

Министерство сельского хозяйства Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования «Горский государственный аграрный университет»
(ФГБОУ ВО Горский ГАУ)

Факультет Межфакультетский центр

Кафедра Информационных технологий

Учебный год 2023-2024

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

АНАЛИЗ БОЛЬШИХ ДАННЫХ

ОБРАЗОВАТЕЛЬНАЯ ПРОГРАММА ВЫСШЕГО ОБРАЗОВАНИЯ -

ПРОГРАММА СПЕЦИАЛИТЕТА

Наименование направления подготовки	38.05.01 Экономическая безопасность
Направленность (профиль)	Экономико-правовое обеспечение экономической безопасности
Реквизиты федерального государственного образовательного стандарта высшего образования	Приказ Минобрнауки России от 14 апреля 2021 г. № 293
Год начала подготовки	2022
Очная форма обучения - учебные планы по годам приема	2023
Заочная форма обучения - учебные планы по годам приема	2022,2023
Очно-заочная форма обучения - учебные планы по годам приема	
Номер по реестру ОП ВО ФГБОУ ВО Горский ГАУ	С-380501-2022
Реквизиты решения ученого совета ФГБОУ ВО Горский ГАУ об утверждении ОП ВО	Протокол от 11 апреля 2023 г. № 6
Реквизиты приказа ректора или уполномоченного лица об утверждении ОП ВО	Приказ врио ректора от 11 апреля 2023 г. № 85/06
Место дисциплины в структуре учебного плана	Обязательная часть
Количество зачетных единиц	4 ЗЕ

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

№ №	Планируемые результаты освоения образовательной программы		Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине	Направление воспитательной работы
	Наименование категории (группы) компетенций	Код и наименование компетенции			
		ОПК-6 Способен использовать современные информационные технологии и программные средства при решении профессиональных задач	ОПК-6. И-1. Понимает базовые принципы использования современных информационных технологий и программных средств	<p>Знает базовые принципы использования современных информационных технологий и программных средств.</p> <p>Умеет анализировать профессиональные задачи, выбирать и использовать подходящие ИТ-решения.</p> <p>Владеет современными программно-техническими платформами и программными средствами, в том числе отечественного производства, способен анализировать профессиональные задачи, выбирая и используя подходящие ИТ-решения.</p>	
			ОПК-6. И-2. Применяет современные информационные технологии и программные средства для решения профессиональных задач.	<p>Знает специализированные пакеты прикладных программ, предназначенные для решения профессиональных задач.</p> <p>Умеет использовать современные информационные технологии и программные средства для решения профессиональных задач.</p> <p>Владеет навыками анализа профессиональных задач и способен применять современных программно-технические платформы и программные средства, в том числе отечественного производства в профессиональной деятельности.</p>	

2. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ ДЕЯТЕЛЬНОСТИ

2.1. Трудоемкость дисциплины по видам учебной деятельности и формам обучения:

Виды учебной деятельности	Всего часов <u>144ч</u> , в том числе часов:		
	Очная форма обучения	Заочная форма обучения	Очно-заочная форма обучения
Лекционные занятия	18	4	–
Практические (лабораторные, др.) занятия	54	12	–
Самостоятельная работа (СР)	72	128	–
Форма промежуточной аттестации	экзамен		

2.2. Трудоемкость дисциплины по (разделам) темам:

№ № п/п	Наименование тем	Всего часов								
		Очная форма обучения			Заочная форма обучения			Очно-заочная форма обучения		
		Лекции	Практические (лабораторные, др.) занятия	СР	Лекции	Практические (лабораторные, др.) занятия	СР	Лекции	Практические (лабораторные, др.) занятия	СР
1.	Введение в большие данные	2	2	5	2	2	15	–	–	–
2.	Современные технологии управления BigData. Методики анализа BigData	2	4	5			15	–	–	–
3.	Системы управления базами данных.	2	6	5			15	–	–	–
4.	Основы SQL.	2	6	7			2	15	–	–
5.	Введение в R и его применение в больших данных.	2	8	10	2	2	15	–	–	–
6.	Введение в Python и его применение в больших данных.	2	8	10			15	–	–	–
7.	Аналитика больших данных и ее инструментарий. Нейронные сети.	4	10	15			15	–	–	–
8.	Системы BI как инструмент анализа и обработки больших данных	2	10	15			2	23	–	–

3. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ, СТРУКТУРИРОВАННОЕ ПО ТЕМАМ

ТЕМА 1. ВВЕДЕНИЕ В БОЛЬШИЕ ДАННЫЕ

На лекции рассматриваются цели и задачи учебной дисциплины. Вводятся основные понятия. Большие данные. Хранение больших данных. Вопросы безопасности BIG DATA. Понятие Data Mining. Программный инструментарий Data Mining и Machine Learning. Задачи анализа больших данных.

ВОПРОСЫ ПРАКТИЧЕСКОГО ЗАНЯТИЯ (ПЗ):

ПЗ: История вопроса. Принципы работы с большими данными. Ведение в машинное обучение и обработку данных. Примеры задач. Постановка основных классов задач.

Для САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПРЕДЛАГАЮТСЯ СЛЕДУЮЩИЕ ВОПРОСЫ:

- 1). Что такое информация и какими свойствами она обладает? Почему понятие «информация» многозначно?
- 2). В чем состоит сущность и каковы особенности больших данных?
- 3). В чем заключается отличие больших данных и Big Data?
- 4). Какие вы можете привести реальные примеры значения Big Data в практике деятельности предприятий и организаций?
- 5). Какие характеристики присущи BIG DATA?
- 6). Какие существуют базовые принципы обработки BIG DATA?
- 7). Что такое потоковые данные?
- 8). Что такое хранилища для ключей?
- 9). Перечислите этапы жизненного цикла данных.
- 10). Перспективы развития больших данных.

ТЕМА 2. СОВРЕМЕННЫЕ ТЕХНОЛОГИИ УПРАВЛЕНИЯ BIG DATA. МЕТОДИКИ АНАЛИЗА BIG DATA.

Обзор источников информации для Big Data. Методики анализа больших данных. Современные технологии обработки больших данных. Хранилища данных. Системы для Больших Данных: конвергенция архитектур. Big Data Система управления большими данными. Модели данных.

ВОПРОСЫ ПРАКТИЧЕСКОГО ЗАНЯТИЯ (ПЗ):

ПЗ: Современные технологии обработки больших данных: машинное обучение, правило ассоциаций, дерево классификаций.

ПЗ: Хранилище данных и озеро данных. Витрины данных

Для САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПРЕДЛАГАЮТСЯ СЛЕДУЮЩИЕ ВОПРОСЫ:

- 1). Вертикальное и горизонтальное масштабирование при работе с Big Data
- 2). Сущность технологий
- 3). Алгоритмы работы Map Reduce: преимущества и недостатки
- 4). Платформа Hadoop в качестве инструмента успешного использования Map Reduce.
- 5). Сравнительный анализ инструментов Big Data – платформ Hadoop и Spark
- 6). Какие аспекты и принципы необходимо учитывать для управления Big Data?
- 7). В чем заключаются задачи управления большими данными на основе модели М. Чисхолма?
- 8). Как можно охарактеризовать инструменты управления Big Data? В чем их значение?
- 9). Каковы предпосылки появления Data-driven организаций?
- 10). Какие действия следует предпринять компании, чтобы правильно организовать управление Big Data?
- 11). С какими проблемами сталкиваются организации при управлении большими данными?

ТЕМА 3. СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ

Реляционные и нереляционные СУБД. Реляционная модель. Первичные ключи, внешние ключи и нормализация базы данных. Запросы к базе данных: синтаксис. Основные типы данных. Сортировка, выборка, фильтрация. Работа в учебной базе данных.

ВОПРОСЫ ПРАКТИЧЕСКОГО ЗАНЯТИЯ:

ПЗ: Изучение файловой системы БД (на примере иерархической модели данных). Работа с типизированным файлом. Реляционные таблицы (принципы построения и работы). Определение домена. Создание таблиц в среде Microsoft Access.

ПЗ: Изучение использования на практике ER-модели (модели «сущность-связь»): концепция ER-модели. Изучение возможности нормализации данных.

ПЗ: Проектирование реляционной базы данных: построение ER-диаграммы, построение реляционной схемы, нормализация таблиц. Определение форм отношений и типов связей.

Для САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПРЕДЛАГАЮТСЯ СЛЕДУЮЩИЕ ВОПРОСЫ:

- 1). Дайте определение СУБД. Состав компонентов СУБД. Основные функции СУБД.
- 2). Классификация СУБД. Критерии выбора СУБД.
- 3). Что представляет собой модель данных? Классификация моделей данных.
- 4). Достоинства и недостатки иерархической модели данных.
- 5). Охарактеризуйте сетевую модель.
- 6). Охарактеризуйте реляционную модель.

7). Каким образом реализуется поддержка целостности базы данных? Что означает избыточность данных?

8). Этапы проектирования базы данных.

9). Объясните, в чем суть процесса нормализации.

ТЕМА 4. ОСНОВЫ SQL

Реляционная модель данных. СУБД PostgreSQL. Выбор данных из базы: оператор SELECT. Фильтрация данных с помощью WHERE. Сортировка данных: ORDER BY. Создание, изменение и удаление таблиц. CREATE TABLE, DROP TABLE. Типы данных. Создание, изменение и удаление данных. Операторы INSERT, UPDATE, DELETE. Группировка данных в SQL: GROUP BY. Функции агрегации SUM, AVG, COUNT, MAX, MIN. Декомпозиция данных в базе. Объединение данных из нескольких таблиц: JOIN. Типы объединений данных в SQL: внутреннее, внешнее (левое, правое, полное), перекрестное. Подзапросы в SQL

ВОПРОСЫ ПРАКТИЧЕСКОГО ЗАНЯТИЯ:

ПЗ: Элементы языка SQL. Классификация команд языка SQL по назначению. Язык запросов DQL. Язык определения данных DDL.

ПЗ: Классификация команд языка SQL по назначению. Язык манипулирования данными DML. Язык управления доступом к данным DCL. Язык обработки транзакций TPL. Язык управления курсором CCL. Организация запросов в форме SQL

ПЗ: Изучение основных понятий языка SQL и работы оператора SELECT. Изучение файловых систем (язык SQL, операторы манипулирования данными). Основные типы данных языка SQL. Работа с базой данных, таблицами, представлениями, индексами, хранимыми процедурами, триггерами.

Для САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПРЕДЛАГАЮТСЯ СЛЕДУЮЩИЕ ВОПРОСЫ:

- 1). Основные возможности языка SQL.
- 2). Общие команды SQL.
- 3). Инструкция SELECT.
- 4). Типы данных, допустимые к использованию в SQL.
- 5). SQL-запросы на объединение.
- 6). SQL-запросы к серверу.
- 7). Создание подчиненного запроса SQL.

ТЕМА 5. ВВЕДЕНИЕ В R И ЕГО ПРИМЕНЕНИЕ В БОЛЬШИХ ДАННЫХ.

Особенности Языка R. Общее представление о задачах и особенностях среды R.

Типы данных языка R. Ввод данных из разных источников. Вывод результатов расчета в файл. Сохранение и вызов программного кода. Вычисление основных

характеристик выборки. Вычисление результирующих величин для таблицы (по строкам и столбцам). Визуализация данных. Использование метода главных компонент. Трехмерная графика

ВОПРОСЫ ПРАКТИЧЕСКОГО ЗАНЯТИЯ:

ПЗ: Классы данных в R. Работа с векторами, со списками, матрицами и таблицами, фреймами данных

ПЗ: Условные конструкции языка R

ПЗ: Операторы цикла WHILE и FOR

ПЗ: Базовые графические возможности R

Для САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПРЕДЛАГАЮТСЯ СЛЕДУЮЩИЕ ВОПРОСЫ:

- 1) Объекты языка R, организация обращения к элементам данных
- 2) Операторы [, [[, \$. Извлечение элементов с учетом свойств.
- 3) Создание скриптов. Запуск скриптов.
- 4) Особенные типы данных (NA, NaN, Inf)
- 5) Создание объектов R
- 6) Создание пользовательских функций в R
- 7) Ввод данных с внешних носителей.
- 8) Графические команды высокого уровня plot(), hist(), boxplot()
- 9) Графические команды низшего уровня lines(), points(), text(), legend() и другие.
- 10) Условный оператор if() {}
- 11) Операторы цикла for() {}, while(){}

ТЕМА 6. ВВЕДЕНИЕ В PYTHON И ЕГО ПРИМЕНЕНИЕ В БОЛЬШИХ ДАННЫХ.

Особенности Python. Python-библиотека Scikit-learn как инструмент машинного моделирования.

ВОПРОСЫ ПРАКТИЧЕСКОГО ЗАНЯТИЯ:

ПЗ: Scikit-learn: модели регрессии. Регрессионный анализ как задача машинного обучения

ПЗ: Scikit-learn: модели классификации. Линейные модели классификации

ПЗ: Классификация как задача машинного обучения.

ПЗ: Scikit-learn: модели кластеризации. Кластерный анализ как задача машинного обучений

Для САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПРЕДЛАГАЮТСЯ СЛЕДУЮЩИЕ ВОПРОСЫ:

- 1) Строки и списки. Файлы и файловая система.
- 2) Пользовательские функции. Словари.
- 3) Связанные структуры данных.

4) Методы pandas

5) Библиотеки визуализации данных Matplotlib, Seaborn, Altair, Plotly Express

ТЕМА 7. АНАЛИТИКА БОЛЬШИХ ДАННЫХ И ЕЕ ИНСТРУМЕНТАРИЙ. НЕЙРОННЫЕ СЕТИ

Методы анализа и обработки больших данных, анализ больших данных в облаках, технологии аналитики в памяти (in memory), практическое применение Big Data, Методы многомерного статистического анализа

ВОПРОСЫ ПРАКТИЧЕСКОГО ЗАНЯТИЯ:

ПЗ: Методы многомерного статистического анализа

ПЗ: Дискриминантный анализ

ПЗ: Кластерный анализ

ПЗ: Дисперсионный анализ

ПЗ: Построение нейросети в MS Excel

Для САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПРЕДЛАГАЮТСЯ СЛЕДУЮЩИЕ ВОПРОСЫ:

- 1). Методы многомерного статистического анализа
- 2). Характеристика группы методов факторного анализа
- 3). Суть кластерного анализа
- 4). Алгоритм применения кластерного анализа
- 5). Методы многомерного шкалирования: основные определения, порядок применения
- 6). Особенности применения дисперсионного анализа
- 7). Анализ нечисловой информации

ТЕМА 8. СИСТЕМЫ ВІ КАК ИНСТРУМЕНТ АНАЛИЗА И ОБРАБОТКИ БОЛЬШИХ ДАННЫХ.

Технологии формирования многомерной интерактивной отчетности. Методы сбора и анализа больших данных и элементов информации бизнес-анализа; маркетинговая и клиентская аналитика больших данных с помощью различных ВІ платформ. Программные решения для OLAP-моделирования; технологии Tableau, Power BI и Qlik Sense; клиентская и маркетинговая аналитика больших данных в платформах Tableau, Power BI и Qlik Sense; интеграция технологий визуальной аналитики и машинного обучения для анализа больших данных.

ВОПРОСЫ ПРАКТИЧЕСКОГО ЗАНЯТИЯ:

ПЗ: Power Query как с ETL-инструмент. Преобразование таблиц в целом.

ПЗ: Преобразования столбцов и добавление новых.

ПЗ: Создание модели данных. Вычисления в модели на языке DAX. Прочие инструменты настройки модели.

ПЗ: Визуализируем данные. Собираем визуализации в отчет. Добавляем интерактивность в отчет. Публикация отчета.

ПЗ: Инструменты Power BI в Excel.

Для САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПРЕДЛАГАЮТСЯ СЛЕДУЮЩИЕ ВОПРОСЫ:

- 1). Загрузка и преобразование данных (ETL-инструменты)
- 2). Многомерная модель данных
- 3). Корпоративное хранилище данных (DWH, Data Warehouse)
- 4). Продвинутая аналитика (Advanced Analytics)
- 5). Предиктивная аналитика (Predictive Analytics)
- 6). Текстовая аналитика (Text Mining) и прочий инструментарий инженеров по данным (Data Scientists)
- 7). Автоматизация отчетности (Reporting)
- 8). Самостоятельное исследование данных (Self-service BI)

4. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

4.1. ОСНОВНАЯ ЛИТЕРАТУРА

- Глебов, В. И., Анализ данных в экономике. Сборник задач.: учебник / В. И.
1. Глебов, С. Я. Криволапов. — Москва : КноРус, 2022. — 578 с. — ISBN 978-5-406-09169-2. — URL: <https://book.ru/book/943011> — Текст : электронный.
Криволапов, С. Я., Статистические вычисления на платформе Jupyter Notebook с использованием Python : учебник / С. Я. Криволапов. — Москва : КноРус, 2022. — 431 с. — ISBN 978-5-406-09739-7. — URL: <https://book.ru/book/943660> — Текст: электронный.
 2. Паршинцева, Л. С., Многомерный анализ данных на Python : учебник / Л. С.
 3. Паршинцева, А. А. Паршинцев. — Москва : КноРус, 2024. — 129 с. — ISBN 978-5-406-12606-6. — URL: <https://book.ru/book/951954> — Текст : электронный.

4.2. ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

- Анализ больших данных : учебное пособие / И. Б. Тесленко, В. Е. Крылов, А. М.
4. Губернаторов [и др.]. — Москва : КноРус, 2023. — 295 с. — ISBN 978-5-406-10550-4. — URL: <https://book.ru/book/950469> — Текст : электронный.
Кондрашов, Ю. Н., Язык SQL. Сборник ситуационных задач по дисциплине «Базы данных» : учебно-практическое пособие / Ю. Н. Кондрашов. — Москва : Русайнс, 2023. — 125 с. — ISBN 978-5-466-02005-2. — URL: <https://book.ru/book/> — Текст : электронный
 - 5.

4.3. СОСТАВ ЛИЦЕНЗИОННОГО И СВОБОДНО РАСПРОСТРАНЯЕМОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ, В ТОМ ЧИСЛЕ ОТЕЧЕСТВЕННОГО ПРОИЗВОДСТВА

1. Microsoft Windows 7 Pro
2. Office 2007 Standard
3. Moodle 3.8
4. Oracle VM VirtualBox 6
5. AutoCAD 2012 AcademicEdition New SLM ML03

4.4. СОВРЕМЕННЫЕ ПРОФЕССИОНАЛЬНЫЕ БАЗЫ ДАННЫХ, ИНФОРМАЦИОННЫЕ СПРАВОЧНЫЕ СИСТЕМЫ, ЭЛЕКТРОННЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ

1. Информационно-правовой портал «Гарант» <http://www.garant.ru/>
2. Справочная правовая система КонсультантПлюс <http://www.consultant.ru/>
3. Федеральный портал «Российское образование» <https://www.edu.ru/>
4. Система автоматизации библиотек ИРБИС64; ООО «ЭйВиДи-систем» <http://support.open4u.ru>
5. Электронная библиотечная система ООО «КноРус медиа» www.book.ru
6. Электронная библиотечная система издательства «Лань» www.e.lanbook.ru

5. МАТЕРИАЛЬНО-ТЕХНИЧЕСКИЕ УСЛОВИЯ, ОБОРУДОВАНИЕ И ТЕХНИЧЕСКИЕ СРЕДСТВА ОБУЧЕНИЯ

Для проведения обучения необходимы:

- лекционная аудитория, оборудованная комплектом мебели, доской, и проекционным мультимедийным оборудованием;
- учебная аудитория с компьютерами, оснащенными операционной системой Windows и пакетом программ Microsoft Office и имеющими доступ к сети Интернет и ЭИОС ГГАУ;
- библиотека с информационными ресурсами на бумажных и электронных носителях.

6. ОЦЕНОЧНЫЕ СРЕДСТВА

6.1 Перечень вопросов к экзамену.

1. Что означает термин «Big Data» в информационных технологиях
2. Характеристики больших данных (3v, 4v)
3. Что является основной целью обработки Big Data
4. Какие главные характеристики Big Data
5. В чем отличие систем хранения больших данных (масштабируемость, отказоустойчивость)
6. Что такое Hadoop.
7. Основные компоненты Hadoop (HDFS, MapReduce)
8. Файловая система распределенного хранения больших данных HDFS
9. Модель распределенных вычислений MapReduce
10. Какие понятия содержит в себе принцип трех "V"
11. Чем характеризуются "Большие данные"
12. Что является главным результатом процесса Business Intelligence
13. Что означает термин «Business Intelligence» в информационных технологиях
14. Расшифруйте аббревиатуру OLAP.
15. Что относится к средствам предоставления Intelligence
16. Что относится к средствам интеграции в «Business Intelligence»
17. Какие цели ставит перед собой Data Science
18. Что такое жизненный цикл аналитики данных

19. Дайте определение термину «предиктивное моделирование»
20. Что такое ETL
21. Какова роль BI-аналитика в проекте
22. Какому основному принципу следует HDFS
23. Какой размер блока по умолчанию в HDFS
24. Какие функции выполняет NameNode в HDFS
25. Какой узел отвечает за репликацию данных в Hadoop
26. Какие компоненты содержит Slave узел в Hadoop
27. Какие компоненты содержит Master узел в Hadoop
28. Какие компоненты являются частями HDFS
29. Какое API было добавлено в Hadoop v2.0
30. Для чего используется автономный режим Hadoop
31. Какой режим необходим для того, чтобы на локальной машине использовать Hadoop как кластер, состоящий из одного узла
32. Что является отличительной особенностью NoSQL
33. В каком случае стоит применять NoSQL хранилища
34. Что, согласно теореме CAP, возможно обеспечить в любой реализации распределенных вычислений
35. Какое свойство означает, что транзакции не нарушают согласованность данных, то есть они переводят базу данных из одного корректного состояния в другое
36. Что относится к преимуществам нереляционных БД
37. Для чего нужна визуализация
38. Как называется один из самых популярных языков сценариев
39. Какие достоинства у Amazon S3
40. Какие традиционные виды визуализации
41. Какие отличия и основные возможности у языка R
42. В чем особенности хранения в Amazon S3
43. Что такое дедупликация данных
44. В чем основные задачи визуализации
45. Какие требования предъявляются к визуализации
46. Какие типы визуализации можно выделить
47. Чем анализ больших данных отличается от традиционного анализа
48. Какие основные типы Data Mining
49. Какие категории Web Mining можно выделить
50. В чем основная задача Web Content Mining
51. В чем основные задачи интеллектуального анализа текстов
52. К каким алгоритмам классификации относится метод ближайших соседей
53. Что является целью кластеризации
54. С помощью какого алгоритма можно найти ассоциативное правило
55. Чем отличаются ошибки первого и второго рода
56. Что является результатом решения задачи регрессии
57. Что такое α -error
58. Какая основная цель статистического анализа
59. Что такое генеральная совокупность
60. Как оценить качество обучения модели?

6.2 Тестовые задания для диагностической работы.

1 Суммарная оперативная память IBM Watson составляет порядка:

- А) 100 гигабайт
- Б) 5000 терабайт
- В) 10 зетабайт
- Г) **15 терабайт**

2 Кто ввел термин Большие данные?

- А) **Клиффорд Линч**
- Б) Алан Тьюринг
- В) Бьерн Страуструп
- Г) Дональд Кнут

3 Какие данные занимают больше мировой памяти относительно остальных?

- А) Structured Data
- Б) **Unstructured Data**
- В) Semi-Structured Data
- Г) Quasi-Structured Data

4 BigData – это ...

- А) Представление фактов, понятий или инструкций в форме, приемлемой для интерпретации, или обработки.
- Б) **Комплексный набор неструктурированных данных колоссальных объемов.**
- В) Колоссальный объем данных, собранных человечеством.
- Г) Класс в Java, предназначенный для хранения данных от 100 Гб

5 Какая компания создала технологию MapReduce?

- А) **Google**
- Б) Yahoo
- В) EMC
- Г) Oracle

6 Данные текстовых файлов с определенными паттернами для их обработки (например, XML) являются:

- А) Структурированными
- Б) **Полуструктурированными**
- В) Квазиструктурированными
- Г) Неструктурированными

7 Что означает термин «Big Data» в информационных технологиях?

- А) Комплексный набор методов для создания файлов большого объема
- Б) **Комплексный набор неструктурированных данных колоссальных объемов.**
- В) Файлы с большим количеством данных.
- Г) Представление времени, дня, месяца и года в качестве значения количества миллисекунд, прошедших с начала нашей эры.

8 Данные имеющие определенный тип, формат и структуру (например, транзакционные данные) являются:

- А) **Структурированными**
- Б) Полуструктурированными
- В) Квазиструктурированными

Г) Неструктурированными

9 Чему примерно равен объем всей существующей на земле информации (в байтах)?

А) 10^{11}

Б) 10^{21}

В) $10^{1010101}$

Г) 10^{171}

10 В каком году впервые был введен термин Большие данные?

А) 2002

Б) 2004

В) 2006

Г) 2008

11 Что является средством анализа в BI?

А) Карты показателей;

Б) Совместная работа и управление рабочими процессами;

В) Информационные панели;

Г) BI инфраструктура.

12 Основное умение исследователя данных?

А) Умение находить наиболее важные элементы в хранимой информации

Б) Уметь прогнозировать исход работы системы

В) Находить скрытые логические связи в системе собранной информации

Г) Отличать неструктурированные данные от структурированных

13 Какой язык программирования из перечисленных является наиболее важным для аналитика?

А) C++

Б) PHP

В) F#

Г) R

14 Что означает термин «Business Intelligence» в информационных технологиях?

А) Комплексный набор методов для создания бизнес планов.

Б) Методы и инструменты для перевода необработанной информации в осмысленную, удобную для восприятия форму.

В) Файлы, содержащие информацию о бизнес плане.

Г) Технологии, направленные на развитие бизнеса.

15 Языком, на котором был разработан RabbitMQ, является:

А) Java

Б) Python

В) C++

Г) Erlang

16 Что является главным результатом процесса Business Intelligence?

А) Возможность принятия решений для бизнеса

Б) Результаты интеллектуального анализа данных

В) Возможность использования искусственного интеллекта

Г) Получение структуризации данных после выполнения всех шагов процесса

17 Что из перечисленного не является средством анализа?

А) Продвинутая визуализация

Б) Reporting

В) Predictive Modelling

Г) Data Mining

18 Что относится к средствам Intelligence»?

А) Генератор нерегламентированных запросов

Б) Совместная работа и управление рабочими процессами

В) Предиктивное моделирование и Data Mining

Г) Карты показателей

19 Процессом создания и выбора модели для предсказания вероятности наступления некоторого события является:

А) OLAP

Б) Data Mining

В) Predictive Modelling

Г) Data Science

20 Что не является целью процесса Business Intelligence?

А) Интерпретация большого количества данных;

Б) Моделирование исходов различных вариантов действий;

В) Модификация существующего программного обеспечения;

Г) Отслеживание результатов решений.