

Вопросы для подготовки к экзамену

1. Что такое большие данные. Характеристики больших данных (3v, 4v)
2. В чем отличие систем хранения больших данных (масштабируемость, отказоустойчивость)
3. Что такое Hadoop. Основные компоненты Hadoop (HDFS, MapReduce)
4. Файловая система распределенного хранения больших данных HDFS
5. Модель распределенных вычислений MapReduce
6. Узлы Hadoop кластера (DataNode, NameNode, TaskTracker, JobTracker)
7. Отказоустойчивость, что если упадет DataNode, NameNode, TaskTracker, JobTracker
8. Основные проблемы Hadoop 1. Основные отличия Hadoop 2 от Hadoop 1
9. Hadoop 2. Изменения в HDFS
10. Hadoop 2. Архитектура Yarn
11. Языки высокого уровня над Hadoop (Pig, Hive)
12. Что такое Spark, основная идея
13. Зачем нужны NoSQL базы данных
14. Типы NoSQL баз данных
15. Модель данных HBase
16. Основные этапы извлечения информации из текстов
17. Подходы к распознаванию сущностей из текстов
18. Язык AQL. Компоненты языка AQL