Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Оренбургский государственный университет»

Кафедра информатики

В.В. Запорожко

**Системы и технологии обработки больших данных**

Методические указания

Рекомендовано к изданию редакционно-издательским советом федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет» для обучающихся по образовательной программе высшего образования по направлению подготовки 09.04.02 Информационные системы и технологии

Оренбург

2020

УДК 004

ББК 32.97- 018я7

З33

Рецензент – доцент, кандидат технических наук В. В. Извозчикова

**Запорожко В. В.**

З33 Системы и технологии обработки больших данных: методические указания / В.В. Запорожко; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2020. – 26 с.

Методические указания для обучающихся по освоению дисциплины «Системы и технологии обработки больших данных» содержат список рекомендуемой литературы по разделам, контрольные вопросы для самоконтроля, задания для выполнения лабораторных работ, вопросы к экзамену.

Методические указания предназначены для обучающихся по образовательной программе высшего образования по направлению подготовки 09.04.02 Информационные системы и технологии.

УДК 004

ББК 32.97- 018я7

 Запорожко В. В., 2020

 ОГУ, 2020

2

**Содержание**

[Введение 4](#_Toc24443778)

[1 Методические указания по освоению дисциплины 6](#_Toc24443779)

[2 Содержание разделов дисциплины, рекомендуемая литература и вопросы для самоконтроля 8](#_Toc24443780)

[3 Методические указания по лекционным занятиям 15](#_Toc24443781)

[4 Лабораторные работы 17](#_Toc24443782)

[4.1 Методические указания по лабораторным занятиям 17](#_Toc24443783)

[4.2 Методические рекомендации по самостоятельной подготовке к лабораторным занятиям 17](#_Toc24443784)

[4.3 Задания к лабораторным работам 18](#_Toc24443785)

[5 Методические указания по самостоятельной работе 19](#_Toc24443786)

[6 Перечень дискуссионных тем для проведения круглого стола 21](#_Toc24443787)

[7 Методические указания по подготовке к рубежному контролю и промежуточной аттестации 22](#_Toc24443788)

[8 Вопросы к экзамену 24](#_Toc24443789)

[Список использованных источников 26](#_Toc24443790)

# 

# Введение

**Цель** освоения дисциплины «Системы и технологии обработки больших данных» состоит в формировании профессиональных компетенций, необходимых для получения фундаментальных знаний в области систем и технологий обработки больших данных, овладения умениями и способами деятельности в процессе освоения программных средств, используемых при обработке огромных объемов данных в сфере научных исследований.

**Задачи:**

1) привить устойчивый интерес к будущей научной деятельности; раскрыть роль и значение информационных технологий используемыми для обработки огромных массивов научных данных;

2) обеспечить сознательное и прочное овладение обучающимися теоретических основ современных информационных технологий получения, хранения, обработки, анализа и визуализации больших объемов данных, систематизировать знания в данной области;

3) создать условия для формирования умений и овладения обучающимися способами деятельности сознательного и рационального использования систем и технологий обработки больших данных при решении задач в научно-исследовательской деятельности.

Выполнение предложенного комплекса заданий по дисциплине «Системы и технологии обработки больших данных» способствует формированию у обучающихся направления подготовки 09.04.02 Информационные системы и технологии профессиональной компетенции *ПК\*-3 Способен выбирать методы и разрабатывать технологии получения, хранения, обработки, анализа и визуализации больших объемов научных данных* [1].

Индикаторы достижения компетенции:

ПК\*-3-В-1 Обоснованно выбирает методы и разрабатывает технологии получения, хранения, обработки, анализа и визуализации больших объемов научных данных.

ПК\*-3-В-2 Решает практические задачи в сфере научных исследований с помощью технологии обработки больших данных.

ПК\*-3-В-3 Способен создавать высокопроизводительные кластеры, применять математические модели параллельных алгоритмов и программ для обработки, анализа и визуализации больших объемов научных данных.

ПК\*-3-В-4 Способен обрабатывать большие объемы научных данных на основе современных технологий высокопроизводительных вычислений.

ПК\*-3-В-5 Владеет методами построения эволюционных моделей.

ПК\*-3-В-6 Выбирает и использует системы управления базами данных для хранения, обработки, анализа и визуализации больших объемов научных данных.

ПК\*-3-В-7 Демонстрирует знания и понимание методов и технологий получения, хранения, обработки, анализа и визуализации больших объемов научных данных.

ПК\*-3-В-8 Использует современные графические пакеты при разработке технологии получения, хранения, обработки, анализа и визуализации больших объемов научных данных.

**Планируемые результаты обучения:**

**Знать:** теоретические основы современных информационных технологий получения, хранения, обработки, анализа и визуализации больших объемов данных.

**Уметь**: выбирать методы и средства получения, хранения, обработки, анализа и визуализации больших объемов данных при решении задач профессиональной деятельности; работать с прикладным программным обеспечением, обеспечивающим технологии получения, хранения, обработки, анализа и визуализации больших объемов данных.

**Владеть:** навыками решения задач, связанных с технологией получения, хранения, обработки, анализа и визуализации больших объемов научных данных.

* 1. Методические указания по освоению дисциплины

В соответствии с учебным планом по направлению подготовки 09.04.02 Информационные системы и технологии освоение учебной дисциплины «Системы и технологии обработки больших данных» предусматривает прослушивание лекций и выполнение ряда семестровых работ. Распределение занятий по часам представлено в утвержденной рабочей программе, доступной на официальном сайте ОГУ (http://www.osu.ru): Главное меню > Университет > Направления подготовки и специальности.

Для изучения теоретической части курса необходимо освоить вопросы, рассматриваемые в лекциях и в ходе учебных дискуссий, проводимых в форме круглого стола. Список рекомендуемой литературы по разделам дисциплины приведен в п. 2. Методические указания по лекционным занятиям приведены в п. 3.

Текущий контроль знаний, проводимый в форме тестирования, предусмотрен по каждому из разделов курса. Для подготовки к его успешному прохождению рекомендуется предварительно ответить на контрольные вопросы, приведенные в п. 2.

По некоторым наиболее сложным разделам курса предусмотрено выполнение лабораторных работ. На лабораторных занятиях обучающимся выдаются задания (см. п. 4.3), обсуждаются вопросы по их выполнению и сроки сдачи, проводится защита работ. Методические указания по лабораторным занятиям и рекомендации по самостоятельной подготовке к лабораторным занятиям приводятся в п. 4.1 и 4.2 соответственно.

Для получения допуска к промежуточной аттестации по дисциплине необходимо выполнить и успешно сдать предусмотренные настоящей рабочей программой лабораторные работы.

Рекомендуется защитить все выполненные в полном объёме семестровые работы в установленные преподавателем сроки: от этого зависит равномерность распределения учебной нагрузки в семестре и успешность процесса обучения в целом.

Внеаудиторная самостоятельная работа определяется учебным планом и является обязательной для каждого обучающегося. В рамках общего объема часов, отведенных для изучения дисциплины, предусматривается выполнение следующих видов самостоятельных работ: проработку и повторение лекционного материала, материала учебников и учебных пособий, подготовку к лабораторным занятиям, рубежному контролю. В процессе выполнения самостоятельной работы, при необходимости, обращайтесь за консультациями к преподавателю, чтобы вовремя скорректировать свою учебную деятельность или обсудить правильность дальнейшего выполнения задания.

Самостоятельная внеаудиторная работа обучающихся может проходить в компьютерном классе (ауд. 1541а) или читальных залах библиотеки университета, оснащенных компьютерной и телекоммуникационной техникой и подключённых к сети Интернет. При работе в компьютерных классах соблюдайте правила техники безопасности и охраны труда.

Электронная информационно-образовательная среда университета обеспечивает доступ к электронным информационным и образовательным ресурсам: электронному каталогу, электронно-библиотечным системам, информационно-справочным системам, системам научного поиска и т.п. (http://www.osu.ru/doc/2763). Рекомендуется ознакомиться со списком рекомендуемой литературы, приведенным в рабочей программе.

В соответствии с графиком учебного процесса два раза в семестр проводится рубежный контроль успеваемости обучающихся: на восьмой и четырнадцатой неделях.

Обучение по учебной дисциплине завершается промежуточной аттестацией в форме экзамена, вопросы к которому приведены в п. 8.

Методические указания по подготовке к рубежному контролю и промежуточной аттестации приведены в п. 7.

Обучающиеся, не прошедшие промежуточную аттестацию по графику сессии, должны ликвидировать академическую задолженность в установленном порядке.

* 1. Содержание разделов дисциплины, рекомендуемая литература и вопросы для самоконтроля

**Раздел № 1 Большие данные. Основы систем больших данных**

Понятие «большие данные». Характеристики больших данных. Принципы работы с большими данными. Большие данные в научных исследованиях. Экосистема больших данных. Распределенные файловые системы. Распределенные фреймворки. Бенчмаркинг. Серверное программирование. Планирование. Системы развертывания. Интеграция данных. Информационная безопасность. Машинное обучение. Базы данных NoSQL и новые SQL базы данных.

**Рекомендуемая литература по разделу:**

1. Базы данных в высокопроизводительных информационных системах / авт.-сост. Е.И. Николаев; Министерство образования и науки РФ, Федеральное государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». - Ставрополь: СКФУ, 2016. - 163 с. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=466799>
2. Дадян, Э.Г. Методы, модели, средства хранения и обработки данных: учебник / Э.Г. Дадян, Ю.А. Зеленков. - М.: Вузовский учебник: ИНФРА-М, 2017. - 168 с. - Режим доступа: <http://znanium.com/catalog/product/543943>
3. Майер-Шенбергер, В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / В. Майер-Шенбергер, К. Кукьер. – М.: Манн, Иванов и Фербер, 2014. - 240 с.
4. Методы и модели исследования сложных систем и обработки больших данных: монография / И.Ю. Парамонов, В.А. Смагин, Н.Е. Косых, А.Д. Хомоненко; под редакцией В.А. Смагина и А.Д. Хомоненко. - Санкт-Петербург: Лань, 2020. - 236 с.
5. Просто о больших данных: пер. с англ. / Д. Гурвиц [и др.]. - М.: Эксмо, 2015. - 400 с.
6. Радченко, И.А. Технологии и инфраструктура Big Data: учебное пособие / И.А. Радченко, И.Н. Николаев. - Санкт-Петербург: Университет ИТМО, 2018. - 55 с.
7. Фрэнкс, Б. Укрощение больших данных: как извлекать знания из массивов информации с помощью глубокой аналитики / Б. Фрэнкс; Пер. с англ. А. Баранов. - М.: Манн, Иванов и Фербер, 2014. - 352 с.

**Контрольные вопросы:**

1. Дайте понятие «большие данные».
2. Назовите основные характеристики, присущие большим данным.
3. Каковы принципы работы с большими данными?
4. Охарактеризуйте основные предпосылки появления и этапы развития больших данных.
5. Опишите экосистему больших данных.
6. Что представляет собой озеро данных?
7. В каких сферах наиболее востребованы технологии больших данных?
8. Раскройте роль и место технологии больших данных в развитии научных исследований.
9. Приведите примеры направлений использования больших данных в научных исследованиях.
10. Раскройте тенденции и перспективы развития больших данных.
11. Кратко опишите современное состояние развития больших данных в России и за рубежом.
12. Перечислите проблемы безопасности больших данных.

**Раздел № 2 Архитектура системы обработки больших данных**

Уровни в системах обработки больших данных. Прием данных (Data Ingestion). Сбор данных (Data Staging). Анализ данных (Analysis Layer). Представление результатов (Consumption Layer).

**Рекомендуемая литература по разделу:**

1. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д.У. Джеффри; Пер. с англ. А.А. Слинкин. - М.: ДМК Пресс, 2016. - 498 с.
2. Базы данных в высокопроизводительных информационных системах / авт.-сост. Е.И. Николаев; Министерство образования и науки РФ, Федеральное государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». - Ставрополь: СКФУ, 2016. - 163 с. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=466799>
3. Дадян, Э.Г. Методы, модели, средства хранения и обработки данных: учебник / Э.Г. Дадян, Ю.А. Зеленков. - М.: Вузовский учебник: ИНФРА-М, 2017. - 168 с. - Режим доступа: <http://znanium.com/catalog/product/543943>
4. Майер-Шенбергер, В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / В. Майер-Шенбергер, К. Кукьер. – М.: Манн, Иванов и Фербер, 2014. - 240 с.
5. Марц, Н. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени / Н. Марц, Д. Уоррен. - М.: Вильямс, 2018. - 368 с.
6. Методы и модели исследования сложных систем и обработки больших данных: монография / И.Ю. Парамонов, В.А. Смагин, Н.Е. Косых, А.Д. Хомоненко; под редакцией В.А. Смагина и А.Д. Хомоненко. - Санкт-Петербург: Лань, 2020. - 236 с.
7. Просто о больших данных: пер. с англ. / Д. Гурвиц [и др.]. - М.: Эксмо, 2015. - 400 с.
8. Радченко, И.А. Технологии и инфраструктура Big Data: учебное пособие / И.А. Радченко, И.Н. Николаев. - Санкт-Петербург: Университет ИТМО, 2018. - 55 с.
9. Форман, Д. Много цифр. Анализ больших данных при помощи Excel / Форман Д.; Пер. с англ. Соколовой А. - М.: Альпина Пабл., 2016. - 461 с. - Режим доступа: <http://znanium.com/catalog/product/551044>
10. Фрэнкс, Б. Укрощение больших данных: как извлекать знания из массивов информации с помощью глубокой аналитики / Б. Фрэнкс; Пер. с англ. А. Баранов. - М.: Манн, Иванов и Фербер, 2014. - 352 с.
11. Ын, А. Теоретический минимум по Big Data. Всё что нужно знать о больших данных / А. Ын, К. Су. - СПб.: Питер, 2019. - 208 с.

**Контрольные вопросы:**

1. Какие уровни можно выделить в архитектуре системы обработки больших данных?
2. Кратко охарактеризуйте каждый из уровней системы обработки больших данных.
3. Назовите основные компоненты архитектуры для обработки больших данных.
4. Назовите источники больших данных.
5. Раскройте специфику сбора структурированных и неструктурированных данных.
6. Поясните, в чем может заключаться сложность обработки структурированных и неструктурированных данных?
7. Какие задачи может решать анализ больших данных?
8. Обозначьте методы анализа, применимые к большим данным.

**Раздел № 3** **Параллельные алгоритмы для работы с данными**

Операторы Map и Reduce. Оператор Map (предварительная обработка). Оператор Reduce (свертка). Лямбда-архитектура. Каппа-архитектура.

**Рекомендуемая литература по разделу:**

1. Базы данных в высокопроизводительных информационных системах / авт.-сост. Е.И. Николаев; Министерство образования и науки РФ, Федеральное государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». - Ставрополь: СКФУ, 2016. - 163 с. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=466799>
2. Дадян, Э.Г. Методы, модели, средства хранения и обработки данных: учебник / Э.Г. Дадян, Ю.А. Зеленков. - М.: Вузовский учебник: ИНФРА-М, 2017. - 168 с. - Режим доступа: <http://znanium.com/catalog/product/543943>
3. Марц, Н. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени / Н. Марц, Д. Уоррен. - М.: Вильямс, 2018. - 368 с.
4. Методы и модели исследования сложных систем и обработки больших данных: монография / И.Ю. Парамонов, В.А. Смагин, Н.Е. Косых, А.Д. Хомоненко; под редакцией В.А. Смагина и А.Д. Хомоненко. - Санкт-Петербург: Лань, 2020. - 236 с.
5. Парфенов, Ю. П. Постреляционные хранилища данных: учебное пособие для вузов / Ю. П. Парфенов; под научной редакцией Н. В. Папуловской. – Екатеринбург: Изд-во Урал. ун-та, 2019. - 121 с.
6. Просто о больших данных: пер. с англ. / Д. Гурвиц [и др.]. - М.: Эксмо, 2015. - 400 с.
7. Радченко, И.А. Технологии и инфраструктура Big Data: учебное пособие / И.А. Радченко, И.Н. Николаев. - Санкт-Петербург: Университет ИТМО, 2018. - 55 с.

**Контрольные вопросы:**

1. Назовите технологии, используемые для сбора и обработки больших данных. Раскройте их отличительные особенности.
2. Объясните назначение операторов Map и Reduce.
3. Каковы достоинства и недостатки модели распределенных вычислений MapReduce?
4. Поясните, что такое лямбда-архитектура?
5. Опишите структуру лямбда-архитектуры.
6. Поясните, что такое каппа-архитектура?
7. Опишите структуру каппа-архитектура.
8. Каковы различия между каппа-архитектурой и лямбда-архитектурой?

**Раздел № 4** **Программные платформы и системы для больших данных**

Системы управления потоками данных. Системы хранения больших данных. Платформы больших данных. Обработка данных в реальном времени. Системы управления большими данными. Аналитические платформы.

**Рекомендуемая литература по разделу:**

1. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д.У. Джеффри; Пер. с англ. А.А. Слинкин. - М.: ДМК Пресс, 2016. - 498 с.
2. Базы данных в высокопроизводительных информационных системах / авт.-сост. Е.И. Николаев; Министерство образования и науки РФ, Федеральное государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». - Ставрополь: СКФУ, 2016. - 163 с. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=466799>
3. Дадян, Э.Г. Методы, модели, средства хранения и обработки данных: учебник / Э.Г. Дадян, Ю.А. Зеленков. - М.: Вузовский учебник: ИНФРА-М, 2017. - 168 с. - Режим доступа: <http://znanium.com/catalog/product/543943>
4. Методы и модели исследования сложных систем и обработки больших данных: монография / И.Ю. Парамонов, В.А. Смагин, Н.Е. Косых, А.Д. Хомоненко; под редакцией В.А. Смагина и А.Д. Хомоненко. - Санкт-Петербург: Лань, 2020. - 236 с.
5. Парфенов, Ю. П. Постреляционные хранилища данных: учебное пособие для вузов / Ю. П. Парфенов; под научной редакцией Н. В. Папуловской. – Екатеринбург: Изд-во Урал. ун-та, 2019. - 121 с.
6. Просто о больших данных: пер. с англ. / Д. Гурвиц [и др.]. - М.: Эксмо, 2015. - 400 с.
7. Радченко, И.А. Технологии и инфраструктура Big Data: учебное пособие / И.А. Радченко, И.Н. Николаев. - Санкт-Петербург: Университет ИТМО, 2018. - 55 с.
8. Форман, Д. Много цифр. Анализ больших данных при помощи Excel / Форман Д.; Пер. с англ. Соколовой А. - М.: Альпина Пабл., 2016. - 461 с. - Режим доступа: <http://znanium.com/catalog/product/551044>

**Контрольные вопросы:**

1. Приведите примеры систем управления потоками больших данных.
2. Приведите примеры систем хранения больших данных.
3. Приведите примеры платформ больших данных.
4. Опишите состав экосистемы Hadoop.
5. Перечислите основные принципы работы фреймворка Hadoop.
6. Приведите примеры систем обработки больших данных в реальном времени.
7. Приведите примеры систем управления большими данными.
8. Приведите примеры аналитических платформ для работы с большими данными.

# Методические указания по лекционным занятиям

Лекция является важным звеном дидактического цикла обучения.

Лекционный материал усваивается студентами в двух формах: в процессе лекционного занятия и во время самостоятельной работы.

Изучение дисциплины начинается с вводной лекции, которая знакомит обучающихся с целью и назначением курса, его ролью и местом в системе учебных дисциплин.

В процессе лекций основной материал может быть представлен в виде презентации и сопровождается пояснениями лектора. Кроме этого, во время традиционной лекции может быть предложено во время занятия вести краткий конспект (допускается электронный). Желательно оставить в рабочих конспектах поля, на которых делать пометки, дополняющие материал прослушанной лекции, а также подчеркивающие особую важность тех или иных теоретических положений или вопросов. Рекомендуется задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.

Проработку лекционного материала рекомендуется проводить систематически, помечая непонятные места в конспекте. Если самостоятельно не удалось разобраться в материале, необходимо сформулировать вопросы и обратиться за консультацией к преподавателю. Перед очередной лекцией необходимо просмотреть материал предыдущей лекции.

Рекомендуемую в рабочей программе дисциплины литературу следует использовать после изучения данной темы в целях дополнительного, более углубленного изучения материала по тем вопросам, которые были даны лектором для самостоятельного изучения.

При изучении ряда тем используются специфические термины и понятия. Если встречается незнакомое понятие или термин, необходимо посмотреть его суть и содержание в словаре или ином источнике, выписать его значение в тетрадь для подготовки к занятиям или скопировать в электронный конспект.

При подготовке материала необходимо обращать внимание на точность определений, последовательность изучения материала, аргументацию, собственные примеры, анализ конкретных ситуаций.

Каждую неделю рекомендуется отводить время для повторения пройденного материала, изучения дополнительного, проверяя свои знания и умения по контрольным вопросам.

# Лабораторные работы

* 1. **Методические указания по лабораторным занятиям**

Лабораторная работа проводится в специально оборудованных учебных лабораториях. Продолжительность лабораторной работы – не менее 2-х академических часов. Перед проведением лабораторной работы преподавателем организуется инструктаж, а по ее окончании – обсуждение итогов.

Проведению лабораторной работы должна предшествовать проверка теоретической подготовленности обучающихся.

В процессе лабораторной работы как вида учебных занятий студенты выполняют одно или несколько лабораторных заданий под руководством преподавателя в соответствии с изучаемым содержанием учебного материала.

После выполнения каждой лабораторной работы студент демонстрирует результат выполнения преподавателю, отвечает на вопросы. Преподаватель оценивает работу.

Лабораторная работа выполняется каждым студентом самостоятельно.

Студенты, пропустившие занятия, выполняют лабораторные работы во внеурочное время.

* 1. **Методические рекомендации по самостоятельной подготовке к лабораторным занятиям**

Самостоятельная работа обучающихся по подготовке к лабораторным занятиям направлена на закрепление, расширение и углубление знаний, формирование и дальнейшее развитие необходимого комплекса умений, входящих в структуру профессиональной компетенции ПК\*-3, приобретение опыта конкретной деятельности.

Самостоятельная работа включает проработку теоретического материала, а также овладение различными системами и технологиями обработки больших данных.

С целью успешного выполнения лабораторных работ во время аудиторных занятий необходима предварительная самостоятельная подготовка, которая заключается в следующем:

– проработке материалов лекций и рекомендуемых источников информации по заявленной теме;

– освоении функциональных возможностей программных средств, используемых при выполнении заданий;

– подготовке к защите лабораторной работе (контрольные вопросы приведены после каждого задания);

– написании и оформлении отчета.

Отчет по лабораторной работе оформляется в соответствии с требованиями действующего стандарта организации СТО 02069024.101–2015 РАБОТЫ СТУДЕНЧЕСКИЕ. Общие требования и правила оформления [2].

Отчет должен содержать титульный лист; задание к лабораторной работе; результаты ее выполнения; основные выводы.

* 1. **Задания к лабораторным работам**

Тематика заданий к лабораторной работе «Система для обработки больших объемов данных Apache Hadoop»:

1. Установка и конфигурирование программного комплекса Apache Hadoop.
2. Работа с файлами в системе HDFS.
3. Импорт и обработка данных в кластере Hadoop.
4. Написание простых MapReduce-приложений.
5. Размещения набора данных по заданной тематике. Построение поисковых запросов на языке Pig.

# Методические указания по самостоятельной работе

Самостоятельная учебная деятельность является необходимым условием успешного образования. Многие профессиональные умения, способность мыслить и обобщать, делать выводы и строить суждения, выступать и слушать других, – все это развивается в процессе самостоятельной работы студентов.

Самостоятельная учебная работа является равноправной формой учебных занятий, наряду с лекциями и лабораторными занятиями, экзаменом, но реализуемая во внеаудиторное время.

Самостоятельная работа – вид учебной деятельности, базирующийся на выполнении студентами комплекса заданий при консультационно-координирующей помощи преподавателя, ориентированный на формирование результатов обучения, выраженных соответствующими компетенциями.

Изучение дисциплины предполагает не только познавательную деятельность, которую студенты осуществляют во время лекций и лабораторных занятий, но и самостоятельную работу, осуществляемую вне аудиторных занятий. Успешное усвоение учебного материала возможно только при комплексном подходе, состоящем в получении новой информации в ходе лекции или лабораторного занятия; ее понимании и обобщении; записи в собственной интерпретации в виде текста, схем, таблиц; самостоятельном изучении и конспектировании рекомендованной учебной литературы; выполнении учебных заданий; подготовке к учебным дискуссиям и рубежному контролю.

Самостоятельная работа студента состоит, во-первых, из деятельности студентов во всех организационных формах учебных занятий и во внеаудиторное время, когда они самостоятельно изучают теоретический материал, определенный содержанием рабочей программы. Во-вторых, выполнения учебных заданий, которые рекомендованы студенту во время обучения.

Самостоятельная внеаудиторная работа студента может включать в себя: типовые виды (выполняются всеми студентами) и нестандартные (выполняются не всеми, участие в этих видах зависит от наличия сформированности тех или иных компетенций).

Подготовка студента к занятиям должна включать в себя не только непосредственное выполнение домашнего задания. Она должна предусматривать тот факт, что последующее занятие будет направлено на изучение нового теоретического и/или практического материала. Такая подготовка предполагает изучение рабочей программы, установление связи с ранее полученными знаниями, выделение наиболее значимых и актуальных проблем, на изучении которых следует обратить особое внимание.

# Перечень дискуссионных тем для проведения круглого стола

1. Предпосылки появления и этапы развития больших данных.
2. Сферы использования технологии больших данных.
3. Роль и место технологии больших данных в развитии научных исследований.
4. Направления использования больших данных в научных исследованиях.
5. Тенденции и перспективы развития больших данных.
6. Современное состояние развития больших данных в России и за рубежом.
7. Большие проблемы больших данных.
8. Большие данные и блокчейн: вместе или врозь?
9. Большие данные и интернет вещей – идеальное сочетание?
10. Оборудование для работы с большими данными.

# Методические указания по подготовке к рубежному контролю и промежуточной аттестации

**Подготовка к рубежному контролю успеваемости**

Рубежный контроль может проводиться в форме тестирования или устного/письменного опроса, включающего в себя ответы на теоретические вопросы и решение задач.

При подготовке к рубежному контролю обучающимся следует придерживаться следующих рекомендаций:

1) готовиться к теоретической части рубежного контроля целесообразно во время изучения соответствующего материала, записывая ответы на вопросы к экзамену;

2) при подготовке к сдаче практической части рубежного контроля целесообразно использовать тщательно разобранные решения заданий лабораторных работ;

3) если подготовка к рубежному контролю вызывает трудности, то допускаются консультации у преподавателя на лекционных или лабораторных занятиях.

Оценка по результатам промежуточной аттестации учитывает результаты рубежного контроля успеваемости.

**Подготовка к промежуточной аттестации в форме экзамена**

Промежуточная аттестация может проводиться как в устной, так и в письменной форме.

При подготовке к экзамену обучающимся следует придерживаться следующих рекомендаций:

1) готовиться к сдаче теоретической части экзамена целесообразно во время изучения соответствующего материала в течение всего семестра, записывая ответы на вопросы к экзамену;

2) при подготовке к сдаче практической части экзамена целесообразно использовать тщательно разобранные решения заданий лабораторных работ;

3) если подготовка к экзамену вызывает трудности, то допускаются консультации у преподавателя на лекционных или лабораторных занятиях.

# Вопросы к экзамену

1. Понятие «большие данные».
2. Характеристики, присущие большим данным.
3. Принципы работы с большими данными.
4. Предпосылки появления и этапы развития больших данных.
5. Экосистема больших данных.
6. Понятие «озеро данных».
7. Сферы использования технологии больших данных.
8. Роль и место технологии больших данных в развитии научных исследований.
9. Направления использования больших данных в научных исследованиях.
10. Тенденции и перспективы развития больших данных.
11. Современное состояние развития больших данных в России и за рубежом.
12. Проблемы безопасности больших данных.
13. Архитектура системы обработки больших данных.
14. Компоненты архитектуры для обработки больших данных.
15. Источники больших данных.
16. Специфика сбора структурированных и неструктурированных данных.
17. Обработка структурированных и неструктурированных данных.
18. Анализ больших данных. Методы анализа, применимые к большим данным.
19. Технологии, используемые для сбора и обработки больших данных.
20. Назначение операторов Map и Reduce. Модель распределенных вычислений MapReduce.
21. Лямбда-архитектура.
22. Каппа-архитектура.
23. Системы управления потоками больших данных.
24. Системы хранения больших данных.
25. Платформы больших данных.
26. Apache Hadoop. Состав экосистемы Hadoop. Основные принципы работы фреймворка Hadoop.
27. Системы обработки больших данных в реальном времени.
28. Системы управления большими данными.
29. Аналитические платформы для работы с большими данными.
30. Оборудование для работы с большими данными.

# Список использованных источников

1. Об утверждении федерального государственного образовательного стандарта высшего образования - магистратура по направлению подготовки 09.04.02 Информационные системы и технологии: Приказ Минобрнауки России от 19.09.2017 № 917 (Зарегистрировано в Минюсте России 16.10.2017 № 48550). – Режим доступа: <http://www.consultant.ru/document/cons_doc_LAW_280720/>
2. СТО 02069024.101–2015 Работы студенческие. Общие требования и [правила оформления](http://www.osu.ru/docs/official/standart/standart_101-2015_.pdf). – Оренбург: ОГУ, 2015. – Режим доступа: <www.osu.ru/docs/official/standart/standart_101-2015_.pdf>
3. Радченко, И.А. Технологии и инфраструктура Big Data: учебное пособие / И.А. Радченко, И.Н. Николаев. – Санкт-Петербург: Университет ИТМО, 2018. – 55 с.
4. Рабочая программа дисциплины «Б1.Д.В.3 Системы и технологии обработки больших данных» / сост. В.В. Запорожко. – Оренбург: ОГУ, 2020. – 7 с.