МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«Национальный исследовательский ядерный университет «МИФИ»

Обнинский институт атомной энергетики –

филиал федерального государственного автономного образовательного учреждения высшего образования
«Национальный исследовательский ядерный университет «МИФИ»

(ИАТЭ НИЯУ МИФИ)

Одобрено на заседании УМС
ИАТЭ НИЯУ МИФИ Протокол
от 30.08.2022 № 2-8/2022

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

«Технологии программмирования для больших данных»

название дисциплины

для студентов направления подготовки

09.04.01 Информатика и вычислительная техника

профиль:
Большие данные и машинное обучение для атомной энергетики

Форма обучения: очная

г. Обнинск 2022 г.

Программа составлена в соответствии с образовательным стандартом высшего образования НИЯУ МИФИ по направлению подготовки 09.04.01 «Информатика и вычислительная техника».

Программу составил:

_____ О.А. Мирзеабасов, доцент отд. ИКС, к.т.н.

Рецензент:

_____

Программа рассмотрена на заседании отделения интеллектуальных кибернетических систем (О)
(протокол № ____ от «____»_____2022 г.)

Руководитель образовательной программы
090401 «Информатика и вычислительная техника»
_____ Старков С.О.
«____»_____2022 г.

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения ОПОП магистратуры обучающийся должен овладеть следующими результатами обучения по дисциплине:

Коды компетенций	Результаты освоения ООП *Содержание компетенций*	Перечень планируемых результатов обучения по дисциплине
ОПК-2	способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	З-ОПК-2 Знать: современные информационные и интеллектуальные технологии и инструментальные средства разработки алгоритмов и программного обеспечения, алгоритмические языки программирования,операционные системы и оболочки, современные среды разработки программного обеспечения У-ОПК-2 Уметь: выбирать современные информационные и интеллектуальные технологии и инструментальные средства разработки алгоритмов и программногообеспечения, составлять алгоритмы, писать и отлаживать коды на языке программирования, тестировать работоспособность программы, интегрировать программные модули В-ОПК-2 Владеть: навыками применения современных информационных и интеллектуальных технологий и инструментальных средств разработки алгоритмов и программного обеспечения, языками программирования, навыками отладки и тестирования работоспособности программ, применяемых для решения профессиональных задач
ОПК-5	способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных	З-ОПК-5 Знать: современные информационные технологии и инструментальные средства разработки программного и аппаратного обеспечения информационных и автоматизированных систем У-ОПК-5 Уметь: выбирать и применять современные инструментальные

	систем	средства разработки программного и аппаратного обеспечения информационных и автоматизированных систем в соответствии с решаемыми задачами В-ОПК-5 Владеть: навыками разработки и модернизации программного и аппаратного обеспечения информационных и автоматизированных систем с применением современных инструментальных средств
ОПК-8	способен осуществлять эффективное управление разработкой программных средств и проектов	З-ОПК-8 Знать: действующее законодательство в области управления разработкой программных средств и проектов, цели, принципы, функции, объекты управления проектами, основные инструменты проведения реинжиниринга бизнес-процессов, методы сбора информации, подходы к организации деятельности специфических служб по управлению проектами, основные методологии управления проектами У-ОПК-8 Уметь: проектироватьорганизационную структуру, осуществлять распределение полномочий и ответственности на основе их делегирования В-ОПК-8 Владеть: современными инструментальными средствами по управлению проектами, навыками организации деятельности по управлению проектами, методами оценки эффективности
ПК-4	разрабатывать, согласовывать и выпускать все виды проектной документации	З-ПК-4 Знать: требования ГОСТ ЕСКД, ЕСТД и ЕСПД по разработке и выпуску всех видов проектной документации в области информатики и вычислительной техники У-ПК-4 Уметь: выполнять разработку, согласование и выпуск всех видов проектной документации В-ПК-4 Владеть: современными инструментальными средствами по разработке и выпуску проектной документации

2. Место дисциплины в структуре ОПОП магистратуры

Дисциплина реализуется в рамках обязательной части.

Для освоения дисциплины необходимы компетенции, сформированные в рамках изучения следующих дисциплин: «Программирование», «Технологии программирования», «Объектно-ориентированное программирование», «Большие данные».

Дисциплины и/или практики, для которых освоение данной дисциплины необходимо как предшествующее: «Обработка и статистический анализ больших данных», «Производственная практика: научно-исследовательская работа», «Высокопроизводительные вычисления», «Производственная практика: технологическая (проектно-технологическая) практика».

Дисциплина изучается на 1 курсе в 1 и 2 семестрах.

3. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам занятий) и на самостоятельную работу обучающихся

Вид учебной работы	Всего часов	Семестры			
		1	2		
Аудиторные занятия (всего)	96	48	48		
В том числе:	-	-	-	-	-
Практические занятия	-	-			
Семинары	-	-			
Лабораторные работы	32	16	16		
В том числе:	-	-	-	-	-
интерактивные формы обучения (лекции)	32	16	16		
интерактивные формы обучения (практические занятия/семинары)	32	16	16		
Самостоятельная работа (всего)	120	60	60		
В том числе:	-	-	-	-	-
Учебный проект (работа)	-	-			
Расчетно-графические работы	-	-			
Реферат	-	-			
Вид промежуточной аттестации (зачет, экзамен)	Экзамен	Зачет	Экзамен		

ОБЩАЯ ТРУДОЕМКОСТЬ					
час	252	108	144		
зач.ед.	7	3	4		

4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины и трудоемкость по видам учебных занятий (в академических часах)

№ п/п	Наименование раздела /темы дисциплины	**Виды учебной работы** в часах				
		Очная форма обучения				
		Лек	**Пр**	**Лаб**	**Внеауд**	**СРО**
1.	Введение в программирование на языке Python	4	4	4		14
2.	Библиотеки анализа данных, визуализации и вычислений для Python	4	4	4		16
3.	Загрузка и анализ данных	4	4	4		14
4.	Введение в программирование на языке R	4	4	4		16
5.	Работа с пакетами расширений: ggplot2, dplyr, dbplyr	4	4	4		16
6.	Автоматическая генерация отчетов в R. Пакет knitr	4	4	4		14
7.	Архитектура Apache Spark	4	4	4		14
8.	Доступ к данным Apache Spark с помощью библиотек PySpark и sparklyr	4	4	4		16
	Всего:	**32**	**32**	**32**		**120**

Прим.: Лек – лекции, Пр – практические занятия / семинары, Лаб – лабораторные занятия, Внеауд – внеаудиторная работа, СРО – самостоятельная работа обучающихся

4.2. Содержание дисциплины, структурированное по разделам (темам)

Лекционный курс

№	Наименование раздела /темы дисциплины	Содержание
1.	Введение в программирование на языке Python	Обзор языка Python, базовые типы данных, преобразования типов. Последовательности: списки, диапазоны, кортежи. Циклы, условные операторы. Строки и словари. Функции. Лямбда-выражения.
2.	Библиотеки анализа данных, визуализации и вычислений для Python	Обзор возможностей библиотек numpy, scipy. Массивы данных. Визуализация в matplotlib.
3.	Загрузка и анализ данных	Библиотека pandas. Таблицы DataFrame. Чтение табличных данных разных форматов. Визуализация в seaborn.
4.	Введение в программирование на языке R	Обзор языка R и среды Rstudio. Типы данных, векторы, индексация. Векторные вычисления. Таблицы данных. Базовая графика, реализация статистических методов в R.
5.	Работа с пакетами расширений: ggplot2, dplyr, dbplyr	Установка и подключение пакетов расширения. Аналитическая графика в ggplot2. Манипуляции с табличными данными в dplyr, конвейеры обработки. Доступ к БД из R, отложенные запросы.
6.	Автоматическая генерация отчетов в R. Пакет knitr	Шаблоны отчетов в формате LaTeX и Rmarkdown. Вставка кода на R в отчет. Генерация отчетов в разных форматах.
7.	Архитектура Apache Spark	Обзор архитектуры и возможностей Apache Spark.
8.	Доступ к данным Apache Spark с помощью библиотек PySpark и sparklyr	Доступ к Apache Spark из кода на языке Python: библиотека PySpark. Пакет sparklyr для доступа к Spark из R.

Практические/семинарские занятия

№	Наименование раздела /темы дисциплины	Содержание
1.	Введение в программирование на языке Python	Объектно-ориентированные возможности Python. Функциональные возможности. Изменяемые и неизменяемые последовательности. Генераторы списков, преобразования последовательностей.

2.	Библиотеки анализа данных, визуализации и вычислений для Python	Вычисления с помощью numpy, scipy. Примеры визуализации с помощью matplotlib
3.	Загрузка и анализ данных	Возможности библиотеки pandas. Табличные данные, статистические характеристики, визуализация.
4.	Введение в программирование на языке R	Загрузка табличных данных в R. Визуализация с помощью стандартной графики.
5.	Работа с пакетами расширений: ggplot2, dplyr, dbplyr	Связь визуализации с табличными данными в пакете ggplot2. Преобразование табличных данных в пакете dplyr
6.	Автоматическая генерация отчетов в R. Пакет knitr	Создание шаблонов отчетов. Вставка кода на R в отчеты, обработка и генерация отчета.
7.	Архитектура Apache Spark	Развертывание и запуск Apache Spark. Доступ из Rstudio.
8.	Доступ к данным Apache Spark с помощью библиотек PySpark и sparklyr	Возможности PySpark и sparklyr.

Лабораторные занятия

№	**Наименование раздела /темы дисциплины**	**Название лабораторной работы**
1.	Введение в программирование на языке Python	*Лабораторная работа №1:* Обработка больших объемов текстовой информации с применением регулярных выражений.
2.	Библиотеки анализа данных, визуализации и вычислений для Python	*Лабораторная работа №2:* Разработать скрипт на Python для решения задачи (генерация случайной матрицы, решение задач линейной алгебры).
3.	Загрузка и анализ данных	*Лабораторная работа №3:* Загрузить табличные данные с помощью pandas, используя заданный URL. Провести разведочный анализ и визуализацию данных
4.	Введение в программирование на языке R	*Лабораторная работа №4:* Загрузить табличные данные в R, используя заданный URL. Провести разведочный анализ и

		визуализацию данных
5.	Работа с пакетами расширений: ggplot2, dplyr, dbplyr	*Лабораторная работа №5:* Для заданных табличных данных провести преобразование и выборку данных с помощью dplyr (или dbplyr), отобразить заданный тип графика в ggplot2
6.	Автоматическая генерация отчетов в R. Пакет knitr	*Лабораторная работа №6:* Для заданных табличных данных создать шаблон отчета по их анализу, сгенерировать отчет в заданном формате.
7.	Архитектура Apache Spark	*Лабораторная работа №7:* Загрузить данные в Apacht Spark
8.	Доступ к данным Apache Spark с помощью библиотек PySpark и sparklyr	*Лабораторная работа №8:* Провести анализ данных, загруженных в Spark, с помощью выбранного пакета.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

В качестве учебно-методических материалов используется рекомендованная литература и рекомендованные ресурсы сети Интернет (разделы 7 и 8).

6. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

6.1. Паспорт фонда оценочных средств по дисциплине

№ п/п	Контролируемые разделы (темы) дисциплины (результаты по разделам)	Код контролируемой компетенции (или её части) / и ее формулировка	Наименование оценочного средства
1-3.	1. Введение в программирование на языке Python 2. Библиотеки анализа данных, визуализации и вычислений для Python 3. Загрузка и анализ данных	ОПК-2 (знать, уметь, владеть) ОПК-5 (знать, уметь, владеть)	**Лабораторные работы №1 - 3** (демонстрация на компьютере выполненного проекта и защита работы в форме собеседования с преподавателем); **Контрольная работа №1** (в форме письменных ответов и устного собеседования на теоретические вопросы); **Экзамен** (в форме письменных ответов и устного собеседования на

			теоретические вопросы)
4-6.	4. Введение в программирование на языке R 5. Работа с пакетами расширений: ggplot2, dplyr, dbplyr 6. Автоматическая генерация отчетов в R. Пакет knitr	ОПК-2 (знать, уметь, владеть) ОПК-8 (знать, уметь, владеть) ПК-4 (знать, уметь, владеть)	**Лабораторные работы №4 - 6** (демонстрация на компьютере выполненного проекта и защита работы в форме собеседования с преподавателем) **Контрольная работа №2** (в форме письменных ответов и устного собеседования на теоретические вопросы); **Экзамен** (в форме письменных ответов и устного собеседования на
7-8.	7. Архитектура Apache Spark 8. Доступ к данным Apache Spark с помощью библиотек PySpark и sparklyr	ОПК-2 (знать, уметь, владеть) ОПК-5 (знать, уметь, владеть)	**Лабораторные работы №7 - 8** (демонстрация на компьютере выполненного проекта и защита работы в форме собеседования с преподавателем) **Экзамен** (в форме письменных ответов и устного собеседования на

6.2. *Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующие этапы формирования компетенций в процессе освоения образовательной программы*

6.2.1. Экзамен

Экзамен проводится в виде письменных ответов на 2 вопроса, с последующим устным собеседованием. Критерий оценки – правильность и полнота ответа на вопросы.
Оценка выставляется в баллах от 0 до 40 в равных долях за каждый вопрос.
Экзамен считается сданным при оценке не ниже 60% от максимального балла.

Список билетов на экзамен:

Экзаменационный билет № 1

1. Архитектура Apache Spark
2. Типы данных R. Векторы, таблицы данных (data.frame). Индексация.
3. numpy arrays: создание, свойства.

Экзаменационный билет № 2

1. Структура библиотеки машинного обучения Apache Spark
2. Вычисление параметров линейной регрессии с помощью функции lm()
3. numpy: генераторы псевдослучайных чисел.

Экзаменационный билет № 3

1. Применение пакета sparklyr для доступа к Spark из R
2. Функции вычисления числовых характеристик выборки в R
3. scipy: задачи линейной алгебры.

Экзаменационный билет № 4

1. Использование PySpark для доступа к Apache Spark
2. Списки в R. Функции split(), sapply(), lapply()
3. scipy: статистические методы.

Экзаменационный билет № 5

1. Архитектура Apache Spark
2. Пакет dplyr. Фильтрация данных, изменение структуры таблиц.
3. pandas Series. Создание, свойства

Экзаменационный билет № 6

1. Структура библиотеки машинного обучения Apache Spark
2. Пакет dplyr. Отбор столбцов, группировка, агрегирование.
3. pandas DataFrame: создание, свойства

Экзаменационный билет № 7

1. Применение пакета sparklyr для доступа к Spark из R
2. Визуализация с помощью пакета ggplot2. Слои geom_point и geom_line
3. pandas DataFrame: индексация, выборки

Экзаменационный билет № 8

1. Использование PySpark для доступа к Apache Spark
2. Визуализация с помощью пакета ggplot2. Слои geom_histogram и geom_boxplot
3. pandas DataFrame: описательная статистика

Экзаменационный билет № 9

1. Архитектура Apache Spark
2. Алгоритм K-средних. Функция kmeans().
3. pandas DataFrame: метод aggregate()

6.2.2. Контрольная работа №1

Контрольная работа предназначена для выявления качества усвоения теоретических знаний по темам:
- Введение в программирование на языке Python
- Библиотеки анализа данных, визуализации и вычислений для Python
- Загрузка и анализ данных

Контрольная работа включает в себя 2 вопроса, на которые студент должен дать исчерпывающий устный ответ. Контрольная работа оценивается в баллах от 0 до 10 и считается сданной при оценке не ниже 60% от максимального балла.

Варианты заданий составляются из двух вопросов: первый вопрос из 1-5, второй вопрос из 6-13.

Вопросы контрольной работы №1:
1. Преобразования типов в Python
2. Списки и кортежи в Python
3. Словари и множества в Python
4. Функции и лямбда-выражения
5. Генераторы списков
6. Массивы numpy
7. Линейная алгебра в numpy.linalg
8. Символьные вычисления в sympy
9. Свойства pandas Series
10. Свойства pandas DataFrame
11. Визуализация в matplotlib
12. Визуализация в seaborn
13. Структура библиотеки scipy

Контрольная работа №2

Контрольная работа предназначена для выявления качества усвоения теоретических знаний по темам:
- Введение в программирование на языке R
- Работа с пакетами расширений: ggplot2, dplyr, dbplyr
- Автоматическая генерация отчетов в R. Пакет knitr

Контрольная работа включает в себя 2 вопроса, на которые студент должен дать исчерпывающий устный ответ. Контрольная работа оценивается в баллах от 0 до 10 и считается сданной при оценке не ниже 60% от максимального балла.

Варианты заданий составляются из двух вопросов: первый вопрос из 1-6, второй вопрос из 7-14.

Вопросы контрольной работы №2:

1. Типы данных в R
2. Векторы. Создание, индексация
3. Таблицы data.frame, подмножества данных
4. Списки в R. Функции split, sapply, lapply
5. Базовая графика R: plot, hist, boxplot
6. Вспомогательные графические функции: abline, lines, points, grid
7. Пакеты расширения. Функции require, library
8. Особенности работы с пакетом ggplot2
9. Преобразования таблиц с помощью dplyr: отбор и фильтрация данных
10. Преобразования таблиц с помощью dplyr: агрегирование данных
11. Пакет knitr. Отчеты в формате Rnw (LaTeX + R)
12. Пакет knitr. Отчеты в формате Rmd (rmarkdown)
13. Доступ к БД из R: DBI
14. Доступ к БД из R: dbplyr. Отложенные запросы

6.2.3. Лабораторные работы №1 - 8

Лабораторные работы предназначены для выработки практических навыков по материалу, полученному в рамках предмета (курс лекций), а также выявления качества усвоения знаний по дисциплине.

По завершению каждой из лабораторных работ студент должен продемонстрировать ее результат на компьютере и защитить в форме собеседования с преподавателем. На собеседование выносятся вопросы, касающиеся теоретических аспектов выполняемой работы, последовательности используемых для решения задачи шагов/процедур, а также анализа полученных результатов.

Критерий оценки – полнота, качество, своевременность выполненной работы и успешная ее защита. Лабораторные работы №1 и №2 оцениваются в баллах от 0 до 10, а лабораторные работы №3 и №4 от 0 до 15. Каждая лабораторная работа считается сданной при получении оценки не ниже 60% от максимального балла.

6.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Рейтинговая оценка знаний является интегральным показателем качества теоретических и практических знаний и навыков студентов по дисциплине и складывается из оценок, полученных в ходе текущего контроля и промежуточной аттестации.

Текущий контроль в семестре проводится с целью обеспечения своевременной обратной связи, для коррекции обучения, активизации самостоятельной работы студентов.

Промежуточная аттестация предназначена для объективного подтверждения и оценивания достигнутых результатов обучения после завершения изучения дисциплины.

Текущий контроль осуществляется два раза в семестр: контрольная точка № 1 (КТ № 1) и контрольная точка № 2 (КТ № 2).

Результаты текущего контроля и промежуточной аттестации подводятся по шкале балльно-рейтинговой системы.

Вид контроля	**Этап рейтинговой системы Оценочное средство**	**Балл**	
		Минимум	Максимум
Текущий	**Контрольная точка № 1**	**18**	**30**
	Лабораторная работа №1	6	10
	Лабораторная работа №2	6	10
	Контрольная работа №1 (2 вопроса – 5 и 5 баллов)	6	10
	Контрольная точка № 2	**18**	**30**
	Лабораторная работа №3	9	15
	Лабораторная работа №4	9	15
Промежуточный	Зачет	24	40
Текущий	**Контрольная точка № 1**	**18**	**30**
	Лабораторная работа №5	6	10
	Лабораторная работа №6	6	10
	Контрольная работа №2 (2 вопроса – 5 и 5 баллов)	6	10
	Контрольная точка № 2	**18**	**30**
	Лабораторная работа №7	9	15
	Лабораторная работа №8	9	15
Промежуточный	**Экзамен**	**24**	**40**
ИТОГО по дисциплине		60	100

За несвоевременную сдачу любого из указанных в таблице оценочных средств оценка может быть снижена от 1 до 2 баллов.

Процедура оценивания знаний, умений, владений по дисциплине включает учет успешности по всем видам заявленных оценочных средств.

Устный опрос проводится на каждом практическом занятии и затрагивает как тематику прошедшего занятия, так и лекционный материал. Ответ оценивается преподавателем.

По окончании освоения дисциплины проводится промежуточная аттестация в виде экзамена, что позволяет оценить совокупность приобретенных в процессе обучения компетенций. При выставлении итоговой оценки применяется балльно-рейтинговая система оценки результатов обучения.

Экзамен предназначен для оценки работы обучающегося в течение всего срока изучения дисциплины и призван выявить уровень и систематичность полученных обучающимся теоретических знаний, приобретенных навыков самостоятельной работы.

Оценка сформированных компетенций на экзамене для тех обучающихся, которые пропускали занятия и не участвовали в проверке компетенций во время изучения дисциплины, проводится после индивидуального собеседования с преподавателем по пропущенным или не усвоенным обучающимся темам с последующей оценкой самостоятельно усвоенных знаний на экзамене.

7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

1. Груздев А. В. , Хейдт М. Изучаем Pandas. Издательство "ДМК Пресс" 700 с. 2019 г.
2. Копырин А. С., Салова Т. Л. Программирование на Python: учебное пособие. Издательство "ФЛИНТА" 48 с. 2021 г.
3. Holden Karau, Rachel Warren. High Performance Spark. 2017. 175 p.
4. Matei Zaharia, Holden Karau, Andy Konwinski, Patrick Wendell. Learning Spark, Lightning-Fast Big Data Analysis. 2015. 276 p.
5. Petar Zecevic. Spark in Action. 2016. 468 p.
6. Mike Frampton. Mastering Apache Spark. 2015. 318 p.

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» (далее - сеть «Интернет»), необходимых для освоения дисциплины

1. Язык программирования Python [Официальный сайт]. — https://www.python.org/
2. Язык программирования R [Официальный сайт]. — http://r-project.org/
3. Apache Spark [Официальный сайт]. — https://spark.apache.org/
4. http://pandas.pydata.org Документация по библиотеке pandas
5. https://seaborn.pydata.org Документация по визуализации в seaborn
6. https://www.sympy.org Символьные вычисления в Python

9. Методические указания для обучающихся по освоению дисциплины

Вид учебного занятия	Организация деятельности студента
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; помечать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удается разобраться в материале,

	необходимо сформулировать вопрос и задать преподавателю на лекции или лабораторной работе. Уделить внимание следующим базовым понятиям: большие данные, масштабирование, распределенная система, целостность данных, репликация данных, шардинг данных, CAP теорема.
Контрольная работа	Работа с конспектами лекций, знакомство с основной и дополнительной литературой, включая справочные издания, зарубежные источники.
ораторная работа	При выполнении лабораторных работ необходимо ориентироваться на конспекты лекций, рекомендуемую литературу. Лабораторная работа считается выполненной после ее успешной защиты, включающей: – демонстрацию на компьютере решаемой задачи с разъяснением разработанного программного кода и демонстрацией выполнения; – собеседование с преподавателем для выявления уровня освоения теоретических основ в области больших данных.
готовка к экзамену	При подготовке к экзамену необходимо ориентироваться на конспекты лекций и лабораторные работы, а также рекомендуемую литературу.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости)

- Операционные системы Windows 7/10, Linux (CentOS / RedHat, OpenSUSE, Ubuntu);
- Среда для программирования на языке Python – Jupyter;
- Среда для программирования на языке R – Rstudio;
- Электронные презентации лекций в формате PDF, демонстрируемые с использованием мультимедийного проектора или дистанционно.

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

• Компьютерный класс сетевых технологий. Класс оснащен 10 компьютерами (Intel Core i5/8GB/1 TB) и 1 компьютером (Intel Celeron 1.6 GHz, 2 GB RAM, 250 GB) с операционной системой Linux, а также мультимедийным проектором. Есть доступ к Wi-Fi.

• Аудиторный класс, оборудованный проекционным экраном, мультимедийным проектором и персональным компьютером (AMD, ATHLON64, 2.7 GHz, 4 GB RAM, 250 GB). Есть доступ к Wi-Fi.

12. Иные сведения и (или) материалы

12.1. Перечень образовательных технологий, используемых при осуществлении образовательного процесса по дисциплине

Лекционные и практические занятия проходят с обсуждением учебного материала, демонстрируемого в форме презентаций на экране с использованием мультимедиа-проектора. Все лабораторные занятия проводятся в интерактивной форме при тесном контакте студентов с преподавателем.

В рамках лабораторных работ студенты выполняют 4 лабораторные работы, призванные дать представление о возможностях применения больших данных, как инструментария для решения самых разнообразных практических задач. Лабораторные работы проводятся при активном взаимодействии студентов и преподавателя, в ходе которого обсуждаются детали создания проекта задачи, проблемы и ошибки, возникающие на всех этапах их разработки, проводится проверка корректности полученных результатов.

12.2. Формы организации самостоятельной работы обучающихся (темы, выносимые для самостоятельного изучения; вопросы для самоконтроля; типовые задания для самопроверки

На самостоятельное изучение студентам предлагается более глубоко рассмотреть темы, кратко затрагиваемые в лекционных курсах. Контроль освоения материала осуществляется в ходе приема лабораторных работ и в рамках экзамена по дисциплине.

№	Тема	Часть, осваиваемая самостоятельно
1.	Введение в программирование на языке Python	Объектно-ориентированное программирование на Python.
2.	Библиотеки анализа данных, визуализации и вычислений для Python	Библиотека scipy
3.	Загрузка и анализ данных	Индексация в dataFrame
4.	Введение в программирование на языке R	Списки в R. Семейство функций apply

№	Тема	Часть, осваиваемая самостоятельно
5.	Работа с пакетами расширений: ggplot2, dplyr, dbplyr	Пакет DBI
6.	Автоматическая генерация отчетов в R. Пакет knitr	Формат TeX
7.	Архитектура Apache Spark	Реализация ML в Spark
8.	Доступ к данным Apache Spark с помощью библиотек PySpark и sparklyr	Отложенные вычисления

Контроль освоения самостоятельно изученного теоретического материала осуществляется в виде собеседования во время защиты лабораторных, в виде устного опроса на экзамене.

Кроме этого, студенты также самостоятельно выполняют большую часть предусмотренных практических работ, промежуточный результат которых представляется на лабораторных занятиях, а конечный результат - на защите лабораторных работ.

Вопросы для самоконтроля:

- Организация Spark кластера.
- Издательская система TeX/ LaTeX
- Язык Python. Последовательности
- Язык R. Списки