

Домашние задания

 - 1 MongoDB 2

Цель: В результате выполнения ДЗ вы настроите реплицирование и шардирование, аутентификацию в кластере и проверите отказоустойчивость.

Необходимо:

 - построить шардированный кластер из 3 кластерных нод(по 3 инстанса с репликацией) и с кластером конфига(3 инстанса);
 - добавить балансировку, нагрузить данными, выбрать хороший ключ шардирования, посмотреть как данные перебалансируются между шардами;
 - настроить аутентификацию и многоуровневой доступ;
 - поронять разные инстансы, посмотреть, что будет происходить, поднять обратно. Описать что произошло.

6 **Couchdb + pouchdb** объяснить основы работы с Couchdb + pouchdb, архитектуру и возможности.

7 **Elasticsearch** объяснить основы работы с Elasticsearch, архитектурой и описать возможности.

Домашние задания

1 Elasticsearch

Необходимо:

- установить Elasticsearch;
- создать индекс с одним полем типа "текст";
- добавить 3 записи в индекс "яблоко", "яблоко", "апельсин";
- написать запрос с условием нечеткого поиска при котором должны находиться и "яблоко" и "яблоко".

Сдавать нужно определение индекса и текст запроса.

1 **Redis. Часть 1** охарактеризовать преимущества и недостатки NoSQL баз данных по сравнению с реляционными; составить стратегии кеширования RDMBS с использованием Redis.

2 **Redis. Часть 2** охарактеризовать преимущества и недостатки NoSQL баз данных по сравнению с реляционными; описать случаи использования NoSQL; составить стратегии кеширования RDMBS с использованием Redis; применять Redis.

Домашние задания

1 Redis

Необходимо:

- сохранить большой жсон (~20МБ) в виде разных структур
- строка, hset, zset, list;
- протестировать скорость сохранения и чтения;
- предоставить отчет.

* настроить редис кластер на 3х нодах с отказоустойчивостью, затюнить таймоуты

3 **Riak**

4 **etcd**

1	Архитектура и применение cassandra	описать основы работы с Cassandra, архитектуру и возможности.
2	Язык запросов cassandra	написать свои запросы к Cassandra.
3	Масштабирование и отказоустойчивость cassandra. Часть 1	<p>выбрать replication factor; выбирать уровень консистентности.</p> <p>Домашние задания</p> <p>1 Масштабирование и отказоустойчивость Cassandra. Часть 1. - Домашнее Задание</p> <p>Цель: Подготовить среду и развернуть Cassandra кластер для дальнейшего изучения возможностей масштабирования и восстановления Cassandra кластеров.</p> <p>Необходимо:</p> <ul style="list-style-type: none">- развернуть Kubernetes кластер в облаке или локально (используя "ПРЕРЕКВИЗИТЫ.docx" из материалов);- поднять 3 узловой Cassandra кластер на Kubernetes (используя "How to Run Cassandra on Azure Kubernetes Service (AKS), part1.pdf" из материалов);- нагрузить кластер при помощи Cassandra Stress Tool (используя "How to use Apache Cassandra Stress Tool.pdf" из материалов).
4	Масштабирование и отказоустойчивость cassandra. Часть 2	<p>определить алгоритм восстановления cassandra и масштабирования кластера.</p> <p>Домашние задания</p> <p>1 Масштабирование и отказоустойчивость Cassandra. Часть 2 - Домашнее Задание</p> <p>Цель: В результате выполнения ДЗ вы изучите сторонние возможности восстановления Cassandra кластеров, замерите производительность вашего кода.</p> <p>Необходимо:</p> <ul style="list-style-type: none">- забэкапить и восстановить используя 3DSnap ваш 3 узловой Cassandra кластер на Kubernetes из прошлого ДЗ(используя "How to Run Cassandra on Azure Kubernetes Service (AKS), part2.pdf" из материалов);- выбрать два на ваш вкус драйвера к Cassandra и сравнить их по производительности и потреблению ресурсов (используя "Apache Cassandra with JPA.pdf" из материалов).

- 5 **Средства мониторинга cassandra** настроить grafana для мониторинга cassandra.
-
- 6 **Clickhouse** поймем основы работы с Clickhouse архитектура и возможности
- Домашние задания

1 ClickHouse

Необходимо, используя tutorial <https://clickhouse.tech/docs/ru/getting-started/tutorial/> :

- развернуть БД;
- выполнить импорт тестовой БД;
- выполнить несколько запросов и оценить скорость выполнения.

Дз сдается в виде миниотчета.

1 Neo4j, часть 1

2 Neo4j, часть 2

Домашние задания

1 Neo4j

Необходимо (на выбор):

- сделать базу данных для своего круга знакомых (кто с кем общается)
- сделать базу данных для авиакомпании (как куда лететь и за какую цену)

- 1 **InterSystems Caché** объяснить основы работы с Clickhouse, архитектуру и возможности

Домашние задания

1 Cache

Необходимо:

- развернуть Каше;
 - разработать БД для простейшего магазина с сущностями: заказы, позиции заказа, заказчик по методичке в материалах.
-

- 2 **Работа с tarantool** строить высоконапряженные решения на базе tarantool.

Домашние задания

1 Работа с tarantool

Необходимо написать на тарантуле биллинг реального времени облачной системы. Должны быть хранимые процедуры:

- добавление денег на баланс;
- изменение расхода денег в секунду.

Когда баланс становится равным нулю, тарантул по http должен сделать GET-запрос на какой-либо внешний урл, где передать userID пользователя, у которого кончились деньги (запрос на отключение виртуальных машин). Этот вызов должен происходить как можно быстрее после окончания денег на счете.

Для реализации рекомендуется использовать библиотеку expirationd.

Использовать шардинг на основе vshard.

- 3 **Kafka**
-

- 4 **RabbitMQ**
-

- 5 **s3**

1 **Google BigQuery** описать характеристики bigQuery - уникального аналитического хранилища данных, которое доступно только в Google Cloud

2 **AWS DynamoDB**

3 **Azure CosmosDB**

Домашние задания

1 Облака

Необходимо:

- одну из облачных БД заполнить данными (любыми из предыдущих дз);
- протестировать скорость запросов.

*сравнить 2-3 облачных NoSQL по скорости загрузки данных и времени выполнения запросов.

- 1 Выбор проектной работы**

выбрать и обсудить тему проектной работы;
спланировать работу над проектом;
ознакомиться с регламентом работы над проектом.

Домашние задания

 - 1 Проектная работа

Цель: 1 Вариант
Разработать и протестировать высоконагруженный отказоустойчивый кластер на любой NoSQL БД
2 Вариант
Взять 2+ NoSQL БД, загрузить 10+ млн записей и провести масштабное исследование по скорости обработки запросов
3 Вариант
Свой проект

- 2 Консультация по проектной работе**

получить ответы на вопросы по проекту, ДЗ и по курсу.

- 3 Защита проектной работы.
Подведение итогов курса**

защитить проект и получить рекомендации экспертов.