

**Вопросы к экзамену по курсу
«Управление разно-структурированными большими данными»:**

1. Что такое большие данные (3v, 4v)
2. Чем знаменит IBM Watson
3. Что называют традиционными архитектурами хранения данных
4. В чем отличие систем хранения больших данных (масштабируемость, отказоустойчивость)
5. Что такое Hadoop (HDFS+MapReduce)
6. На каком языке написан Hadoop
7. Какие технологии, связанные с Hadoop, Вы знаете
8. Хорошо ли использовать Hadoop для транзакций
9. Что такое блок в hadoop
10. Что такое репликация
11. Назовите любую команду для работы с Hadoop
12. Какие узлы есть в Hadoop кластере
13. Что такое NameNode
14. Что такое JobTracker
15. TaskTracker
16. DataNode
17. Что на вход принимает map
18. Что на вход принимает reduce
19. Что происходит между Map и reduce
20. Отказоустойчивость, что если упадет DataNode, NameNode, TaskTracker, JobTracker
21. Зачем нужен Combiner
22. Какие языки высокоуровневого программирования Вы знаете
23. Что такое Pig
24. Что такое Hive, пример команды на Hive
25. Что такое Jaql
26. Что такое Yarn, основная идея
27. Что такое Spark, основная идея
28. Зачем нужны NoSQL базы данных
29. Модель данных HBase
30. Какие средства для распределенной обработки данных на языке R вы знаете
31. Основные этапы извлечения информации из текстов
32. Подходы к распознаванию сущностей из текстов
33. Язык AQL. Компоненты языка AQL