**Список вопросов для подготовки к экзамену по дисциплине «Системы и технологии интеллектуальной обработки данных»**

1. Причины возникновения, структура и особенности хранилищ данных
2. Свойства и модели хранилищ данных. OLAP, пример использования
3. Основные понятия OLAP, архитектура
4. Понятие datamining, задачи , основные этапы
5. Методы datamining, области применения
6. Метод Деревья решений, алгоритмы реализации
7. Задача классификации, алгоритмы реализации
8. Кластеризация, общая схема
9. Алгоритмы кластеризации
10. Особенности проектов datamining. Методика извлечения знаний (KDD)
11. Рабочий цикл проектов datamining, Общая схема анализа
12. Подходы к обработке больших объемов данных в datamining, Способы повышения производительности
13. Типовая структура аналитических систем
14. Назначение и типовая структура BI – решений.
15. ETL – назначение, примеры использования
16. Рынок BigData, области применения
17. Определение BigData. отличия от других систем хранения и обработки данных
18. Типовые архитектурные решения BigData для предприятий
19. Проблемы и подходы к масштабированию распределенных систем хранения данных
20. Задачи **Batch Processing в** BigData.
21. Архитектура ядра Hadoop, основные компоненты
22. Жизненный цикл обработки на кластерах Hadoop
23. Типовая Архитектура кластера Hadoop
24. Сравнение Hadoop и реляционных СУБД
25. Паттерн MapReduce , особенности использования, терминология
26. Архитектура Hadoop Distributed File System, организация доступа к файлам
27. Использование PIG, HIVE в Hadoop
28. Примеры Задач для Big Data Machine Learning
29. Свойства **ACID для реляционных СУБД**
30. **Суть и определение NoSQL систем**
31. Теорема **CAP**
32. **Сравнение реляционных СУБД и NoSQL**
33. Классы задач для использования **NoSQL**
34. **Ключевые особенности NoSQL**
35. **Виды NoSQL-систем, Key-Value системы**
36. **Виды NoSQL-систем, Document-based системы**
37. **Виды NoSQL-систем, Column-based системы**
38. **Виды NoSQL-систем, Graph-based системы область применения**
39. Near real-time системы обработки данных (Storm, Spark): архитектура, примеры задач