OTUS
онлайн-образование

Полная программа

NoSQL

NoSQL

Длительность курса: 150 часов

Модуль 1. Введение в NOSQL

Тема 1
Введение в NOSQL. NOSQL vs RDBMS

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

Краткое содержание
NoSQL и RDBMS;
история NoSQL;
программа курса;

Тема 2
CAP теорема

Цель занятия
объяснить глубину дилеммы на примере проблемы византийских генералов;
описать алгоритм консенсуса.

Краткое содержание
CAP теорема;
CP системы. RAFT, PAXOS;
PA системы. GOSSIP, CasPAXOS.

Домашние задания

CAP теорема
Цель
В результате выполнения ДЗ вы научитесь работать с гитом.

Модуль 2. Document Store

Тема 1
Базовые возможности mongodb

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

Краткое содержание
понятие коллекции, документов и полей;
CRUD-операции;
фильтрация по полям;
sort, skip, limit;
запросы по по документам.

Домашние задания

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

Тема 2
Кластерные возможности mongodb

Цель занятия
шардировать в mongodb: mongos;
правильно выбрать ключ шардирования.

Краткое содержание
варианты репликации;
Replica set;
концепция кворума;
балансировка;
подходы к шардированию в БД.

Тема 3
Mongodb. Aggregation Framework.

Цель занятия
строить сложные запросы;
научится группировать и аггрегировать данные в mongodb;
научиться объеденять коллекции;
познакомиться с geo, view и transaction в mongodb;

Краткое содержание
aggregation Framework. Знакомство и фичи.
группирование данных $group.
объединение коллекций $lookup.
свои функции аггрегирование данных.
работа с Geo данными.
mongoDB View.
mongoDB ChangeStream.
MongoDB Transaction.

Тема 4
Оптимизация производительности mongodb (ДЗ)

Цель занятия
построить индексы;
оптимизировать CRUD;
профилирование запросов

Краткое содержание
виды индексов в mongodb;
профилирование;
дисковые движки;
GridFS;
производительность кластеров.

Домашние задания

MongoDB 2
Цель
В результате выполнения ДЗ вы настроите реплицирование и шардирование, аутентификацию в кластере и проверите отказоустойчивость.

Тема 5
Couchbase (ДЗ)

Цель занятия
объяснить основы работы с Couchbase, архитектуру и возможности.

Краткое содержание
основы архитектуры;
отказоустойчивый кластер;
анализ производительности.

Домашние задания

Кластер Couchbase
Цель
В результате выполнения ДЗ вы построение отказоустойчивый кластер Couchbase.

Модуль 3. Колоночные

Тема 1
Cassandra: Distributed Key Value, Architecture

Цель занятия
описать основы работы с Cassandra, архитектуру и возможности.

Краткое содержание
CAP
Data Partition
Data Replication
Consistency Level
Inconsistency resolution
System architecture diagram

Тема 2
Cassandra: System Components (ДЗ)

Цель занятия
знать особенности архитектуры, построения кластеров и распределения данных между кольцами узлов;
использовать встроенный языку запросов CQL;

Краткое содержание
Coordinator;
Token ring;
Handling failures;
Write path;
Read Path;
Primary Key;
Column Family.

Домашние задания

Масштабирование и отказоустойчивость Cassandra
Цель
В результате выполнения ДЗ вы подготовите среду и развернете Cassandra кластер для дальнейшего изучения возможностей маштабирования и восстанавления Cassandra кластеров.

Тема 3
Clickhouse (ДЗ)

Цель занятия
описать основы работы с Clickhouse, архитектура и возможности.

Краткое содержание
когда Clickhouse полезен, а когда не стоит его использовать;
основные движки таблиц/БД Clickhouse и их особенности;
движок MergeTree и его специфика;
использование совместно с Kafka и PostgreSQL
протоколы доступа (http, tcp и базовые библиотеки);
установка и настройка.

Домашние задания

Установка ClickHouse
Цель
В результате выполнения ДЗ в развернете БД.

Тема 4
Clickhouse. Масштабирование и манипуляции с данными

Цель занятия
познакомиться с базовым SQL синтаксисом и типами данных;
разобрать основные движки БД и таблиц;
изучить принципы построения отказоустойчивого кластера, репликацию и шардирование.

Краткое содержание
разбор основных принципов производительности в ClickHouse и инструментов для ее измерения;
изучение приемов оптимизации запросов и настройки схем данных для обеспечения высокой производительности;
практика применения движков класса MergeTree Family.

Модуль 4. Key-Value

Тема 1
Redis. Часть 1

Цель занятия
охарактеризовать преимущества и недостатки NoSQL баз данных по сравнению с реляционными;
составить стратегии кеширования RDMBS с использованием Redis.

Краткое содержание
области применения;
внутренние структуры;
очереди, транзакции.

Тема 2
Redis. Часть 2 (ДЗ)

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

Краткое содержание
кластеризация;
redis sentinel;
redis cluster.

Домашние задания

Redis
Цель
В результате выполнения ДЗ вы поработаете с Redis.

Тема 3
etcd

Цель занятия
познакомиться с etcd и понять спектр ее применения.

Краткое содержание
алгоритм Raft;
отличия etcd от прочих аналогичных решений;
основные моменты установки и настройки кластера etcd;
бэкап и репликация кластера.

Тема 4
Consul (ДЗ)

Цель занятия
настроить DNS с healthcheck на базе консула.

Краткое содержание
установка;
настройка;
кластеризация;
федерация;
бэкап;
service discovery;
DNS.

Домашние задания

DCS
Цель
В результате выполнения ДЗ вы развернете отказоустойчивый DCS.

Тема 5
Foundation db

Цель занятия
научиться работать с foundation db.

Краткое содержание
foundation db - обзор и архитектура
транзакции;
масштабируемость и отказоустойчивость;
ограничения.

Тема 6
Занятие в формате вопрос-ответ

Цель занятия
получить ответы на вопросы по темам 1 и 2 модуля.

Краткое содержание
ответы на вопросы по темам занятий и домашних заданий.

Модуль 5. Графовые

Тема 1
Neo4j, часть 1 (ДЗ)

Цель занятия
рассмотреть особенности графовых баз данных;
использовать графовые базы данных в продакшене;
тюнинг графовой БД.

Краткое содержание
разминка: графы;
способы структурирования графов;
Neo4j в действии.

Домашние задания

Сравнение с неграфовыми БД
Цель
В результате выполнения ДЗ вы поймете зачем и когда может понадобиться графовая база данных.

Тема 2
Neo4j, часть 2 (ДЗ)

Цель занятия
рассмотреть основные юзкейсы для графовой БД;
рассмотреть архитектуру Neo4J.

Краткое содержание
Cypher: tips and tricks;
обзор графов х БД;
Neo4J: Operations side.

Домашние задания

Использование neo4j
Цель
1. Научиться представлять данные в neo4j
2. Научиться использовать Cypher для манипуляции данными в neo4j
3. Научиться смотреть план запроса
4. Создание индексов

Тема 3
Elasticsearch (ДЗ)

Цель занятия
объяснить основы работы с Elasticsearch, архитектуру и возможности.

Краткое содержание
основные модели использования ElasticSearch;
архитектура и компоненты ElasticSearch;
происхождение "ELK стек" и его актуальность;
установка и настройка ElasticSearch;
создание и настройка индексов;
типы запросов ElasticSearch и особенности их применения;
Kibana - основы интерфейса и использования.

Домашние задания

Знакомство с ElasticSearch
Цель
В результате выполнения ДЗ вы научитесь разворачивать ES в AWS и использовать полнотекстовый нечеткий поиск

Тема 4
Opensearch

Цель занятия
научиться использовать основные компоненты OpenSearch;
ориентироваться в экосистеме этой БД;
проектировать сложные пайплайны для подготовки данных.

Краткое содержание
особенности OpenSearch;
работа с Data Prepper.

Модуль 6. Другие NoSQL

Тема 1
Kafka (ДЗ)

Цель занятия
объяснить архитектуру kafka, плюсы и минусы;
сравнить в другими брокерами;
познакомиться с Kafka APIs.

Краткое содержание
kafka - обзор и архитектура;
kafka Топики. Партишены и Реплики;
основные операции на кластере;
понимание кластерных метрик и того как их интерпретировать;
понимание Producer API Kafka;
понимание Consumer API Kafka;
понимание Streams API Kafka;
возможность интеграции Kafka с внешними системами данных с использованием Kafka Connect.

Домашние задания

Kafka
Цель
научится запускать Kafka, работать с ее утилитами;

Тема 2
RabbitMQ (ДЗ)

Цель занятия
конфигурировать rabbitMQ кластер.

Краткое содержание
протокол AMPQ;
задача и функциональность брокера сообщений;
особенности кластера rabbitMQ;
запуск кластера;
проверка отказоустойчивости.

Домашние задания

Работаем с RabbitMQ
Цель
научиться запускать RabbitMQ, работать через UI в браузере и с использованием программного кода

Тема 3
Hadoop File System

Цель занятия
рассмотреть основную идея и архитектуру распределенной файловой системы;
рассмотреть форматы данных и их особенности + знакомство с фичами Hive SQL доступа.

Краткое содержание
Hadoop File System;
форматы данных;
SQL + Hive;
bash + Java/Scala.

Тема 4
s3 (ДЗ)

Цель занятия
понять, что такое S3;
увидеть реализации S3;
рассмотреть дополнительные возможности версий S3.

Краткое содержание
протокол S3;
CEPH;
фичи и ограничения s3: версионирование и другие.

Домашние задания

Запустить сервис S3
Цель
научиться работать с S3 подобными сервисами;

Тема 5
Prometheus. Grafana

Цель занятия
prometheus;
grafana;
alertManager;
promQL.

Краткое содержание
метрики Prometheus;
PromQL;
service Monitors;
demo.

Тема 6
Архитектура Tarantool (ДЗ)

Цель занятия
использовать возможности Tarantool;

Краткое содержание
архитектура ядра Tarantool;
Tarantool как БД и Сервер приложений;
сравнение с другими БД;
кейсы использования;
установка и процесс запуска Tarantool;
основные операции с данными.

Домашние задания

Работа с tarantool
Цель
настроить базовое взаимодействие с Tarantool для хранения и выборки данных, используя индексы и встроенные функции.

Тема 7
Масштабирование Tarantool

Цель занятия
масштабировать Tarantool и настраивать репликацию.

Краткое содержание
масштабирование Tarantool;
репликация в Tarantool;
хранение данных при масштабировании;
устройство асинхронной репликации;
устройство синхронной репликации;
шардирование.

Тема 8
NATS

Цель занятия
познакомить с возможностями сервиса NATS;
рассмотреть варианты использования и настроить кластер.

Краткое содержание
что такое NATS;
варианты использования;
JetStream;
запускаем кластер;
пример подключения и использования.

Модуль 7. NoSQL в облаках

Тема 1
Google BigQuery

Цель занятия
описать характеристики bigQuery - уникального аналитического хранилища данных, которое доступно только в Google Cloud.

Краткое содержание
обзор NoSQL сервисов управления структурированными данными в Google Cloud Platform;
демонстрация уникальных возможностей Google BigQuery.

Тема 2
Nosql в Яндекс облаке (ДЗ)

Цель занятия
разобрать варианты NoSQL.

Краткое содержание
узнаем, какие существуют nosql решения в ЯО.

Домашние задания

Облака
Цель
В результате выполнения ДЗ вы поработает с облаками.

Тема 3
Nosql в cloud

Цель занятия
разобрать варианты nosql в cloud.

Краткое содержание
узнаем, какие существуют nosql решения в cloud.

Модуль 8. Проект

Тема 1
Выбор проектной работы

Цель занятия
выбрать и обсудить тему проектной работы;
спланировать работу над проектом;
знакомиться с регламентом работы над проектом.

Краткое содержание
правила работы над проектом и специфика проведения итоговой защиты;
требования к результату проекта и итоговой документации.

Домашние задания

Проектная работа
Цель
Защитить проект и получить рекомендации экспертов.

Тема 2
Консультация по проектной работе

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

Краткое содержание
вопросы по улучшению и оптимизации работы над проектом;
затруднения при выполнении ДЗ;
вопросы по программе.

Тема 3
Подведение итогов курса и защита проектной работы

Цель занятия
защитить проект и получить рекомендации экспертов;
узнать, как получить сертификат об окончании курса, как взаимодействовать после окончания курса с OTUS и преподавателями, какие вакансии и позиции есть для выпускников и на какие компании стоит обратить внимание.

Краткое содержание
презентация проектов перед комиссией;
вопросы и комментарии по проектам.
организационные вопросы;
рынок вакансий по направлению;
статистика курса и вопросы по курсу.