ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«ЛЕНИНГРАДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**ИМЕНИ А.С. ПУШКИНА»**

УТВЕРЖДАЮ

Проректор по учебно-методической

работе

\_\_\_\_\_\_\_\_\_\_\_\_ С.Н.Большаков

РАБОЧАЯ ПРОГРАММА

дисциплины

**Б1.В.03.ДВ.01.02 РЕШЕНИЯ НА ОСНОВЕ ТЕХНОЛОГИИ BIG DATA**

Направление подготовки **09.04.03 Прикладная информатика**

Направленность (профиль) **Прикладная информатика в цифровой экономике**

(год начала подготовки – 2022)

Санкт-Петербург

2022

**1. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ:**

Процесс изучения дисциплины направлен на формирование следующих компетенций:

|  |  |  |
| --- | --- | --- |
| Индекс компетенции | Содержание компетенции  (или ее части) | Индикаторы компетенций (код и содержание) |
| УК-1 | Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий | ИУК-1.1. Знает процедуры критического анализа, методики анализа результатов исследования и разработки стратегий проведения исследований, организации процесса принятия решения.  ИУК-1.2. Умеет принимать конкретные решения для повышения эффективности процедур анализа проблем, принятия решений и разработки стратегий.  ИУК-1.3. Владеет методами установления причинно-следственных связей и определения наиболее значимых среди них; методиками постановки цели и определения способов ее достижения; методиками разработки стратегий действий при проблемных ситуациях. |
| ПК-3 | Способен проектировать информационные процессы и системы с использованием инновационных инструментальных средств | ИПК-3.1. Знает инновационные инструментальные средства проектирования ИС.  ИПК-3.2. Умеет проектировать информационные процессы и системы, адаптировать современные ИКТ.  ИПК-3.3. Владеет способностью проектировать информационные процессы и системы с использованием инновационных инструментальных средств, адаптировать современные ИКТ к задачам прикладных ИС. |
| ПК-4 | Способен принимать эффективные проектные решения в условиях неопределенности и риска | ИПК-4.1. Знает условия неопределенности и риска проектных решений.  ИПК-4.2. Умеет принимать эффективные проектные решения в условиях неопределенности.  ИПК-4.3. Владеет способностью принимать эффективные проектные решения в условиях неопределенности и риска. |
| ПК-6 | Способен управлять информационными ресурсами и ИС | ИПК-6.1. Знает структуру деятельности по управлению сложными объектами, суть понятий информационные ресурсы и информационные системы.  ИПК-6.2. Умеет формулировать управляющие воздействия на информационные ресурсы и информационные системы.  ИПК-6.3. Владеет одним из средств компьютерного мониторинга состояния информационных ресурсов и информационных систем. |
| ПК-8 | Способен использовать и развивать методы научных исследований и инструментария в области проектирования и управления информационными системами в прикладных областях | ИПК-8-1. Знает методы научных исследований и инструментария в области проектирования и управления ИС.  ИПК-8-2. Умеет выбирать и использовать методы научных исследований и инструментария в области проектирования и управления ИС.  ИПК-8.3. Владеет способами применения методов научных исследований и инструментария в области проектирования и управления ИС в прикладных областях. |

**2. Место дисциплины в структуре ОП:**

Цель дисциплины: знакомство обучающихся с основными технологиями решения задач обработки больших данных, формирование умений применять методы анализа больших данных, использовать программные решения для аналитики больших данных.

Задачи дисциплины:

* изучение базовых понятий технологии Big Data;
* ознакомление с аналитическими методами обработки данных;
* определение массивов больших данных и проведение анализа больших данных;
* изучение современных технологий создания и обслуживания больших данных.

Дисциплина относится к дисциплинам по выбору части, формируемой участниками образовательных отношений, модуль Современные технологии представления и передачи данных.

Освоение дисциплины и сформированные при этом компетенции необходимы в последующей деятельности.

**3. Объем дисциплины и виды учебной работы:**

Общая трудоемкость освоения дисциплины составляет 3 зачетные единицы, 108 академических часов *(1 зачетная единица соответствует 36 академическим часам).*

Очная форма обучения

|  |  |  |
| --- | --- | --- |
| Вид учебной работы | Трудоемкость в акад.час | |
|  |  | Практическая подготовка |
| **Контактная работа (аудиторные занятия) (всего):** | 20 | |
| в том числе: |  | |
| Лекции | 10 | - |
| Лабораторные работы / Практические занятия (в т.ч. зачет) | 10/- | 4/- |
| **Самостоятельная работа (всего)** | 88 | |
| **Вид промежуточной аттестации (экзамен):** | - | |
| контактная работа | - | |
| самостоятельная работа по подготовке к экзамену | - | |
| **Общая трудоемкость дисциплины (в час. /з.е.)** | 108/3 | |

Заочная форма обучения

|  |  |  |
| --- | --- | --- |
| Вид учебной работы | Трудоемкость в акад.час | |
|  |  | Практическая подготовка |
| **Контактная работа (аудиторные занятия) (всего):** | 16 | |
| в том числе: |  | |
| Лекции | 4 | - |
| Лабораторные работы/ Практические занятия | 12/- | 4/- |
| **Самостоятельная работа (всего)** | 88 | |
| **Вид промежуточной аттестации (зачет):** | 4 | |
| контактная работа | 0,25 | |
| самостоятельная работа по подготовке к зачету | 3,75 | |
| **Вид промежуточной аттестации (экзамен):** | - | |
| контактная работа | - | |
| самостоятельная работа по подготовке к экзамену | - | |
| **Общая трудоемкость дисциплины (в час. /з.е.)** | 108/3 | |

**4.СОДЕРЖАНИЕ ДИСЦИПЛИНЫ:**

При проведении учебных занятий обеспечивается развитие у обучающихся навыков командной работы, межличностной коммуникации, принятия решений, лидерских качеств (включая при необходимости проведение интерактивных лекций, групповых дискуссий, ролевых игр, тренингов, анализ ситуаций и имитационных моделей, составленных на основе результатов научных исследований, проводимых организацией, в том числе с учетом региональных особенностей профессиональной деятельности выпускников и потребностей работодателей**).**

**4.1 Блоки (разделы) дисциплины.**

|  |  |
| --- | --- |
| № п/п | Наименование блока (раздела) дисциплины |
| 1 | Введение в науку и технологию больших данных |
| 2 | Структуры и методы представления больших данных |
| 3 | Первичная обработка данных |
| 4 | Инструменты обработки данных |
| 5 | Визуализация данных |
| 6 | Анализ и преобразование данных |
| 7 | Работа с временными рядами |
| 8 | Системы управления базами данных |
| 9 | Проектирование структурированных данных |
| 10 | Запросы на языке SQL |
| 11 | NoSQL хранилища |
| 12 | СУБД Redis |
| 13 | СУБД MongoDB |
| 14 | СУБД Cassandra |
| 15 | СУБД Neo4j |

**4.2. Примерная тематика курсовых работ (проектов):**

Курсовая работа по дисциплине не предусмотрена учебным планом.

**4.3. Перечень занятий, проводимых в активной и интерактивной формах, обеспечивающих развитие у обучающихся навыков командной работы, межличностной коммуникации, принятия решений, лидерских качеств. Практическая подготовка\*.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№ п/п** | **Наименование блока (раздела) дисциплины** | **Занятия, проводимые в активной и интерактивной формах** | | **Практическая подготовка\*** |
| **Форма проведения занятия** | **Наименование видов занятий** |
| 1 | Введение в науку и технологию больших данных | лекционное занятие | лекция-дискуссия |  |
| 2 | Структуры и методы представления больших данных | лекционное занятие | лекция-дискуссия |  |
| 3 | Первичная обработка данных | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах |  |
| 4 | Инструменты обработки данных | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах |  |
| 5 | Визуализация данных | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах |  |
| 6 | Анализ и преобразование данных | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах |  |
| 7 | Работа с временными рядами | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах |  |
| 8 | Системы управления базами данных | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах |  |
| 9 | Проектирование структурированных данных | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах |  |
| 10 | Запросы на языке SQL | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах |  |
| 11 | NoSQL хранилища | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах |  |
| 12 | СУБД Redis | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах | Инсталляция и настройка СУБД Redis |
| 13 | СУБД MongoDB | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах | Инсталляция и настройка СУБД MongoDB |
| 14 | СУБД Cassandra | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах | Инсталляция и настройка СУБД Cassandra |
| 15 | СУБД Neo4j | лекционное занятие | лекция-дискуссия |  |
| лабораторное занятие | работа в группах | Инсталляция и настройка СУБД Neo4j |

**\***Практическая подготовка при реализации учебных предметов, курсов, дисциплин (модулей) организуется путем проведения практических занятий, практикумов, лабораторных работ и иных аналогичных видов учебной деятельности, **предусматривающих участие обучающихся в выполнении отдельных элементов работ, связанных с будущей профессиональной деятельностью.**

**5. Учебно-методическое обеспечение для самостоятельной работы обучающихся по дисциплине:**

**5.1. Темы конспектов:**

Тема 1. Введение в науку и технологию больших данных.

Тема 2. Структуры и методы представления больших данных.

Тема 3. Первичная обработка данных.

Тема 4. Инструменты обработки данных.

Тема 5. Визуализация данных.

Тема 6. Анализ и преобразование данных.

Тема 7. Работа с временными рядами.

Тема 8. Системы управления базами данных.

Тема 9. Проектирование структурированных данных.

Тема 10. Запросы на языке SQL.

Тема 11. NoSQL хранилища.

Тема 12. СУБД Redis.

Тема 13. СУБД MongoDB.

Тема 14. СУБД Cassandra.

Тема 15. СУБД Neo4j.

**5.2. Темы для творческой самостоятельной работы обучающегося:**

Темы для творческой самостоятельной работы студента формулируются обучающимся самостоятельно, исходя из перечня тем занятий текущего семестра.

**5.3. Темы рефератов:**

* 1. Подходы к структурированию и типы данных в DataScience.
  2. Машинные данные и данные на естественном языке.
  3. Этапы в технологии представления больших данных.
  4. Экосистемы больших данных
  5. Распределенные файловые системы
  6. Научные проблемы в области больших данных
  7. Применение технологии больших данных в прогнозировании и предвидении в социально-политических процессах
  8. Понятие машинного обучения, его цели и сферы применения.
  9. Типы машинного обучения
  10. Базы данных в DataScience
  11. Многомерные статистические методы в экономике, управлении и финансах
  12. Реализация статистических методов в пакетах прикладных программ
  13. Математические основы многомерных статистических методов
  14. Методы множественного корреляционно-регрессионного анализа.
  15. Кластерный анализ
  16. кластерного анализа
  17. Постановка задач снижения размерности
  18. Факторный анализ
  19. Введение в теорию качественных признаков и нечисловой информации
  20. Интеграция данных из различных источников.

**5.4. Вопросы для подготовки к практическим занятиям (лабораторным работам):**

*Лабораторная работа №1. Первичная обработка данных*

При подходе к остановке у каждого трамвая фиксируется время. Но иногда датчик дает сбой, и время прохождения остановки не фиксируется. Необходимо восстановить пропущенное время прохождения остановок, определив пропорции прохождения каждой неопределенной остановки маршрута, усреднив время прохождения этой же остановки другими трамваями этого маршрута без пропусков.

Изображение выглядит как стол

Автоматически созданное описание

*Лабораторная работа №2. Инструменты обработки данных*

С сайта Московской фьючерсной биржи необходимо загрузить данные по контракту SPFB.RTS-12.18 (RIZ8) с 1 сентября по 31 декабря 2018 года с минутным интервалом. В файле должны присутствовать следующие поля:

<DATE> Дата проведения торгов

<TIME> Время, соответствующее началу минутного интервала

<OPEN> Цена открытия, т.е. цена первой сделки

<HIGH> Максимальная цена сделки за выбранный интервал

<LOW> Минимальная цена сделки за выбранный интервал

<CLOSE> Цена закрытия, т.е. цена последней сделки

<VOL> Объем торгов за выбранный интервал

Для каждого минутного интервала сосчитайте среднюю цену (PRICE), усреднив цену открытия, закрытия, минимума и максима.

Для каждого минутного интервала вычислите оборот по совершенным сделкам, умножив среднюю цену (PRICE) на объем (VOL).

*Лабораторная работа №3. Визуализация данных*

С сайта Московской фьючерсной биржи необходимо загрузить данные по контракту SPFB.RTS-12.18 (RIZ8) с 1 сентября по 31 декабря 2018 года с минутным интервалом. В файле должны присутствовать следующие поля:

<DATE> Дата проведения торгов

<TIME> Время, соответствующее началу минутного интервала

<OPEN> Цена открытия, т.е. цена первой сделки

<HIGH> Максимальная цена сделки за выбранный интервал

<LOW> Минимальная цена сделки за выбранный интервал

<CLOSE> Цена закрытия, т.е. цена последней сделки

<VOL> Объем торгов за выбранный интервал

Преобразуйте эти данные к часовым интервалам и постройте график «японские свечи» для даты 4 сентября.

*Лабораторная работа №4. Анализ и преобразование данных*

Поступив в университет, Василий решил не дожидаться, пока ему предоставят место в общежитии, а снять себе квартиру. В таблице представлены возможные варианты квартир, которые рассматривал Василий. В первом столбце ID квартиры, во втором - расстояние от метро пешком (в минутах). В третьем столбце количество остановок на метро, которые надо проехать, чтобы добраться до места учебы. И последний столбец - стоимость аренды в месяц (в тысячах рублей).

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **DISTANCE** | **STOP\_COUNT** | **COST** |
| 1 | 7 | 3 | 23.4 |
| 2 | 22 | 12 | 27.0 |
| 3 | 21 | 11 | 28.1 |
| 4 | 10 | 2 | 23.7 |
| 5 | 15 | 4 | 24.1 |
| 6 | 14 | 9 | 19.5 |
| 7 | 18 | 9 | 29.4 |
| 8 | 21 | 12 | 27.5 |
| 9 | 6 | 6 | 29.0 |
| 10 | 5 | 10 | 27.7 |
| 11 | 22 | 9 | 23.8 |
| 12 | 14 | 4 | 16.6 |
| 13 | 21 | 8 | 23.7 |
| 14 | 5 | 11 | 23.0 |
| 15 | 14 | 11 | 20.7 |
| 16 | 20 | 7 | 23.6 |
| 17 | 9 | 8 | 24.6 |
| 18 | 24 | 1 | 15.6 |
| 19 | 23 | 3 | 18.2 |
| 20 | 5 | 10 | 28.6 |

*Лабораторная работа №5. Работа с временными рядами*

Задан временной ряд Y из 20 элементов. Ряд содержит шумы. Проведите экспоненциальное сглаживание временного ряда и укажите сглаженное значение для последней (20-й) точки ряда. При сглаживании используйте коэффициент сглаживания α = 0.2. При вводе ответа значение округлите до целого.

*Y* = [33, 49, 38, 44, 48, 52, 37, 55, 39, 59, 41, 42, 54, 63, 53, 50, 47, 57, 58, 56]

*Лабораторная работа №6.*  Системы управления базами данных

Изображение выглядит как стол

Автоматически созданное описание

*Лабораторная работа №7. Проектирование структурированных данных*

База данных содержит таблицу, созданную командой:

CREATE TABLE T1(

ATTR\_1 INTEGER PRIMARY KEY,

ATTR\_2 INTEGER CHECK(ATTR\_2 > 11),

ATTR\_3 CHAR(1) UNIQUE,

ATTR\_4 VARCHAR(15)

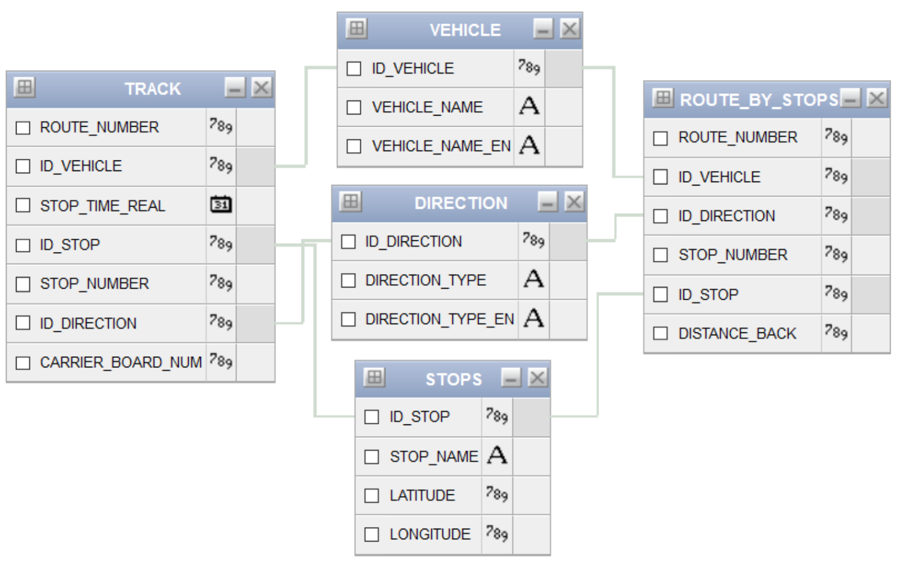
);

Укажите те команды, которые позволят добавить записи в таблицу

* INSERT INTO T1 VALUES(55, 15, 23, 'AFFFF')
* INSERT INTO T1 VALUES(21, 'B', 100, 55)
* INSERT INTO T1(ATTR\_1, ATTR\_3) VALUES(33, 'C')
* INSERT INTO T1(ATTR\_1, ATTR\_2) VALUES(155, 233)
* INSERT INTO T1 VALUES(23, 33, 17, '24.03.2020 12:00:00')
* INSERT INTO T1 VALUES(155, 'X', 13 , 'Tom')
* INSERT INTO T1(ATTR\_1, ATTR\_2) VALUES(155, 235)

*Лабораторная работа №8. Запросы на языке SQL*

Напишите запросы к представленной базе данных:



* Позволяющий определить сколько остановок в прямом направлении есть на маршруте №3 (вид транспорта — автобус).
* Позволяющий определить максимальное расстояние между соседними остановками на маршруте автобуса номер №1.

*Лабораторная работа №9. NoSQL хранилища*

Как называется формат, в котором задан следующий документ:

{

"CartoonName": "The Lion King",

"Year": 1994,

"Country": "USA",

"Duration": 88,

"FilmDirector": [ "Roger Allers", "Rob Minkoff"]

}

Какое из приведенных ниже определений подходит для определения понятия вертикального масштабирования:

* использование декларативных языков управления данными
* увеличение количества объединенных в кластер серверов при нехватке CPU, памяти или дискового пространства
* увеличение количества доступных ресурсов за счет увеличения мощности применяемых серверов
* уменьшение количества объединенных в кластер серверов при нехватке CPU, памяти или дискового пространства
* уменьшение количества доступных ресурсов за счет увеличения мощности применяемых серверов
* использование низкоуровневых языков для управления данными
* представление данных в виде таблиц
* представление данных в виде графов
* представление данных в виде агрегатов
* увеличение количества таблиц, используемых для хранения данных
* увеличение количества операторов в языках запросов

Для решения каких задач используется технология MapReduce:

* копирование данных при их обновлении на другие сервера
* увеличение ресурсов используемого сервера
* эффективное распараллеливания задач по кластеру
* повышения точности вычислений
* обеспечение отказоустойчивости
* агрегирование промежуточных результатов вычислений

*Лабораторная работа №10.* СУБД Redis

Укажите структуры данных, существующие в REDIS:

* Таблицы
* Списки
* Графы
* Строки
* Множества
* Деревья
* Упорядоченные множества
* Хеши

Укажите команды, предназначенные для работы с хеш-структурами:

* SET
* KEYS
* HMSET
* HGET
* GET
* HVALS
* HDEL
* HKEYS

Ключу «user\_list» соответствует список посетителей сайта. Каждый новый посетитель добавлялся командой RPUSH. Напишите запрос, который выведет и удалит из списка последнего добавленного посетителя.

Ключам «friends:Elena» и «friends:Boris» соответствуют множества, содержащие сведения сведения о друзьях Елены и Бориса. Напишите запрос, который сформирует (и выведет) множество всех общих друзей Елены и Бориса (то есть только тех, кто одновременно является друзьями Елены и Бориса).

*Лабораторная работа №11. СУБД MongoDB*

Дана таблица исходных данных:

Изображение выглядит как стол

Автоматически созданное описание

Как можно задать JSON документ с ключами и значениями, указанными в таблице?

Какая команда MongoDB используется для переключения баз?

Какая команда позволяет просмотреть список доступных баз?

*Лабораторная работа №12. СУБД Cassandra*

Как называется язык запросов в СУБД Cassandra?

Какой фрагмент в предложении CREATE TABLE явно указывает на возможную сортировку по указанному полю?

Какой оператор в CQL используется для выборки данных из базы?

Какой оператор в CQL используется для добавления данных в базу?

*Лабораторная работа №13. СУБД Neo4j*

Выберите правильные утверждения:

* Отношения в графовых базах связывают таблицы.
* Отношения в графовых базах связывают множества.
* Отношения в графовых базах связвают узлы.
* У отношений в графовых базах могут существовать атрибуты.
* У отношений в графовых базах не может быть атрибутов.
* Добавление новых отношений к графовой базе требует перестройки старых отношений.
* Добавление новых отношений к графовой базе не требует перестройки старых отношений.
* Язык CQL используется для создания графовых баз.
* Язык Cypher используется для создания графовых баз.

Выберите правильные утверждения:

* Язык SQL используется для обхода графовых баз.
* Язык CQL используется для обхода графовых баз.
* Язык Cypher используется для обхода графовых баз.
* Сортировка выбранных узлов и отношений не предусмотрена в Neo4j.
* Сортировка выбранных узлов и отношений возможна в Neo4j.
* В графовых базах возможен обход графа на произвольную глубину.
* В графовых базах нельзя обойти граф на произвольную глубину.

**6 Оценочные средства для текущего контроля успеваемости:**

**6.1. Текущий контроль**

|  |  |  |
| --- | --- | --- |
| №п/п | № и наименование блока (раздела) дисциплины | Форма текущего контроля |
| 1 | Разделы 1-15 | Проверка наличия конспектов.  Устный опрос, тест |
| 2 | Разделы 3-15 | Защита отчета по результатам выполнения лабораторных работ. |

**6.2. Примеры оценочных средств для текущего контроля успеваемости**

***Темы конспектов.***

Представлены в разделе 5.1.

***Перечень практических заданий.***

Представлены в п. 5.4.

***Примеры тестовых заданий.***

1. Что больше: выборка или генеральная совокупность?

*Укажите в правильном порядке основные этапы анализа данных.*

* построение моделей
* описание изучаемых объектов
* интерпретация
* очистка данных
* выявление переменных и формулировка гипотез
* сбор данных для формирования генеральной совокупности или выборки
* трансформация данных

2. Какой шкале соответствуют значения переменной: Пекин, Москва, Ханой, Париж, Лондон?

* Номинальная
* Порядковая
* Интервальная
* Относительная

3. Представлена статистика получения водительских прав в городе N-ск за 2017 год: Мужчины 15473; Женщины 9466; Студенты 3118. Какие данные можно агрегировать в представленной таблице?

* Мужчины
* Женщины
* Студенты

4. Как следует преобразовать данные в таблице для последующего анализа?

Изображение выглядит как стол

Автоматически созданное описание

*Выберите все возможные варианты.*

* Удалить повторяющиеся значения
* Преобразовать все расстояния к километрам
* Проверить регулярные выражения
* Преобразовать все расстояния к милям
* Заполнить пропуски
* Преобразовать все расстояния к метрам
* Преобразовать все расстояния к сантиметрам

5. Какие программные средства относятся к электронным таблицам:

* Notepad
* Microsoft Excel
* MySQL
* Tableau
* Oracle
* Google Sheets
* Microsoft PowerPoint
* Microsoft Access
* Word Online

6. Выберите верные высказывания про настройки доступа Google-таблиц:

* Настройки доступа позволяют просматривать и редактировать файл монопольно
* Создавая Google-таблицу, доступ к ней на просмотр автоматически предоставляется любому пользователю
* Настройки доступа позволяют предоставить доступ на просмотр файла любому, у кого есть ссылка на таблицу
* Настройки доступа позволяют предоставить доступ на редактирование файла любому, у кого есть ссылка
* Настройки доступа позволяют предоставить доступ на редактирование файла только тем, у кого есть почта на Gmail
* Настройки доступа позволяют пригласить к совместному просмотру или редактированию таблицы конкретных пользователей, указав их имена или электронную почту

7. С помощью регулярных выражений можно:

* Писать степенные функции
* Упорядочить данные таблицы
* Искать текст по заданному образцу
* Отображать регулярные изменения в таблицах

8. Фильтрация данных позволяет:

* Удалить часть данных из таблицы
* Отобразить часть данных из таблицы
* Упорядочить данные таблицы
* Разбить данные по столбцам
* Проверить корректность значений в ячейках

9. Инструмент обработки данных электронных таблиц, служащий для обобщения и агрегирования данных, называется:

* Аналитическая таблица
* Группирующий фильтр
* Сводная таблица
* Итоговый инструмент

10. К примерам визуализации данных можно отнести:

* Графики
* Линейные диаграммы
* Гистограммы
* Карты
* Диаграммы Ганта
* Диаграмма структуры организации
* Преобразование Фурье

11. Линейный график подходит:

* Для небольших наборов дискретных данных
* Для отображения пропорций в распределении данных
* Для отображения количественного значения в течение непрерывного интервала
* Тип визуализации, в которой цвет выступает в качестве дополнительного измерения, называют:
* Цветная раскраска
* Тепловая карта
* Глубинная схема
* Диаграмма Ганта

12. Нормированная диаграмма нужна для того, чтобы:

* Отображать значений двух переменных в виде точек на плоскости
* Оценить долевое участие каждого из параметров в совокупном результате
* Произвести очистку данных

13. Какой вид диаграмм лучше подойдет визуализации данных, которые отличаются на порядок:

* Линейный график
* Столбчатая диаграмма
* Комибнированная диаграмма
* Пузырьковая диаграмма

14. Выберите верные высказывания:

* Мода всегда больше медианы
* Мода всегда равна медиане
* Средневзвешенное значение всегда равно моде
* Средневзвешенное значение всегда равно медиане
* Все предыдущие утверждения неверны

15. Какие значения являются экстремальными выбросами в следующем наборе данных:



16. Выберите преобразования, которые целесообразно применить для визуализации данных с большим разбросом положительных значений:

* Логарифмирование
* Умножение на коэффициент
* Возведение в квадрат
* Извлечение квадратного корня
* Сдвиг по оси Ox

17. Укажите основные назначения нормировки

* Обеспечение возможности для сравнения, агрегации и визуализации значений нескольких переменных из различных шкал
* Заполнить пропущенные значения
* Очистить данные от значений, не являющихся нормальными

18. Что такое оптимизация?

* Это поиск наиболее значимых параметров
* Это поиск тех возможных значений параметров, при которых функция принимает максимум или минимум
* Это нормировка параметров

**7. ПЕРЕЧЕНЬ УЧЕБНОЙ ЛИТЕРАТУРЫ:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № п/п | Наименование | Авторы | Место издания | Год издания | Наличие | |
| печатные издания | ЭБС (адрес в сети Интернет) |
| 1. | Наука о данных: базовый курс: учебное пособие | Д. Келлехер, Б. Тирни | М.: Альпина Паблишер | 2020 |  | <http://biblioclub.ru> |
| 2. | Базы данных в высокопроизводительных информационных системах: учебное пособие | Николаев Е.И. | Ставрополь: СКФУ | 2016 |  | <http://biblioclub.ru> |
| 3. | Администрирование баз данных и компьютерных сетей : учебное пособие | Беспалов Д.А. | Ростов-на-Дону: ЮФУ | 2020 |  | <http://biblioclub.ru> |
| 4. | Модели баз данных: учебное пособие | О.Е. Аврунев, В.М. Стасышин | Новосибирск: НГТУ | 2018 |  | <http://biblioclub.ru> |
| 5. | Анализ данных: учебник для вузов | Мхитарян В.С. | М.: Издательство Юрайт | 2022 |  | <https://urait.ru/bcode/489100> |
|  | Введение в анализ данных: учебник и практикум | Миркин Б. Г. | М.: Издательство Юрайт | 2022 |  | <https://urait.ru/bcode/469306> |

**8. Ресурсы информационно-телекоммуникационной сети «Интернет»:**

1. «НЭБ». Национальная электронная библиотека. – Режим доступа: [http://нэб.рф/](http://www.biblioclub.ru/)

2. «eLibrary». Научная электронная библиотека. – Режим доступа: [https://elibrary.ru](https://elibrary.ru/)

3. «КиберЛенинка». Научная электронная библиотека. – Режим доступа: <https://cyberleninka.ru/>

4. ЭБС «Университетская библиотека онлайн». – Режим доступа: [http://www.biblioclub.ru/](http://www.knigafund.ru/)

5. Российская государственная библиотека. – Режим доступа: <http://www.rsl.ru/>

**9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ:**

В ходе осуществления образовательного процесса используются следующие информационные технологии:

- средства визуального отображения и представления информации (LibreOffice) для создания визуальных презентаций как преподавателем (при проведении занятий) так и обучаемым при подготовке докладов для семинарского занятия.

- средства телекоммуникационного общения (электронная почта и т.п.) преподавателя и обучаемого.

- использование обучаемым возможностей информационно-телекоммуникационной сети «Интернет» при осуществлении самостоятельной работы.

**9.1. Требования к программному обеспечению учебного процесса**

Для успешного освоения дисциплины, обучающийся использует следующие программные средства:

* Windows 10 x64
* MicrosoftOffice 2016
* LibreOffice
* Firefox
* СУБД Redis (свободно распространяемое ПО)
* СУБД MongoDB (свободно распространяемое ПО)
* СУБД Cassandra (свободно распространяемое ПО)
* СУБД Neo4j (свободно распространяемое ПО)

**9.2. Информационно-справочные системы (при необходимости):**

Не используются

**10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

Для проведения занятий лекционного типа предлагаются наборы демонстрационного оборудования и учебно-наглядных пособий.

Для изучения дисциплины используется следующее оборудование: аудитория, укомплектованная мебелью для обучающихся и преподавателя, доской, ПК с выходом в интернет, мультимедийным проектором и экраном.

Для самостоятельной работы обучающихся используется аудитория, укомплектованная специализированной мебелью и техническими средствами (ПК с выходом в интернет и обеспечением доступа в электронно-информационно-образовательную среду организации).