

Hadoop – это...	набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах
R не является...	языком программирования для статистической обработки данных и работы с графикой, созданный на основе языка S
Алгоритм — это:	система правил, описывающая последовательность действий, которые необходимо выполнить для решения задачи
Аналитик это ...	специалист в области анализа и моделирования
Бизнес процесс – это:	Взаимосвязанные виды деятельности, преобразующие входы в выходы
Большинство методов Data mining были разработаны в рамках ...	теории искусственного интеллекта
В 2016 году программа AlphaGo обыграла одного из мировых по шахматам чемпионов Ли Седоля. Какая компания разработала ИИ AlphaGo?	Facebook
В каких из приведенных ниже случаев речь идет структурированных данных:	Данные о продажах фирмы, представленные в виде отчетов в MS Excel
В каком из приведённых примеров наиболее эффективны NoSQL решения типа ключ-значение?	потоковая обработка логов кластера серверов и быстрого сохранения без требования оперативной аналитики
В каком случае система будет являться и распределенной и сосредоточенной одновременно	системы, решающие одинаковые задачи
В краудсорсинге вклад в проект каждого участника осуществляется...	лишь финансами
Выберите неверное высказывание...	большие данные – это данные объёма свыше 1 Tb
Выберите технологию потоковой обработки событий в режиме реального времени	Apache Kafka
Горизонтальная масштабируемость при обработке Big Data – это...	расширение механизма обработки данных при росте объема данных
Дайте определение Big Data	Комплексный набор инструментов обработки структурированных данных колоссальных объемов
Данные могут быть получены в результате...	Измерений
Деревья решений относятся к группе (группам) ...	статистических методов
Для каких систем предикаты местоположения элементов или групп элементов играют существенную роль	для сосредоточенных
Для чего предназначены криптографические протоколы	для установления режимов приема/передачи данных между удаленными объектами
До появления Data Mining невозможно было найти связи...	студентов вуза и их успеваемости по предмету
Заказчик бизнес-процесса – это должностное лицо:	Имеет в своем распоряжении средства для заказа выхода бизнес - процесса
Закончите следующее предложение: «С точки зрения машины, информация становится структурированной, если...	Машина проинструктирована, каким образом её обрабатывать
Изначально Big Data применяли лишь в...	геологии
Интеллектуальный анализ данных или Data Mining...	Информация, которая организована и проанализирована с целью сделать ее понятной и применимой для решения задачи или принятия решений.
Искусственные нейронные сети (ИНС) — модели машинного обучения, использующие комбинации распределенных простых операций, зависящих от обучаемых параметров, для обработки входных данных. Какого вида ИНС не существует?	Наивные
Итак, вы решили работать с большими данными. Какой из этих инструментов вам вряд ли пригодится?	SQL
К основным характеристикам Big Data относятся:	Virtualization, Volume, Variability, Velocity
Как можно получить помощь в R:	В R-Studio можно поставить курсор на название функции и нажать F1
Как называется в R тип данных «строки»?	Моделирование (Modeling)
Как называется в R тип данных «целые числа»?	Понимание данных (Data Understanding)
Как называется в R тип данных «числа с плавающей точкой»?	В R-Studio
Как происходит обучение нейронной сети?	эксперты настраивают нейронную сеть
Какая компания создала технологию MapReduce?	Google
Какая функция dplyr служит для вертикального объединения таблиц?	bind()
Какие из задач решаются Big Data?	Мониторинг оборудования
Какие модули по умолчанию входят в состав проекта Apache Hadoop?	HDFS
Каким образом можно объявить переменную «a» на языке R:	a =
Какими могут быть распределенные системы	непрерывными
Какое API было добавлено в Hadoop v2.0?	YARN
Какое из нижеперечисленных понятий не относится к перечню необходимых критериев для создания проекта, связанного с Большими данными?	Географическое положение

Какое множество состоит из сосредоточенных элементов	V
Какое название получила технология использования глобальной сети компьютеров для решения сложной задачи	Web
Какой язык программирования из перечисленных является наиболее важным для аналитика?	C++
Краудсорсинг большого проекта невозможен без...	соединения его с другими проектами, композиции
Локальность данных Big Data – это:	Расширение механизма обработки данных при росте объема данных
Медианой для выборки 1,_,3,7,10,15,16,18 является:	7,714286
На каком из этапов жизненного цикла данных по методологии CRISP-DM происходит проверка гипотез?	Понимание бизнеса (Business understanding)
На основе какого языка был создан R?	C
Назовите "естественные распределенные системы"	государственное и муниципальное управление
Назовите основные факторы распределенности алгоритма	сбор исходных данных из разных точек
Назовите основные характеристики непрерывной распределенной системы	максимальная удаленность элементов друг от друга
Назовите сложности иерархических методов кластеризации:	ограничение объема набора данных
Неверно, что дендрограмма – является результатом работы ...	дивизимного кластерного анализа
Неверно, что к достоинствам алгоритма кластеризации k-средних относится:	простота использования и быстрота использования
Неверно, что Классификация относится к:	контролируемому обучению
Объём накопленных человечеством цифровых данных на 2022 год измеряется:	Петабайтами
Одна из главных целей Big Data – это...	привлечение все больше пользователей
Одна из главных целей Data Mining – это получение...	связей малых выборок данных, распространение их на большие массивы
Определите виды зависимости в множестве Ud	от источников информации, имеющих определенное местоположение
Определите тип отношений распределенной системы	бинарные
Отказоустойчивость Big Data – это, когда:	Активируются до 1000 компьютеров
Отметьте верное понимание Variety в контексте характеристик Big Data...	высокая скорость генерирования данных
Отметьте причину создания NoSQL баз данных...	высокая стоимость горизонтальной масштабируемости RDBMS при сохранении требования высокой доступности
Отметьте те из вариантов, в которых данные структурированы...	данные о продажах компании, представленные в виде ежемесячных отчетов в формате MS Word
Первый этап жизненного цикла данных в соответствии с методологией CRISP-DM:	Внедрение (Deployment)
Примером применения Big Data не может быть...	зондирование Земли из космоса
Принципом Big Data не является...	горизонтальная масштабируемость
Программа от Google научилась рисовать на основе эскизов, сделанных людьми. Что при этом учитывала программа?	стиль типичный для похожих изображений
Процессы маркетинга относят к группе:	Процессы управления
Процессы подбора персонала относят к группе:	Процессы управления
Расписание движения поездов может рассматриваться как пример...	Табличной модели
Решение задачи прогнозирования ...	возможно без обучающей выборки данных
Сколько в 1 зеттабайте тарабайт?	1,073742•109
Сколько Петабайт в Зеттабайте?	1024
Технология Web mining	информации, содержащейся на Web-узлах
Технология Web mining применяет технологию Data Mining для анализа:	неструктурированной информации
Традиционные методы визуализации могут находить следующее применение...	Представлять пользователю информацию в наглядном виде
У машинного обучения есть ряд задач. Как называется та, что направлена на предсказание значения той или иной непрерывной числовой величины для входных данных?	Кластеризация
Укажите фактор, способствовавший появлению тренда больших данных...	маркетинговые кампании крупных корпораций
Установите соответствие между алгоритмами объединения двух кластеров и их характеристиками:	Метод дальнего соседа, метод средней связи и метод медианной связи
Установите соответствие между наиболее распространенными подходами обработки данных и их характеристиками:	SQL, MapReduce, SAP HANA
Установите соответствие между способами обучения нейронной сети и их характеристиками:	Обучение без учителя, обучение с учителем, обучение с подкреплением
Что из перечисленного помогает следить за эволюцией документа, над созданием которого работает одновременно большое количество авторов?	Пространственный поток
Что не является целью процесса Business Intelligence?	Интерпретация большого количества данных

Что означает термин «Big Data» в информационных технологиях?	Комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов
Что означает термин NoSQL?	Не SQL
Что такое Business Intelligence (BI):	Синоним понятия «бизнес-анализ»
Что такое SQL?	Реляционная база данных
Эффективно применять Data Mining с целью защиты от...	мошенников