

По умолчанию для Аналитическая обработка данных в задачах информационной безопасности (09.04.01, 2021)/ПК-30.1

1. Apache Spark может рассматриваться, как:

- a. Фреймворк для обслуживания всего цикла обработки данных
- b. Кластерное, распределенное приложение
- c. Всё перечисленное

Apache Spark может рассматриваться, как: (Множественный выбор / Только один ответ)

2. Spark Streaming реализует следующий подход к обработке данных

- a. распределенный
- b. микро-пакетный
- c. потоковый

Spark Streaming реализует следующий подход к обработке данных (Множественный выбор / Только один ответ)

3. Алгоритмы нейронных сетей, входы и выходы которых могут принимать только значения {0, 1}

- a. Бинарные алгоритмы
- b. Древовидные алгоритмы
- c. Алгоритмы Monoplan

Алгоритмы нейронных сетей, входы и выходы которых могут принимать только значения (0, 1) (Множественный выбор / Только один ответ)

4. Анализ тональности это

- a. обработка музыки
- b. обработка цвета
- c. обработка естественного языка

Анализ тональности это (Множественный выбор / Только один ответ)

5. Байесовская сеть использует следующую вероятностную схему

- a. Условная вероятность
- b. Бинарная вероятность
- c. Не является вероятностной моделью

Байесовская сеть использует следующую вероятностную схему (Множественный выбор / Только один ответ)

6. В какой модели машинного обучения модель должна приспосабливаться к новым, поступающим данным?

- a. Модель с учителем
- b. Модель с частичным обучением
- c. Модель с динамическим обучением

В какой модели машинного обучения модель должна приспосабливаться к новым, поступающим данным?
(Множественный выбор / Только один ответ)

- 7. В какой структуре данных Спарк нужно определять схему данных вручную?**
- Dataset
 - RDD
 - DataFrame

В какой структуре данных Спарк нужно определять схему данных вручную (Множественный выбор / Только один ответ)

- 8. В чём отличие системы хранения и анализа данных Hadoop от реляционной СУБД? Выберите все подходящие варианты ответа.**
- Реляционные СУБД используют как интерактивный, так и пакетный режим доступа, в то время как Hadoop - только пакетный.
 - Реляционные СУБД позволяют выполнять многократное чтение и запись данных, в то время как Hadoop - однократную запись и многократное чтение.
 - Оба перечисленных

В чём отличие системы хранения и анализа данных Hadoop от реляционной СУБД? Выберите все подходящие варианты ответа. (Множественный выбор / Только один ответ)

- 9. В чём состоит свойство расширяемости записей СУБД?**
- В любую таблицу СУБД можно добавить новую колонку, предварительно изменив структуру этой таблицы
 - Повышение отказоустойчивости системы при добавлении новых записей в СУБД
 - СУБД имеет чёткую, но расширяемую структуру, в каждую запись можно добавить новую колонку, так же как и узнать значение любой записи по добавленной колонке

В чём состоит свойство расширяемости записей СУБД? (Множественный выбор / Только один ответ)

- 10. Возможно ли в рамках проекта Hadoop запустить задание (Job), написанное на языке отличном от Java?**
- Да
 - Нет
 - Такой функции нет

Возможно ли в рамках проекта Hadoop запустить задание (Job), написанное на языке отличном от Java?
(Множественный выбор / Только один ответ)

- 11. Где хранится информация о местонахождении блоков в файловой системе HDFS?**
- На узле имён
 - На узлах данных в форме списков блоков, которые на них хранятся
 - В локальной файловой системе вторичного узла имён (Secondary Name Node)

Где хранится информация о местонахождении блоков в файловой системе HDFS? (Множественный выбор / Только один ответ)

- 12. Для задач машинного обучения механизм shuffle является неотъемлемой частью:**
- Алгоритма Expectation Maximization

- b. Алгоритма word2vec
- c. Оба варианта правильные

Для задач машинного обучения механизм shuffle является неотъемлемой частью (Множественный выбор / Только один ответ)

13. Для какого вида нейронных сетей требуется значительное число обучающих данных для приемлемого результата?

- a. Персептрон
- b. Многослойный персептрон
- c. RBF-сети

Для какого вида нейронных сетей требуется значительное число обучающих данных для приемлемого результата? (Множественный выбор / Только один ответ)

14. Для какого шаблона взаимодействия с данными была разработана файловая система HDFS?

- a. Один раз записали, один раз прочитали
- b. Много раз поменяли данные, много раз прочитал
- c. Один раз записали, много раз прочитали

Для какого шаблона взаимодействия с данными была разработана файловая система HDFS? (Множественный выбор / Только один ответ)

15. Для чего предназначена технология HDFS Federation?

- a. Для записи функций отображения и свёртки на других языках, отличных от Java
- b. Для использования сокетов в качестве каналов, по которым трекер задач взаимодействует с процессами
- c. Для масштабирования кластера за счёт добавления узлов имён, каждый из которых управляет частью пространства имён файловой системы

Для чего предназначена технология HDFS Federation? (Множественный выбор / Только один ответ)

16. Есть ли в Spark интерфейс SQL?

- a. Да
- b. Нет
- c. Находится в стадии разработки

Есть ли в Spark интерфейс SQL? (Множественный выбор / Только один ответ)

17. Зачем нужен механизм Graceful Shutdown?

- a. обеспечивает автоматический перезапуск Spark-заданий в случае сбоя
- b. организует непрерывное выполнение задач Spark-конвейера
- c. реализует плавное завершение текущих задач Spark-приложения, предотвращая потерю данных из-за немедленной остановки контекста в случае сбоя или запланированного перерыва

Зачем нужен механизм Graceful Shutdown? (Множественный выбор / Только один ответ)

18. Имеет ли практическое применение теорема Колмогорова для реализации нейронных сетей?

- a. Да
- b. Нет
- c. Не имеет отношения к нейронным сетям

Имеет ли практическое применение теорема Колмогорова для реализации нейронных сетей? (Множественный выбор / Только один ответ)

19. Интегрировать приложение Apache Spark с внешними хранилищами данных для считывания из них нужной информации можно с помощью

- a. импорта и экспорта файлов и таблиц с данными
- b. Коннекторов Spark SQL API
- c. напрямую через механизм SQL-запросов обмена сообщениями

Интегрировать приложение Apache Spark с внешними хранилищами данных для считывания из них нужной информации можно с помощью (Множественный выбор / Только один ответ)

20. Информация, которую система воспринимает от окружающей среды называется

- a. входной
- b. выходной
- c. внутрисистемной

Информация, которую система воспринимает от окружающей среды называется (Множественный выбор / Только один ответ)

21. Какая из нижеперечисленных нейронных сетей есть сеть с обратными связями?

- a. Сеть Кохонена
- b. Сеть Хемминга
- c. Выходная звезда Гроссберга

Какая из нижеперечисленных нейронных сетей есть сеть с обратными связями? (Множественный выбор / Только один ответ)

22. Какая из следующих СУБД подходит для организации высоко-доступного и распределённого хранилища

- a. Cassandra
- b. CouchDB
- c. Обе перечисленные

Какая из следующих СУБД подходит для организации высоко-доступного и распределённого хранилища (Множественный выбор / Только один ответ)

23. Какая компания разработала открытую библиотеку CatBoost?

- a. Amazon
- b. Google
- c. Яндекс

Какая компания разработала открытую библиотеку CatBoost? (Множественный выбор / Только один ответ)

24. Какая структура данных в Apache Spark обрабатывается медленнее

- a. DataFrame
- b. Dataset
- c. RDD

Какая структура данных в Apache Spark обрабатывается медленнее (Множественный выбор / Только один ответ)

25. Какая структура данных Спарк подойдет для низкоуровневых преобразований неструктурированных данных на языке R или Python

- a. Dataset
- b. RDD
- c. DataFrame

Какая структура данных Спарк подойдет для низкоуровневых преобразований неструктурированных данных на языке R или Python (Множественный выбор / Только один ответ)

26. Какая технология не относится к технологиям больших данных?

- a. Распределенное программирование
- b. Сбор данных
- c. Машинное обучение

Какая технология не относится к технологиям больших данных? (Множественный выбор / Только один ответ)

27. Какие модели используются в компьютерном зрении?

- a. Мульти-эмбединги
- b. Графы сцены
- c. Рекурсивные сети

Какие модели используются в компьютерном зрении? (Множественный выбор / Только один ответ)

28. Какие подходы наиболее подходят для машинного перевода

- a. Языковые модели
- b. Статистические модели
- c. Гибридные модели

Какие подходы наиболее подходят для машинного перевода (Множественный выбор / Только один ответ)

29. Каков размер блока HDFS по умолчанию?

- a. 512 байт
- b. 128 Мбайт
- c. 1024 Мбайт

Каков размер блока HDFS по умолчанию? (Множественный выбор / Только один ответ)

30. Какова основная функция вторичного узла имён (Secondary Name Node)?

- a. Периодическое слияние образа пространства имён с журналом изменений.
- b. Резервное восстановление функциональности узла имён после отказа главного узла имён.
- c. Хранение дополнительных метаданных кластера.

Какова основная функция вторичного узла имён (Secondary Name Node)? (Множественный выбор / Только один ответ)

31. Какое приложение экосистемы Hadoop отвечает за систему управления ресурсами кластеров?

- a. YARN
- b. HCatalog
- c. Flume

Какое приложение экосистемы Hadoop отвечает за систему управления ресурсами кластеров? (Множественный выбор / Только один ответ)

32. Какое приложение экосистемы Hadoop отвечает за столбцовое хранение данных?

- a. Oozie
- b. Hive
- c. HBase

Какое приложение экосистемы Hadoop отвечает за столбцовое хранение данных? (Множественный выбор / Только один ответ)

33. Какой инструмент входит в инфраструктуру Spark?

- a. Spark Streaming
- b. GraphX
- c. Оба перечисленных инструмента

Какой инструмент входит в инфраструктуру Spark? (Множественный выбор / Только один ответ)

34. Какой принцип обучения использует сеть Кохонена?

- a. С учителем
- b. Без учителя
- c. Метод критики

Какой принцип обучения использует сеть Кохонена? (Множественный выбор / Только один ответ)

35. Какой тип NoSQL решения наиболее эффективен для потоковой обработки логов кластера серверов и быстрого сохранения без требования оперативной аналитики?

- a. Документоориентированные
- b. Ключ-значение
- c. Графовая

Какой тип NoSQL решения наиболее эффективен для потоковой обработки логов кластера серверов и быстрого сохранения без требования оперативной аналитики? (Множественный выбор / Только один ответ)

36. Какую базу данных желательно выбрать для данных со сложной структурой, имеющих небольшой размер?

- a. Графовые базы данных
- b. Реляционные базы данных
- c. Хранилище «столбец-значение»

Какую базу данных желательно выбрать для данных со сложной структурой, имеющих небольшой размер? (Множественный выбор / Только один ответ)

37. Какую функцию выполняет входной слой многослойного персептрона?

- a. Передает входной вектор сигналов на первый скрытый слой
- b. Удаляет "шум" из сигнала
- c. Вычисляет производную

Какую функцию выполняет входной слой многослойного персептрона? (Множественный выбор / Только один ответ)

38. Когда полезна предварительная обработка данных при использовании нейронных сетей?

- a. В случае числовых данных
- b. Практически всегда
- c. Не полезна

Когда полезна предварительная обработка данных при использовании нейронных сетей? (Множественный выбор / Только один ответ)

39. Колоночная БД с открытым исходным кодом, решающая вопрос хранения мелких файлов в HDFS называется

- a. Apache HBase
- b. ApacheSolr
- c. JanusGraph

Колоночная БД с открытым исходным кодом, решающая вопрос хранения мелких файлов в HDFS называется (Множественный выбор / Только один ответ)

40. Компонент Apache Spark, ответственный за контроль над расходом вычислительных ресурсов

- a. Driver
- b. Executor
- c. Node

Компонент Apache Spark, ответственный за контроль над расходом вычислительных ресурсов (Множественный выбор / Только один ответ)

41. Компонент Apache Spark, ответственный за отслеживание состояния обработки и генерацию задач

- a. Driver
- b. Executor
- c. Node

Компонент Apache Spark, ответственный за отслеживание состояния обработки и генерацию задач (Множественный выбор / Только один ответ)

42. Компонент Spark, управляющий непосредственно запуском вычислений над данными

- a. DAGScheduler
- b. RangePartitioner
- c. SparkContext

Компонент Spark, управляющий непосредственно запуском вычислений над данными (Множественный выбор / Только один ответ)

43. Матрица ошибок используется

- a. Логи ошибок
- b. Для визуализации качества алгоритма классификации
- c. Для улучшения качества метрик

Матрица ошибок используется (Множественный выбор / Только один ответ)

44. Метод построения нейронной сети, при котором исходной нейронной сетью является уже обученная сеть

- a. Метод редукции
- b. Генетический алгоритм
- c. Эволюционный метод

Метод построения нейронной сети, при котором исходной нейронной сетью является уже обученная сеть (Множественный выбор / Только один ответ)

45. Метод построения нейронной сети, при котором новые нейроны последовательно добавляются и уже не изменяются после настройки

- a. Метод NetLines
- b. Метод динамического добавления узлов
- c. Каскадная корреляция

Метод построения нейронной сети, при котором новые нейроны последовательно добавляются и уже не изменяются после настройки (Множественный выбор / Только один ответ)

46. Методика снижения размерности при предварительной обработке данных для нейронных сетей

- a. Нормализация входных данных
- b. Анализ основных компонент
- c. Ни один из перечисленных вариантов

Методика снижения размерности при предварительной обработке данных для нейронных сетей (Множественный выбор / Только один ответ)

47. Модель, представленная компанией Google используемая для параллельных вычислений над очень большими, вплоть до нескольких петабайт называется

- a. MPI
- b. OpenAcc
- c. Map reduce

Модель, представленная компанией Google используемая для параллельных вычислений над очень большими, вплоть до нескольких петабайт называется (Множественный выбор / Только один ответ)

48. Нейронная сеть от Google, позволяющая создавать программы с ИИ для обработки естественного языка

- a. BERT
- b. TensorFlow
- c. jBNC

Нейронная сеть от Google, позволяющая создавать программы с ИИ для обработки естественного языка (Множественный выбор / Только один ответ)

49. Область модели памяти Spark, размер которой жестко закодирован:

- a. Зарезервированная память
- b. Пользовательская память
- c. Память Spark

Область модели памяти Spark, размер которой жестко закодирован (Множественный выбор / Только один ответ)

50. Открытая программная библиотека, разработанная компанией Яндекс и реализующая уникальный патентованный алгоритм построения моделей машинного обучения

- a. CatBoost
- b. BERT
- c. Яндекс Толока

Открытая программная библиотека, разработанная компанией Яндекс и реализующая уникальный патентованный алгоритм построения моделей машинного обучения (Множественный выбор / Только один ответ)

51. Отметьте причины создания NoSQL баз данных:

- a. Высокая стоимость горизонтальной масштабируемости RDBMS при сохранении требования высокой доступности
- b. Невозможность хранить большие объёмы данных
- c. Дороговизна лицензий RDBMS

Отметьте причины создания NoSQL баз данных (Множественный выбор / Только один ответ)

52. Отметьте ту характеристику, которая объединяют СУБД Redis и CouchDB

- a. NoSQL
- b. Колоночная ориентированность
- c. Высокая доступность

Отметьте ту характеристику, которая объединяют СУБД Redis и CouchDB (Множественный выбор / Только один ответ)

53. Пакет Python для применения машинного обучения к большим данным

- a. NLTK
- b. Dispy
- c. Hadoop

Пакет Python для применения машинного обучения к большим данным (Множественный выбор / Только один ответ)

54. Парадигма обучения нейронных сетей, при которой для каждого примера из обучающей выборки заранее известен желаемый результат?

- a. Обучение «с учителем»
- b. Обучение «без учителя»
- c. Обучение методом критики

Парадигма обучения нейронных сетей, при которой для каждого примера из обучающей выборки заранее известен желаемый результат? (Множественный выбор / Только один ответ)

55. Перед каким этапом обработки данных фреймворком Apache Spark идет этап персистинга данных?

- a. Выгрузка данных во внешнее хранилище
- b. Загрузка данных в кластер
- c. Применение служебных преобразований

Перед каким этапом обработки данных фреймворком Apache Spark идет этап персистинга данных? (Множественный выбор / Только один ответ)

56. Повысить производительность Спарк-кластера можно с помощью увеличения

- a. объема RAM
- b. числа операций перетасовки данных (shuffle)
- c. количества исполнителей и процессорных ядер

Повысить производительность Спарк-кластера можно с помощью увеличения (Множественный выбор / Только один ответ)

57. Почему Hbase не подходит для большинства веб-приложений с высокой посещаемостью?

- a. Там требуется высокая доступность, а HBase не обладает этим свойством
- b. Из-за по колоночного хранения данных
- c. Там требуется распределённость, а HBase не обладает этим свойством

Почему Hbase не подходит для большинства веб-приложений с высокой посещаемостью? (Множественный выбор / Только один ответ)

58. При каком способе проверки модели решается проблема потери части данных для обучения?

- a. Перекрестная проверка модели
- b. Проверка с отложенными данными
- c. «Наивная» проверка

При каком способе проверки модели решается проблема потери части данных для обучения? (Множественный выбор / Только один ответ)

59. Распределенная графовая база данных с открытым исходным кодом находящаяся под The Linux Foundation

- a. InfiniteGraph
- b. JanusGraph
- c. NitrosBase

Распределенная графовая база данных с открытым исходным кодом находящаяся под The Linux Foundation (Множественный выбор / Только один ответ)

60. Распределенную работу Спарк-кластера обеспечивает

- a. Apache Zookeeper
- b. GraphX
- c. YARN

Распределенную работу Спарк-кластера обеспечивает (Множественный выбор / Только один ответ)

61. Система обработки больших данных с открытым исходным кодом называется

- a. Apache Commons
- b. Apache Hadoop

- c. Apache Tika

Система обработки больших данных с открытым исходным кодом называется (Множественный выбор / Только один ответ)

62. Сколько в приложении Apache Spark компонентов Driver?

- a. Один
- b. Три
- c. Несколько

Сколько в приложении Apache Spark компонентов Driver? (Множественный выбор / Только один ответ)

63. Сколько в приложении Apache Spark компонентов Executor?

- a. Один
- b. Два
- c. Несколько

Сколько в приложении Apache Spark компонентов Executor? (Множественный выбор / Только один ответ)

64. Spark-приложения выполняют вычисления MapReduce

- a. в оперативной памяти
- b. в кэше
- c. на жестком диске

Spark-приложения выполняют вычисления MapReduce (Множественный выбор / Только один ответ)

65. Сфера использования модели анализа текстов BERT

- a. Проверка фактов
- b. Анализ тональности
- c. Оба варианта

Сфера использования модели анализа текстов BERT (Множественный выбор / Только один ответ)

66. Ускорить выполнение Spark SQL-запроса можно с помощью

- a. инструментов SQL-on-Hadoop, например, Apache Hive или Cloudera Impala
- b. переменных связывания (bind variable)
- c. кэширования данных

Ускорить выполнение Spark SQL-запроса можно с помощью (Множественный выбор / Только один ответ)

67. Файловая система для хранения и обработки больших объемов данных с открытым исходным кодом называется

- a. GFS
- b. NTFS
- c. HDFS

Файловая система для хранения и обработки больших объемов данных с открытым исходным кодом называется (Множественный выбор / Только один ответ)

68. Фреймворк с открытым исходным кодом для реализации распределённой обработки неструктурированных и слабоструктурированных данных, входящий в экосистему проектов Hadoop

- a. Apache Kafka
- b. Apache Spark
- c. Apache Nifi

Фреймворк с открытым исходным кодом для реализации распределённой обработки неструктурированных и слабоструктурированных данных, входящий в экосистему проектов Hadoop (Множественный выбор / Только один ответ)

69. Цифровой след личности с течением времени...

- a. Не меняется
- b. Может изменяться
- c. Невозможно узнать

Цифровой след личности с течением времени... (Множественный выбор / Только один ответ)

70. Чему равен коэффициент репликации для каталогов в файловой системе HDFS?

- a. 3
- b. Каталоги не реплицируются
- c. 1

Чему равен коэффициент репликации для каталогов в файловой системе HDFS? (Множественный выбор / Только один ответ)

71. Что необходимо сделать на первом шаге использования моделей машинного обучения с учителем?

- a. Выбрать гиперпараметры модели
- b. Выбрать класс модели
- c. Перечисленные варианты не подходят

Что необходимо сделать на первом шаге использования моделей машинного обучения с учителем? (Множественный выбор / Только один ответ)

72. Что такое HadoopStreaming?

- a. Способ написания Java программ в модели вычислений MapReduce, при котором выходные данные одной программы являются входом для другой
- b. Технология, использующая стандартные потоки ввода/вывода Unix для организации взаимодействия Hadoop с другими программами.
- c. Такое понятие как HadoopStreaming вообще не существует.

Что такое HadoopStreaming? (Множественный выбор / Только один ответ)

73. Что такое HDFS?

- a. Распределённая столбцово-ориентированная база данных.
- b. Язык управления потоком данных и исполнительная среда для анализа очень больших наборов данных.
- c. Распределённая файловая система, работающая на больших кластерах из стандартных машин.

Что такое HDFS? (Множественный выбор / Только один ответ)

74. Эффективность модели машинного обучения на проверочном наборе данных сопоставима с ее эффективностью на обучающей последовательности:

- a. Для моделей с высокой дисперсией
- b. Для моделей со значительной систематической ошибкой
- c. Оба варианта правильные

Эффективность модели машинного обучения на проверочном наборе данных сопоставима с ее эффективностью на обучающей последовательности (Множественный выбор / Только один ответ)

75. Является ли модель программирования MapReduce линейно масштабируемой?

- a. Да
- b. Нет
- c. Понятие не применимо к данной модели

Является ли модель программирования MapReduce линейно масштабируемой? (Множественный выбор / Только один ответ)

По умолчанию для Аналитическая обработка данных в задачах информационной безопасности (09.04.01, 2021)/ПК-30.2

1. Apache NiFi используется для

- a. Маршрутизации потоков Big Data и построения ETL-конвейеров
- b. Оптимизации SQL-запросов к DWH
- c. Эффективного хранения больших данных

Apache NiFi используется для (Множественный выбор / Только один ответ)

2. Hadoop это

- a. Набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах
- b. Распределённая СУБД, позволяющая обрабатывать большие данные
- c. Распределённая файловая система, предназначенная для хранения файлов большого объёма

Hadoop это (Множественный выбор / Только один ответ)

3. TensorFlow — фреймворк, который хорошо подходит для

- a. Описание изображений
- b. Обучение нейросетей
- c. Оба варианта

TensorFlow — фреймворк, который хорошо подходит для (Множественный выбор / Только один ответ)

4. Автоматизировать запуск пакетных задач в рамках конвейера обработки больших данных по расписанию можно с

- a. Apache Hive
- b. Apache Hadoop
- c. Apache AirFlow

Автоматизировать запуск пакетных задач в рамках конвейера обработки больших данных по расписанию можно с (Множественный выбор / Только один ответ)

5. Алгоритм Хебба используется следующим образом

- a. Обучение «с учителем»
- b. Обучение «без учителя»
- c. Обучение методом критики

Алгоритм Хебба используется следующим образом (Множественный выбор / Только один ответ)

6. Анализировать данные, хранящиеся в Apache Hadoop, с помощью стандартного инструментария SQL-запросов

- a. Можно
- b. Нельзя
- c. Не представляется возможности

Анализировать данные, хранящиеся в Apache Hadoop, с помощью стандартного инструментария SQL-запросов (Множественный выбор / Только один ответ)

7. В какие игры нейросеть еще не научилась обыгрывать человека?

- a. Го
- b. «Марио»
- c. Бридж

В какие игры нейросеть еще не научилась обыгрывать человека? (Множественный выбор / Только один ответ)

8. В какой фазе технологии MapReduce документы делятся на пары «ключ-значение»?

- a. Фаза отображения
- b. Фаза свертки
- c. Ни в одной из перечисленных

В какой фазе технологии MapReduce документы делятся на пары «ключ-значение»? (Множественный выбор / Только один ответ)

9. В какой фазе технологии MapReduce экземпляры группируются?_____ –

- a. Фаза отображения
- b. Фаза свертки
- c. Ни в одной из перечисленных

В какой фазе технологии MapReduce экземпляры группируются? (Множественный выбор / Только один ответ)

10. В каком виде осуществляется передача данных между Map и Reduce?

- a. Ключ/значение
- b. Массив байт
- c. Текстовые данные

В каком виде осуществляется передача данных между Map и Reduce? (Множественный выбор / Только один ответ)

11. В чем заключается существенный недостаток нахождения решения с помощью логической нейронной сети?

- a. Дистрибутивное представление формального описания СПР
- b. В необходимости нахождения нейрона выходного слоя, имеющего максимальную величину возбуждения
- c. Сложность корректного формального представления системы принятия решений

В чем заключается существенный недостаток нахождения решения с помощью логической нейронной сети?
(Множественный выбор / Только один ответ)

12. В чём преимущество колоночно- ориентированных СУБД?

- a. Они позволяют выполнять более сложные SQL-запросы по сравнению с реляционными СУБД
- b. Они позволяют динамически дополнять содержание записей новыми полями
- c. Они имеют более гибкие возможности аналитики

В чём преимущество колоночно- ориентированных СУБД? (Множественный выбор / Только один ответ)

13. В чем суть закона Мура?

- a. 20% усилий дают 80% результата, а остальные 80% усилий — лишь 20% результата
- b. Каждое следующее поколение компьютеров работает в 2,5 раза быстрее
- c. Если все слова языка или длинного текста упорядочить по убыванию частоты их использования, то частота n-го слова в таком списке окажется приблизительно обратно пропорциональной его порядковому номеру n

В чем суть закона Мура? (Множественный выбор / Только один ответ)

14. Верное утверждение о RDD в Spark:

- a. RDD делится на партиции, которые являются атомарными частями данных
- b. Партиции RDD хранятся на одном сервере
- c. RDD хранит данные только в оперативной памяти

Верное утверждение о RDD в Spark (Множественный выбор / Только один ответ)

15. Выберите верное утверждение:

- a. Data Warehouse создаются для проверки гипотез при анализе больших данных
- b. "Песочница" используется для снижения нагрузки на основной Data Warehouse
- c. "Песочница" необходима для любого процесса аналитики

Выберите верное утверждение (Множественный выбор / Только один ответ)

16. Выберите верный ответ

- a. Большие данные – это обработка или хранение более 1 ТБ информации
- b. Большие данные – это огромная PR-акция крупных вендоров и не более того
- c. Проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки сущностная обработка данных затруднена или невозможна

Выберите верный ответ (Множественный выбор / Только один ответ)

17. Выберите неверное высказывание

- a. Большие объёмы данных приводят к слабой их структуризации, поэтому появляется такое разнообразие данных
- b. Увеличившаяся производительность телекоммуникационных каналов привела к росту объёмов передаваемой информации
- c. Удешевление систем хранения на единицу информации привело к росту рынка больших данных

Выберите неверное высказывание (Множественный выбор / Только один ответ)

18. Выберите неверное утверждение

- a. Алгоритм градиентного спуска позволяет изменить веса по всей сети, разворачивая гиперплоскость каждого нейрона в наилучшее положение
- b. При обучении сети мы «идем» по направлению градиента — вектора наискорейшего возрастания критериев в пространстве коэффициентов
- c. В процессе обучения нейронная сеть решает математическую задачу оптимизации

Выберите неверное утверждение (Множественный выбор / Только один ответ)

19. Выберите одно неверное высказывание про MapReduce:

- a. MapReduce был придуман разработчиками Hadoop
- b. MapReduce – это две операции: распределения и сборки данных
- c. MapReduce был анонсирован разработчиками Google

Выберите одно неверное высказывание про MapReduce (Множественный выбор / Только один ответ)

20. Выберите технологию потоковой обработки событий в режиме реального времени

- a. Spark Streaming
- b. Apache Kafka
- c. Apache Hadoop

Выберите технологию потоковой обработки событий в режиме реального времени (Множественный выбор / Только один ответ)

21. Для машинного обучения подходят данные

- a. Предварительно подготовленные, очищенные от ошибок, пропусков и выбросов, а также нормализованные и представленные в виде числовых векторов
- b. Любых форматов в цифровом виде
- c. Числовые типа int

Для машинного обучения подходят данные (Множественный выбор / Только один ответ)

22. Для полнотекстового интеллектуального поиска и аналитики по полуструктурированным данным в формате JSON отлично подходит СУБД

- a. Elasticsearch
- b. Cassandra
- c. HBase

Для полнотекстового интеллектуального поиска и аналитики по полуструктурированным данным в формате JSON отлично подходит СУБД (Множественный выбор / Только один ответ)

23. Для распределенного глубокого машинного обучения (Deep Learning) больше подходит фреймворк

- a. Flask
- b. TensorFlow
- c. Scikit-learn

Для распределенного глубокого машинного обучения (Deep Learning) больше подходит фреймворк (Множественный выбор / Только один ответ)

24. Для реализации микросервисной архитектуры и интеграции разрозненных систем подходит

- a. Apache Spark
- b. Apache Airflow
- c. Apache Kafka

Для реализации микросервисной архитектуры и интеграции разрозненных систем подходит (Множественный выбор / Только один ответ)

25. Для чего используется ранжирование?

- a. Компьютерное зрение
- b. Генерация изображений
- c. Поисковые запросы

Для чего используется ранжирование? (Множественный выбор / Только один ответ)

26. Если в модель линейной регрессии добавить не значимый признак, это может привести:

- a. К увеличению R-квадрат
- b. К уменьшению R-квадрат
- c. Ни к тому, ни к другому

Если в модель линейной регрессии добавить не значимый признак, это может привести (Множественный выбор / Только один ответ)

27. Если коэффициент корреляции Пирсона между двумя случайными величинами равен нулю, их значения всё равно могут быть связаны друг с другом. Это утверждение верно?

- a. Да
- b. Нет
- c. Не имеет смысла

Если коэффициент корреляции Пирсона между двумя случайными величинами равен нулю, их значения всё равно могут быть связаны друг с другом. Это утверждение верно? (Множественный выбор / Только один ответ)

28. Инструмент машинного обучения в инфраструктуре Spark

- a. TensorFlow
- b. MLlib
- c. Keras

Инструмент машинного обучения в инфраструктуре Spark (Множественный выбор / Только один ответ)

29. К машинным данным относятся:

- a. Запись детализации звонков
- b. Журнал сетевых событий
- c. Журнал веб-сервера
- d. Всё, перечисленное выше

К машинным данным относятся (Множественный выбор / Только один ответ)

30. Как называется задача машинного обучения, что направлена на предсказание значения той или иной непрерывной числовой величины для входных данных?

- a. Регрессия
- b. Кластеризация
- c. Переобучение

Как называется задача машинного обучения, что направлена на предсказание значения той или иной непрерывной числовой величины для входных данных? (Множественный выбор / Только один ответ)

31. Как называлась первая нейросеть?

- a. Перцептрон
- b. Оракул
- c. Гадалка

Как называлась первая нейросеть? (Множественный выбор / Только один ответ)

32. Как осуществляется доступ к файловой системе HDFS?

- a. Прямой доступ или через проху-сервер
- b. Прямой доступ
- c. Через проху-сервер

Как осуществляется доступ к файловой системе HDFS? (Множественный выбор / Только один ответ)

33. Как расшифровывается YARN?

- a. Yet Another Runner NewJob
- b. Yet Another Resource Navigator
- c. Yet Another Resource Negotiator

Как расшифровывается YARN? (Множественный выбор / Только один ответ)

34. Какая технология относится к технологиям больших данных?

- a. Моделирование
- b. Автоматизация
- c. Базы данных

Какая технология относится к технологиям больших данных? (Множественный выбор / Только один ответ)

35. Какие из следующих технологий СУБД не используют принцип MapReduce

- a. Cassandra
- b. HDInsight
- c. Redis

Какие из следующих технологий СУБД не используют принцип MapReduce (Множественный выбор / Только один ответ)

36. Какие свойства отличают большие данные от данных, встречающихся в радиционных средствах управления данными?

- a. Объем
- b. Скорость
- c. Разнообразиие
- d. Всѐ, перечисленное выше

Какие свойства отличают большие данные от данных, встречающихся в радиционных средствах управления данными? (Множественный выбор / Только один ответ)

37. Какое приложение экосистемы Hadoop отвечает за инфраструктуру распределенных вычислений?

- a. MapReduce
- b. Pig
- c. HDFS

Какое приложение экосистемы Hadoop отвечает за инфраструктуру распределенных вычислений? (Множественный выбор / Только один ответ)

38. Какое явление имеет место, если модель машинного обучения обучена шуму в данных?

- a. Переобучение
- b. Недообучение
- c. Ни одно из перечисленных

Какое явление имеет место, если модель машинного обучения обучена шуму в данных? (Множественный выбор / Только один ответ)

39. Какой алгоритм является детерминированным: PCA или k- средних?

- a. PCA
- b. k-средних
- c. Ни один из них

Какой алгоритм является детерминированным: PCA или k- средних? (Множественный выбор / Только один ответ)

40. Какой вид нейронной сети относится к сетям прямого распространения?

- a. Сеть Элмана
- b. RBF-сети
- c. ART-сети

Какой вид нейронной сети относится к сетям прямого распространения? (Множественный выбор / Только один ответ)

41. Какой вид нейронной сети относится к сетям с обратными связями?

- a. Сети Хопфида
- b. Сети каскадной корреляции
- c. Многослойный персептрон

Какой вид нейронной сети относится к сетям с обратными связями? (Множественный выбор / Только один ответ)

42. Какой главный недостаток сгенерированного нейросетью изображения?

- a. Формфактор
- b. Цвет
- c. Глубина

Какой главный недостаток сгенерированного нейросетью изображения? (Множественный выбор / Только один ответ)

43. Какой из видов машинного обучения основывается на взаимодействии обучаемой системы со средой?

- a. С учителем
- b. Без учителя
- c. Обучение с подкреплением

Какой из видов машинного обучения основывается на взаимодействии обучаемой системы со средой? (Множественный выбор / Только один ответ)

44. Какой кластер программ из перечисленных связан с Apache Hadoop?

- a. Anaconda
- b. SciPy
- c. Hortonworks Sandbox

Какой кластер программ из перечисленных связан с Apache Hadoop? (Множественный выбор / Только один ответ)

45. Какой компонент содержится в высокоуровневой платформе для запуска MapReduce задач Pig?

- a. Компилятор программ на PigLatin
- b. Клиент Pig
- c. Распределенная файловая система HDFS

Какой компонент содержится в высокоуровневой платформе для запуска MapReduce задач Pig? (Множественный выбор / Только один ответ)

46. Какой метод обучения используется в модели CatBoost?

- a. Деревья решений
- b. Метод опорных векторов
- c. Градиентный бустинг

Какой метод обучения используется в модели CatBoost? (Множественный выбор / Только один ответ)

47. Какую проблему MapReduce решает Spark?

- a. Низкая производительность
- b. Проблема управления ресурсами
- c. Нехватка памяти для хранения файлов

Какую проблему MapReduce решает Spark? (Множественный выбор / Только один ответ)

48. Компонент Apache Spark, ответственный за хранение данных и предоставление доступа к ним

- a. Driver
- b. Executor
- c. Node

Компонент Apache Spark, ответственный за хранение данных и предоставление доступа к ним (Множественный выбор / Только один ответ)

49. Кто создал первую модель искусственных нейронных сетей?

- a. Ян Лекун
- b. Мак-Каллок и Питтс
- c. Фрэнк Розенблатт

Кто создал первую модель искусственных нейронных сетей? (Множественный выбор / Только один ответ)

50. Модель анализа текста CatBoost может быть запущена из:

- a. Python
- b. Kotlin
- c. C

Модель анализа текста CatBoost может быть запущена из (Множественный выбор / Только один ответ)

51. Модель анализа текстов BERT обучается на задаче:

- a. Предсказания следующего предложения
- b. Генерации пропущенного токена
- c. Одновременно на двух задачах

Модель анализа текстов BERT обучается на задаче (Множественный выбор / Только один ответ)

52. Неверное утверждение об распределенных файловых системах:

- a. Способны хранить файлы, размер которых превышает размер диска отдельного компьютера
- b. Файлы автоматически реплицируются на нескольких серверах
- c. Система сложно масштабируется

Неверное утверждение об распределенных файловых системах (Множественный выбор / Только один ответ)

53. Область модели памяти Spark для хранения объектов, необходимых во время выполнения задач Spark:

- a. Память хранения
- b. Пользовательская память
- c. Память исполнения

Область модели памяти Spark для хранения объектов, необходимых во время выполнения задач Spark (Множественный выбор / Только один ответ)

54. Область модели памяти Spark для хранения структуры данных пользователя, которые создаются в процессе вычислений с помощью трансформаций RDD:

- a. Резервированная память
- b. Пользовательская память
- c. Память Spark

Область модели памяти Spark для хранения структуры данных пользователя, которые создаются в процессе вычислений с помощью трансформаций RDD (Множественный выбор / Только один ответ)

55. Пакет Python для применения машинного обучения к большим данным

- a. RPy2
- b. PySpark
- c. StatsModels

Пакет Python для применения машинного обучения к большим данным (Множественный выбор / Только один ответ)

56. Пакет Python, который помогает решить проблему нехватки памяти при работе с большими данными и хранит массивы в сжатой форме

- a. Dask
- b. Blaze
- c. Bcolz

Пакет Python, который помогает решить проблему нехватки памяти при работе с большими данными и хранит массивы в сжатой форме (Множественный выбор / Только один ответ)

57. Парадигма обучения нейронных сетей, при которой алгоритм обучения сети подстраивает веса по своему усмотрению?

- a. Обучение «с учителем»
- b. Обучение «без учителя»
- c. Обучение методом критики

Парадигма обучения нейронных сетей, при которой алгоритм обучения сети подстраивает веса по своему усмотрению? (Множественный выбор / Только один ответ)

58. Парадигма обучения нейронных сетей, при которой имеется возможность оценивать правильность работы сети и указывать желаемое направление обучения?

- a. Обучение «с учителем»
- b. Обучение «без учителя»
- c. Обучение методом критики

Парадигма обучения нейронных сетей, при которой имеется возможность оценивать правильность работы сети и указывать желаемое направление обучения? (Множественный выбор / Только один ответ)

59. Перечислите четыре основных характеристики Big Data:

- a. Virtualization, Volume, Variability, Vehicle
- b. Variety, Velocity, Volume, Value
- c. Verification, Volume, Velocity, Visualization

Перечислите четыре основных характеристики Big Data (Множественный выбор / Только один ответ)

60. Повысить производительность Apache Kafka можно с помощью

- a. Замены HDD-дисков на SSD
- b. Увеличения плотности разделов на каждом брокере
- c. Повышения коэффициента репликации

Повысить производительность Apache Kafka можно с помощью (Множественный выбор / Только один ответ)

61. Подходит ли MapReduce для интерактивного анализа?

- a. Не применимо к анализу вовсе
- b. Да
- c. Нет

Подходит ли MapReduce для интерактивного анализа? (Множественный выбор / Только один ответ)

62. Предпочтительнее использовать HBase вместо реляционной базы данных в случае:

- a. Большого объема данных
- b. В случае, когда система должна поддерживать транзакции
- c. Нужно объединять несколько таблиц по ключу

Предпочтительнее использовать HBase вместо реляционной базы данных в случае (Множественный выбор / Только один ответ)

63. Признак F1 может принимать значения A, B, C, D, E и F и отображает оценку ученика колледжа. Какое из следующих утверждений верно?

- a. F1 номинальная переменная
- b. F1 ординальная переменная
- c. Ни одно не верно

Признак F1 может принимать значения A, B, C, D, E и F и отображает оценку ученика колледжа. Какое из следующих утверждений верно? (Множественный выбор / Только один ответ)

64. Пример нейронной сети, применяемой в рекомендательных системах

- a. GraphSage
- b. PinSage
- c. Оба варианта

Пример нейронной сети, применяемой в рекомендательных системах (Множественный выбор / Только один ответ)

65. Пример распределенной файловой системы:

- a. Apache Thrift
- b. Red Hat GlusterFS
- c. Reddis

Пример распределенной файловой системы (Множественный выбор / Только один ответ)

66. Пример распределенной файловой системы:

- a. Onyx
- b. Apache Flume
- c. HDFS

Пример распределенной файловой системы (Множественный выбор / Только один ответ)

67. Пример структурированных данных:

- a. Файл Excel
- b. Журнал веб-серверов

- c. Электронное письмо, написанное человеком

Пример структурированных данных (Множественный выбор / Только один ответ)

68. Тип баз данных, эффективно реализующий хранение данных с высокой связностью

- a. Хранилища «ключ-значение»
- b. Хранилища документов
- c. Графовые базы данных

Тип баз данных, эффективно реализующий хранение данных с высокой связностью (Множественный выбор / Только один ответ)

69. Увеличение какого из этих гиперпараметров может привести к переобучению случайного леса?

- a. Скорость обучения
- b. Количество деревьев
- c. Глубина дерева

Увеличение какого из этих гиперпараметров может привести к переобучению случайного леса? (Множественный выбор / Только один ответ)

70. Формат Parquet считается

- a. строковым
- b. колоночным
- c. неструктурированным

Формат Parquet считается (Множественный выбор / Только один ответ)

71. Чем нейронная сеть похожа на естественные структуры мозга?

- a. Только название
- b. Полностью повторяет
- c. Представляет собой связанные нейроны

Чем нейронная сеть похожа на естественные структуры мозга? (Множественный выбор / Только один ответ)

72. Чем отличается глубокое обучение от машинного?

- a. Машинное обучение — это обучение компьютерных алгоритмов без специального свода правил, а глубокое обучение использует все возможные известные виды обучения
- b. Машинное обучение и глубокое обучение по сути одно и то же
- c. Машинное обучение изучает алгоритмы, обучающиеся без использования специального свода правил, а глубокое обучение — это машинное обучение нейронных сетей с более чем тремя скрытыми слоями нейронов

Чем отличается глубокое обучение от машинного? (Множественный выбор / Только один ответ)

73. Что происходит, если выходит из строя диск с промежуточными данными между Map и Reduce?

- a. Задача завершается с ошибкой
- b. Данные восстанавливаются из других реплик

- c. Перезапускается mapper, который создал эти данные

Что происходит, если выходит из строя диск с промежуточными данными между Map и Reduce? (Множественный выбор / Только один ответ)

74. Что такое веса в ИНС?

- a. Веса — это коэффициенты, определяющие, насколько результат, выдаваемый сетью, отличен от реального
- b. Веса — это коэффициенты, определяющие, какой вклад вносят передаваемые по связям между нейронами сигналы
- c. Веса — это показатели, характеризующие обучаемость той или иной архитектуры сети

Что такое веса в ИНС? (Множественный выбор / Только один ответ)

75. Эффективность модели машинного обучения на проверочном наборе данных значительно хуже ее эффективности на обучающей последовательности:

- a. Для моделей с высокой дисперсией
- b. Для моделей со значительной систематической ошибкой
- c. Оба варианта правильные

Эффективность модели машинного обучения на проверочном наборе данных значительно хуже ее эффективности на обучающей последовательности (Множественный выбор / Только один ответ)
