

МИНОБРНАУКИ РОССИИ
Ярославский государственный университет им. П.Г. Демидова

Кафедра цифровых технологий и машинного обучения

УТВЕРЖДАЮ

Декан физического факультета



И.С. Огнев

«21» мая 2024 г.

Рабочая программа дисциплины
«Большие данные и их обработка»

Направление подготовки
03.03.03 Радиофизика

Направленность (профиль)
Технологии беспроводной связи

Форма обучения
Очная

Программа одобрена
на заседании кафедры

от «26» апреля 2024 года, протокол № 8

Программа одобрена НМК
физического факультета

протокол № 5 от «30» апреля 2024 года

Ярославль

1. Цели освоения дисциплины

Цель дисциплины «**Большие данные и их обработка**» заключается в формировании у студентов фундаментальных знаний, практических навыков и аналитических умений, необходимых для работы с большими объемами данных в различных профессиональных контекстах.

Цели освоения дисциплины:

1. Понимание природы больших данных: студенты должны понять особенности больших данных, такие как объем, скорость, разнообразие и ценность. Освоение концептуальных основ работы с большими данными, включая подходы к их обработке, хранению и анализу.

2. Развитие навыков системного мышления: формирование у студентов способности рассматривать большие данные в рамках целостных информационных систем, учитывать взаимодействие между различными компонентами систем обработки данных, а также их влияние на производительность и эффективность.

3. Обучение методам обработки и анализа больших данных: ознакомление с методологиями обработки больших данных, от этапа сбора и предобработки до анализа и интерпретации результатов.

4. Формирование аналитических умений: студенты должны научиться извлекать ценные инсайты из данных, проводить анализ на основе различных подходов и формулировать выводы, которые могут быть применены для принятия решений.

5. Подготовка к работе с современными технологиями: создание базы знаний, необходимой для работы с большими объемами данных в условиях динамично развивающихся технологий и появления новых методов обработки и анализа.

6. Осознание этических аспектов работы с данными: формирование у студентов осознания этических и юридических аспектов работы с данными, включая защиту конфиденциальной информации, соблюдение законодательства о данных и социальную ответственность аналитика.

Задачи дисциплины:

1. Ознакомить студентов с основными принципами и подходами к обработке больших данных.

2. Научить оценивать сложность и требования к обработке данных в зависимости от поставленных задач.

3. Развить навыки критического мышления для выбора оптимальных методов анализа данных.

4. Способствовать формированию умений интерпретации результатов анализа данных для их последующего использования в профессиональной деятельности.

2. Место дисциплины в структуре ОП бакалавриата

Данная дисциплина относится к дисциплинам обязательного цикла. Для успешного освоения курса требуется знание основ программирования, работы с базами данных и математической статистики.

Знания и навыки, полученные при изучении данной дисциплины, будут востребованы при выполнении курсовых и выпускных квалификационных работ, а также в последующей трудовой деятельности обучающихся.

3. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

Процесс изучения дисциплины направлен на формирование следующих элементов компетенций в соответствии с ФГОС ВО, ОП ВО и приобретения следующих знаний, умений, навыков и (или) опыта деятельности:

Формируемая компетенция (код и формулировка)	Индикатор достижения компетенции (код и формулировка)	Перечень планируемых результатов обучения
Общепрофессиональные компетенции		
ОПК-4. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ИД_ОПК-4.2. Демонстрирует умение проектировать решение конкретных задач профессиональной деятельности, выбирая оптимальные способы их решения с использованием современных информационных технологий.	<p>Знать:</p> <ul style="list-style-type: none"> - Основные принципы работы современных информационных технологий (ИТ), включая архитектуру вычислительных систем и сети. - Методы анализа и обработки больших данных, подходы к оптимизации рабочих процессов. - Основные платформы и инструменты для обработки данных, такие как Hadoop, Spark, и NoSQL базы данных. - Подходы к проектированию информационных систем, включая методы моделирования данных и архитектуры. <p>Уметь:</p> <ul style="list-style-type: none"> - Проектировать решения профессиональных задач с использованием современных ИТ, выбирая подходящие инструменты и технологии. - Анализировать и структурировать данные, определять требования к решению задачи. - Оценивать производительность, надежность и масштабируемость проектируемых решений. - Настраивать и использовать платформы обработки данных (Hadoop, Spark) и базы данных. <p>Владеть:</p> <ul style="list-style-type: none"> - Навыками работы с инструментами обработки данных, включая настройку кластеров и анализ производительности.

		<ul style="list-style-type: none"> - Методами визуализации данных и построения дашбордов с использованием современных библиотек (например, Tableau, matplotlib, seaborn). - Навыками управления проектами и документооборота при проектировании информационных систем. - Практическими навыками использования современных технологий ИТ в прикладных задачах.
ОПК-5. Способен разрабатывать алгоритмы и компьютерные программы, пригодные для практического применения	ИД_ОПК-5.1. Знает основные приемы разработки алгоритмов и компьютерных программ для решения теоретических и экспериментальных задач.	<p>Знать:</p> <ul style="list-style-type: none"> - Основные алгоритмы обработки данных, включая сортировку, поиск, кластеризацию и методы машинного обучения. - Подходы к проектированию алгоритмов, такие как динамическое программирование, жадные алгоритмы, и методы ветвей и границ. - Основы языков программирования (Python, SQL) для реализации алгоритмов. - Принципы тестирования и отладки программного кода. <p>Уметь:</p> <ul style="list-style-type: none"> - Разрабатывать алгоритмы и программы для решения прикладных задач, таких как обработка данных, поиск оптимальных решений, и построение прогнозных моделей. - Анализировать эффективность алгоритмов с точки зрения временной и пространственной сложности. - Реализовывать алгоритмы на языках программирования (Python, SQL) для работы с данными. - Использовать библиотеки и фреймворки для автоматизации и ускорения разработки программ. <p>Владеть:</p> <ul style="list-style-type: none"> - Практическими навыками написания, тестирования и оптимизации программного кода.

		<ul style="list-style-type: none"> - Навыками работы с инструментами разработки, отладки и контроля версий (Git, IDE, CI/CD). - Подходами к разработке программ с учетом требований надежности, безопасности и масштабируемости. - Навыками адаптации алгоритмов к специфике профессиональных задач.
--	--	---

4. Объем, структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 3 зачетных единицы, 108 акад. часов.

Очное отделение

№ п/п	Темы (разделы) дисциплины, их содержание	Семестр	Виды учебных занятий, включая самостоятельную работу студентов, и их трудоемкость (в академических часах)						Формы текущего контроля успеваемости Форма промежуточной аттестации (по семестрам)
			Контактная работа						
			лекции	практические	лабораторные	консультации	аттестационные испытания	самостоятельная работа	
1	Введение в большие данные: концепции и области применения	6	4	4				4,5	Тестирование, задание для самостоятельной работы
2	Инфраструктура обработки данных: Hadoop и HDFS	6	4	4				4,5	Тестирование, задание для самостоятельной работы
3	Распределенные вычисления: основы Apache Spark	6	4	4				4,5	Тестирование, задание для самостоятельной работы
4	Хранение данных: основы NoSQL баз данных (MongoDB, Cassandra)	6	4	4				4,5	Тестирование, задание для самостоятельной работы
5	Методы анализа данных: кластеризация и регрессия	6	5	5				4,5	Тестирование, задание для самостоятельной работы
6	Визуализация данных: инструменты и подходы	6	5	5				4,5	Тестирование, задание для самостоятельной работы
7	Этика и конфиденциальность в анализе данных	6	4	4				4,5	Тестирование, задание для самостоятельной работы

8	Технологии обработки данных в реальном времени	6	4	4				4,5	Тестирование, задание для самостоятельной работы
						2	0,3	1,7	Зачет
	Всего		34	34		2	0,3	37,7	

Содержание разделов дисциплины

Тема №1

Введение в большие данные: концепции и области применения

Определение больших данных, их ключевые характеристики (объем, скорость, разнообразие). Обзор основных этапов работы с большими данными: сбор, хранение, обработка, анализ. Примеры применения в бизнесе, медицине, науке. Современные вызовы и перспективы развития технологий обработки данных.

Тема №2

Инфраструктура обработки данных: Hadoop и HDFS

Архитектура Hadoop: компоненты и принципы работы. HDFS как распределенная файловая система: основные операции, репликация, устойчивость к сбоям. Принципы работы MapReduce: концепция разделения задач, использование кластеров. Настройка и развертывание Hadoop-кластера.

Тема №3

Распределенные вычисления: основы Apache Spark

Основы Apache Spark: архитектура и компоненты. Преимущества Spark по сравнению с MapReduce. Работа с RDD (Resilient Distributed Datasets) и DataFrame. Инструменты для обработки данных в Spark: Spark SQL, Spark Streaming, MLlib. Примеры реальных сценариев использования.

Тема №4

Хранение данных: основы NoSQL баз данных (MongoDB, Cassandra)

Ключевые особенности NoSQL баз: гибкость схем данных, масштабируемость, отказоустойчивость. Основы MongoDB: работа с документами и коллекциями. Основы Cassandra: архитектура, принципы распределенного хранения данных. Примеры использования NoSQL баз в современных системах.

Тема №5

Методы анализа данных: кластеризация и регрессия

Обзор методов анализа данных. Кластеризация: постановка задачи, методы К-средних, иерархической кластеризации. Регрессия: линейная и полиномиальная, оценка качества моделей. Примеры применения методов анализа данных в различных отраслях.

Тема №6

Визуализация данных: инструменты и подходы

Значение визуализации в анализе данных. Обзор популярных инструментов: Tableau, Matplotlib, Seaborn. Основные подходы к построению графиков и дашбордов. Примеры визуализации сложных данных и интерпретации результатов анализа.

Тема №7

Этика и конфиденциальность в анализе данных

Основные этические аспекты работы с большими данными: конфиденциальность, защита персональных данных, прозрачность анализа. Проблемы предвзятости данных и алгоритмов. Обзор законодательства и стандартов в области защиты данных. Влияние анализа данных на общество.

Тема №8

Технологии обработки данных в реальном времени

Обзор технологий для обработки потоковых данных: Apache Kafka, Spark Streaming, Flink. Основные концепции: очереди сообщений, обработка событий, микропакеты. Примеры применения: анализ веб-трафика, мониторинг систем, обработка финансовых данных. Реализация и настройка систем потоковой обработки. Вызовы и ограничения реального времени.

5. Образовательные технологии, в том числе технологии электронного обучения и дистанционные образовательные технологии, используемые при осуществлении образовательного процесса по дисциплине

В процессе обучения используются следующие образовательные технологии:

Вводная лекция – дает первое целостное представление о дисциплине и ориентирует студента в системе изучения данной дисциплины. Студенты знакомятся с назначением и задачами курса, его ролью и местом в системе учебных дисциплин и в системе подготовки в целом. Дается краткий обзор курса, история развития науки и практики, достижения в этой сфере, имена известных ученых, излагаются перспективные направления исследований. На этой лекции высказываются методические и организационные особенности работы в рамках данной дисциплины, а также дается анализ рекомендуемой учебно-методической литературы.

Академическая лекция с элементами лекции-беседы – последовательное изложение материала, осуществляемое преимущественно в виде монолога преподавателя. Элементы лекции-беседы обеспечивают контакт преподавателя с аудиторией, что позволяет привлекать внимание студентов к наиболее важным темам дисциплины, активно вовлекать их в учебный процесс, контролировать темп изложения учебного материала в зависимости от уровня его восприятия.

При чтении лекций используется технология проблемного обучения (последовательное и целенаправленное выдвижение перед студентом познавательных задач, разрешая которые студенты активно усваивают знания). Курс построен на принципах системного подхода к отбору программного материала и определению последовательности его изучения студентами, что предусматривает глубокое изучение предмета.

Практическое занятие – занятие, посвященное освоению конкретных умений и навыков и закреплению полученных на лекции знаний.

Консультации – вид учебных занятий, являющийся одной из форм контроля самостоятельной работы студентов. На консультациях по просьбе студентов рассматриваются наиболее сложные моменты при освоении материала дисциплины,

преподаватель отвечает на вопросы студентов, которые возникают у них в процессе самостоятельной работы.

Контроль качества усвоения программного материала проводится с учетом работы студентов на лекциях и лабораторных занятиях.

Самостоятельная работа студентов заключается в проработке конспекта лекций, изучении рекомендованной литературы, выполнении домашних и индивидуальных заданий.

6. Перечень лицензионного и (или) свободно распространяемого программного обеспечения, используемого при осуществлении образовательного процесса по дисциплине

В процессе осуществления образовательного процесса по дисциплине используются:

1. Apache Hadoop – платформа для распределенного хранения и обработки данных.
2. Apache Spark – инструмент для обработки данных в распределенных системах и работы с потоковыми данными.
3. MongoDB – NoSQL база данных для хранения документов.
4. Apache Cassandra – NoSQL база данных для распределенного хранения данных.
5. Python – язык программирования с библиотеками pandas, matplotlib, seaborn для анализа и визуализации данных.
6. Tableau Public – инструмент для визуализации данных (бесплатная версия для учебных целей).
7. Jupyter Notebook – среда для выполнения кода и визуализации данных.
8. Apache Kafka – система для обработки потоковых данных.
9. VS Code – редактор кода с плагинами для Python и работы с базами данных.
10. Docker – инструмент для контейнеризации приложений, позволяющий развертывать среды разработки.
11. Linux OS – операционная система для развертывания кластеров Hadoop и Spark.
12. Git – система контроля версий для управления проектами.
13. Google Colab – облачная платформа для анализа и визуализации данных.

Этот перечень включает популярное и доступное ПО, широко применяемое в профессиональной практике.

7. Перечень современных профессиональных баз данных и информационных справочных систем, используемых при осуществлении образовательного процесса по дисциплине (при необходимости)

В процессе осуществления образовательного процесса по дисциплине используются:

Автоматизированная библиотечно-информационная система «БУКИ-NEXT»
http://www.lib.uniyar.ac.ru/opac/bk_cat_find.php

8. Перечень основной и дополнительной учебной литературы, ресурсов информационно-телекоммуникационной сети «Интернет» (при необходимости), рекомендуемых для освоения дисциплины

а) основная литература

1. Макшанов, А. В. Большие данные. Big Data : учебник для вузов / А. В. Макшанов, А.Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2021. — 188 с. — ISBN 978-5-8114-6810-2. — Текст: электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165835>

2. Python: Искусственный интеллект, большие данные и облачные вычисления Дейтел Пол, Дейтел Харви / Пол Дейтел, Харви Дейтел. URL: <https://ibooks.ru/reading.php?short=1&productid=371701>.
3. Григорьев Ю. А., Плутенко А. Д., Плужникова О. Ю. Реляционные базы данных и системы NoSQL : учеб. пособие / Григорьев Ю. А., Плутенко А. Д., Плужникова О. Ю. ; Амурский гос. ун-т, МГТУ им. Н. Э. Баумана. - Благовещенск : Изд-во Амурского гос. ун-та, 2018. - 420 с. : ил. - Библиогр.: с. 417-420. - ISBN 978-5-93493-308-2.

б) дополнительная литература

1. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учеб. пособие для вузов / Гмурман В. Е. - 3-е изд., перераб. и доп. - М. : Высш. шк., 1979. - 399 с.
2. Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/131721>
3. АНАЛИЗ ДАННЫХ. Учебник для вузов / Под ред. Мхитаряна В.С. - 2021. - URL: <https://urait.ru/book/54898B0C-E25B-4962-9BD4-96129F3BEC2B>.
4. Изучаем Spark: молниеносный анализ данных / Х. Карау, Э. Конвински, П. Венделл, ISBN 978-5-97060-323-9. М. Захария. — Москва : ДМК Пресс, 2015. — 304 с. — ISBN 978-5-97060-323-9. — Текст : электронный // Лань : электронно-библиотечная система. URL: <https://e.lanbook.com/book/90118>

в) ресурсы сети «Интернет»:

Электронная библиотека учебных материалов ЯрГУ
http://www.lib.uni-yar.ac.ru/opac/bk_cat_find.php).

9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине включает в свой состав специальные помещения:

- учебные аудитории для проведения занятий лекционного типа;
- учебные аудитории для проведения практических работ;
- учебные аудитории для проведения групповых и индивидуальных консультаций,
- учебные аудитории для проведения текущего контроля и промежуточной аттестации;
- помещения для самостоятельной работы;
- помещения для хранения и профилактического обслуживания технических средств обучения.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

Число посадочных мест в аудитории для занятий лекционного типа больше либо равно списочному составу группы обучающихся.

Автор:

Доцент кафедры цифровых технологий
 И машинного обучения, к.т.н.

_____ И.В. Апальков

**Приложение №1 к рабочей программе дисциплины
«Большие данные и их обработка»**

**Фонд оценочных средств
для проведения текущей и промежуточной аттестации студентов
по дисциплине**

**1. Типовые контрольные задания или иные материалы,
необходимые для оценки знаний, умений, навыков и (или) опыта деятельности,
характеризующих этапы формирования компетенций**

**1.1. Контрольные задания и иные материалы,
используемые в процессе текущей аттестации**

Задания для самостоятельной работы

1. Определите ключевые характеристики больших данных (объем, скорость, разнообразие) на примере выбранной отрасли (например, медицина, финансы, транспорт).
2. Найдите три практических применения больших данных и кратко объясните, как они помогают улучшить процессы.
3. Настройте Hadoop-кластер на виртуальной машине (с использованием Docker или VirtualBox). Подготовьте отчет с описанием конфигурации.
4. Выполните задачу по обработке данных с помощью HDFS: загрузите файл, выполните операции чтения и удаления.
5. Реализуйте простой сценарий MapReduce для подсчета слов в большом текстовом файле.
6. Используйте Apache Spark для выполнения задания по подсчету среднего значения набора данных. Предоставьте код и результаты выполнения.
7. Создайте коллекцию в MongoDB и выполните основные операции (вставка, обновление, удаление).
8. Разработайте запрос для получения аналитики по данным, хранящимся в Cassandra (например, среднее значение по столбцу, группировка данных).
9. Выполните кластеризацию набора данных (например, Iris) с использованием алгоритма К-средних. Проанализируйте результаты.
10. Постройте линейную регрессию для прогноза значений в выбранном наборе данных. Оцените качество модели.
11. Постройте график распределения данных (гистограмма, диаграмма рассеяния) с использованием matplotlib или seaborn.
12. Создайте интерактивный дашборд с визуализацией ключевых метрик выбранного набора данных с помощью Tableau Public.
13. Опишите, какие аспекты конфиденциальности необходимо учитывать при работе с большим набором персональных данных (например, в здравоохранении).
14. Разработайте рекомендации по минимизации предвзятости в алгоритмах обработки данных.

1.2. Список вопросов и (или) заданий для проведения аттестации

Список вопросов к зачету

1. Что такое большие данные? Опишите их основные характеристики.
2. В чем отличие больших данных от традиционных данных?
3. Какие основные этапы обработки больших данных?
4. Приведите примеры применения технологий больших данных в реальных задачах.
5. Какие основные вызовы связаны с обработкой больших данных?
6. Что такое Hadoop? Опишите его основные компоненты.
7. Для чего используется HDFS? В чем его преимущества?
8. Как работает MapReduce? Объясните концепцию на примере.
9. Какие существуют аналоги Hadoop? Сравните их основные особенности.
10. Какие ограничения имеет Hadoop?
11. Что такое Apache Spark? Как он отличается от Hadoop MapReduce?
12. Опишите основные компоненты Spark: RDD, DataFrame, DataSet.
13. Для чего используется Spark SQL?
14. Как Spark обрабатывает потоковые данные?
15. Какие задачи решаются с помощью Spark MLlib?
16. В чем ключевые отличия NoSQL баз данных от реляционных?
17. Какие модели данных поддерживают NoSQL базы?
18. Что такое MongoDB? Для каких задач она подходит?
19. Как устроена архитектура Cassandra?
20. Какие существуют методы обеспечения отказоустойчивости в NoSQL базах?
21. Какие задачи решаются с помощью кластеризации?
22. Что такое алгоритм К-средних? Какие у него ограничения?
23. Как проводится оценка качества кластеризации?
24. Что такое линейная регрессия? В каких задачах она применяется?
25. В чем отличие линейной регрессии от логистической?
26. Почему визуализация данных важна для анализа?
27. Какие основные виды графиков используются для визуализации?
28. Что такое дашборд, и как он используется?
29. Какие библиотеки Python подходят для визуализации данных?
30. В чем преимущества Tableau перед другими инструментами визуализации?
31. Какие основные этические проблемы возникают при обработке больших данных?
32. Что такое предвзятость данных, и как с ней бороться?
33. Как обеспечивается конфиденциальность персональных данных?
34. Какие правовые аспекты регулируют обработку больших данных?
35. Какие риски несет неправильная интерпретация данных?
36. Что такое потоковые данные? Приведите примеры.
37. Для чего используется Apache Kafka?
38. Как Spark Streaming обрабатывает потоковые данные?
39. Какие вызовы связаны с обработкой данных в реальном времени?
40. Как обеспечить отказоустойчивость в потоковой обработке?
41. Что такое архитектура Lambda? В чем ее преимущества?
42. Какие метрики используются для оценки производительности систем обработки данных?
43. Как выполняется масштабирование Hadoop-кластера?
44. Что такое Data Lake? В чем его отличие от Data Warehouse?
45. Какие инструменты используются для мониторинга систем больших данных?
46. В чем различия между распределенным и параллельным вычислением?

47. Как происходит интеграция Hadoop с другими системами?
48. Что такое обработка данных на краю сети (Edge Computing)?
49. Какие проблемы возникают при переносе больших данных между системами?
50. Как выбрать подходящую технологию для обработки больших данных?
51. Что такое регуляризация, и как она используется в анализе данных?
52. Какие преимущества дает использование кластеров в обработке больших данных?
53. Что такое автоматическое масштабирование, и как его настроить?
54. Как работают функции шардирования в MongoDB?
55. Какие шаги необходимы для предобработки данных?
56. Что такое распределенная файловая система, и как она организована?
57. Какие алгоритмы используются для потоковой кластеризации данных?
58. В чем разница между OLAP и OLTP системами?
59. Что такое CAP-теорема, и как она связана с базами данных?
60. Какие технологии используются для анализа больших графов?

Критерии оценивания ответов на вопросы билета

Критерий	Пороговый уровень (на «удовлетворительно»)	Продвинутый уровень (на «хорошо»)	Высокий уровень (на «отлично»)
Соответствие ответа вопросу	Хотя бы частичное (не относящееся к вопросу не подлежит проверке)	Полное	Полное
Наличие примеров	Имеются отдельные примеры	Много примеров	Есть практически ко всем утверждениям
Содержание ответа	Понятийные вопросы изложены с классификациями, проблемные с постановкой проблемы и изложением различных точек зрения. Имеются ошибки или пробелы.	Ответ почти полный, без ошибок, не хватает отдельных элементов и тонкостей	Исчерпывающий полный ответ

2. Описание процедуры выставления оценки

Изучение дисциплины заканчивается зачетом. Для подготовки ответа на вопрос билета отводится не менее 40 минут.

Оценка «зачтено» выставляется, если ответ на вопрос билета дан не ниже, чем на пороговом уровне.

Оценка «не зачтено» выставляется, если ответ на вопрос билета дан ниже, чем на пороговом уровне.

Приложение №2 к рабочей программе дисциплины «Большие данные и их обработка»

Методические указания для студентов по освоению дисциплины

Одной из основных форм усвоения учебного материала по дисциплине «**Большие данные и их обработка**» является самостоятельная работа студента, причем в достаточно большом объеме. По всем темам предусмотрены задания самостоятельной работы, на которых происходит закрепление изученного материала и отработка навыков анализа и синтеза систем на базе методов искусственного интеллекта.

Изучение дисциплины заканчивается зачетом. Оценка выставляется на основании уровня сформированности указанных компетенций, который оценивается как средняя оценка по совокупности параметров: оценки за самостоятельные задания и ответы на вопросы билета.

Освоить вопросы дисциплины «**Большие данные и их обработка**» самостоятельно студенту достаточно сложно. Посещение всех предусмотренных лекционных занятий и занятий по выполнению лабораторных работ является совершенно необходимым. Без упорных и регулярных самостоятельных занятий в течение семестра сдать зачет практически невозможно.