

Министерство науки и высшего образования Российской Федерации

федеральное государственное бюджетное образовательное учреждение
высшего образования
РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ
УНИВЕРСИТЕТ

Кафедра Прикладной информатики

Рабочая программа дисциплины

Введение в Big Data

Основная профессиональная образовательная программа
высшего образования по направлению подготовки

38.03.05 «Бизнес-информатика»

Направленность (профиль):
Бизнес-аналитика

Уровень:
Бакалавриат

Форма обучения
Очная/заочная/очно-заочная

Согласовано
Руководитель ОПОП

Степанов С.Ю.

Председатель УМС
Палкин И.И.

Рекомендована решением
Учебно-методического совета РГГМУ
24 июня 2021 г., протокол № 9

Рассмотрена и утверждена на заседании кафедры
11 мая 2021 г., протокол № 6

Зав. кафедрой Истомин Е.П.

Авторы-разработчики:

Истомин Е.П.
Степанов С.Ю.

Санкт-Петербург 2021

**Рассмотрено и рекомендовано к использованию в учебном процессе
на 2022/2023 учебный год без изменений***

**Протокол № 2 заседания кафедры Прикладной информатики от
17.03.2022г.**

1. Цель и задачи освоения дисциплины

Цель освоения дисциплины – дать углубленное представление о технологиях Big Data и особенностях и перспективах их использования на практике

Задачи:

- ознакомить студентов с основными концепциями и методами анализа данных
- научить применять полученные знания использования Big Data на практике
- изучение основных структур и форм хранения данных

2. Место дисциплины в структуре основной профессиональной образовательной программы

Дисциплина «Введение в Big Data» относится к обязательной части блока «Дисциплины».

В соответствии с учебным планом, обучение проводится в 6 семестре для очной формы обучения, в 7 семестре для очно-заочной формы обучения и на 4 году обучения для заочной формы. Освоения дисциплины проводится на основе знаний, полученных во время изучения дисциплины “Информационные и системы” и “Алгоритмизация и программирование” и “Анализ данных”.

3. Перечень планируемых результатов обучения

Процесс изучения дисциплины направлен на формирование компетенций:

ОПК-4.1

Общепрофессиональные компетенции

Код и наименование общепрофессиональной компетенции	Код и наименование индикатора достижения общепрофессиональной компетенции	Результаты обучения
ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.	<i>Знать: основные способы хранения данных и извлечения больших данных</i> <i>Уметь: анализировать потоки данных; делать выводы из полученных результатов анализа.</i> <i>Владеть: терминологией Больших данных; методикой хранения и извлечения больших данных.</i>

4. Структура и содержание дисциплины

4.1. Объем дисциплины

Объем дисциплины составляет 3 зачетных единицы, 108 академических часа.

Таблица 4.

Объем дисциплины по видам учебных занятий в академических часах

Объём дисциплины	Всего часов		
	Очная форма обучения	Очно-заочная форма обучения	Заочная форма обучения
Объем дисциплины	108	108	108
Контактная работа обучающихся с преподавателем (по видам аудиторных учебных занятий) – всего:	42	28	12
в том числе:	-	-	-
лекции	14	8	4
занятия семинарского типа:			
практические занятия			
лабораторные занятия	28	20	8
Самостоятельная работа (далее – СРС) – всего:	66	80	96
в том числе:	-	-	-
курсовая работа			
контрольная работа			
Вид промежуточной аттестации	экзамен	экзамен	экзамен

4.2. Структура дисциплины

Таблица 5.

Структура дисциплины для очной формы обучения

№	Раздел / тема дисциплины	С е м е с т р	Виды учебной работы, в т.ч. самостоятельная работа студентов, час.			Формы текущего контроля успеваемости	Формируемые компетенции	Индикаторы достижения компетенций
			Ле кц ии	Пра ктич ески е заня тия	СР С			
1	Что такое Big Data: введение в большие данные	6	4	8	18	устный опрос	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информацион-	ОПК-4.1. Использует математические модели и методы обработки данных.

							аналитическо й поддержки принятия управленческ их решений.	
2	Основы технологий больших данных	6	6	8	24	коллоквиум	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информацион но- аналитическо й поддержки принятия управленческ их решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
3	Практическое применение технологий больших данных	6	4	12	24	Лабораторная работа	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информацион но- аналитическо й поддержки принятия управленческ их решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
	ИТОГО	-	14	28	66	-	-	-

Таблица 6.

Структура дисциплины для очно-заочной формы обучения

№	Раздел / тема дисциплины	С е м е с т р	Виды учебной работы, в т.ч. самостоятельная работа студентов, час.			Формы текущего контроля успеваемости	Формируемые компетенции	Индикаторы достижения компетенций
			Ле кц ии	Пра ктич ески е заня тия	СР С			
1	Что такое Big Data: введение в большие данные	7	4	4	20	устный опрос	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
2	Основы технологий больших данных	7	2	4	20	коллоквиум	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
3	Практическое применение технологий больших данных	7	2	12	40	Лабораторная работа	ОПК-4. Способен использовать информацию, методы и	ОПК-4.1. Использует математические модели и методы

							программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	обработки данных.
	ИТОГО	-	8	20	80	-	-	-

Таблица 7.

Структура дисциплины для заочной формы обучения

№	Раздел / тема дисциплины	Год	Виды учебной работы, в т.ч. самостоятельная работа студентов, час.			Формы текущего контроля успеваемости	Формируемые компетенции	Индикаторы достижения компетенций
			Лекции	Практические занятия	СРС			
1	Что такое Big Data: введение в большие данные	4	2	2	32	устный опрос	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
2	Основы технологий больших данных	4	1	2	32	коллоквиум	ОПК-4. Способен использовать информацию, методы и программные средства ее	ОПК-4.1. Использует математические модели и методы обработки данных.

							сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	
3	Практическое применение технологий больших данных	4	1	4	32	Лабораторная работа	ОПК-4. Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений.	ОПК-4.1. Использует математические модели и методы обработки данных.
	ИТОГО	-	4	8	96	-	-	-

4.3. Содержание разделов/тем дисциплины

Что такое Big Data: введение в большие данные

Техника анализа большого количества прибывающей информации. Специфика больших данных. Традиционные методы работы с массивами информации. Программные инструменты для извлечения данных из внешних источников и обработки огромных массивов. Набор признаков. Источники больших данных. Методы анализа.

Основы технологий больших данных

Методы и техники анализа. NoSQL. MapReduce. Hadoop. R. Аппаратные решения.

Практическое применение технологий больших данных

Теория и практика больших данных в отраслях. Big Data в промышленности. Big Data в логистике. Big Data в ритейле. Big Data в финансах. Big Data в HR. Big Data в образовании. Big Data в маркетинге. Перспективы Big Data.

4.4. Содержание занятий семинарского типа

Таблица 11.

Содержание лабораторных занятий для очной формы обучения

№ темы дисциплины	Тематика лабораторных занятий	Всего часов
Что такое Big Data: введение в большие данные	“Введение в большие данные”	8
Основы технологий больших данных	“Технологии больших данных”	8
Практическое применение технологий больших данных	“Обнаружение значимых корреляций” “Определение отношений между разнородными данными”	12

Таблица 12.

Содержание лабораторных занятий для очно-заочной формы обучения

№ темы дисциплины	Тематика лабораторных занятий	Всего часов
Что такое Big Data: введение в большие данные	“Введение в большие данные”	4
Основы технологий больших данных	“Технологии больших данных”	4
Практическое применение технологий больших данных	“Обнаружение значимых корреляций” “Определение отношений между разнородными данными”	12

Таблица 13.

Содержание лабораторных занятий для заочной формы обучения

№ темы дисциплины	Тематика лабораторных занятий	Всего часов
Что такое Big Data: введение в большие данные	“Введение в большие данные”	2
Основы технологий больших данных	“Технологии больших данных”	2
Практическое применение технологий	“Определение отношений между разнородными данными”	4

больших данных		
----------------	--	--

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

Moodle

6. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации по итогам освоения дисциплины

6.1. Текущий контроль

Типовые задания, методика выполнения и критерии оценивания текущего контроля по разделам дисциплины представлены в Фонде оценочных средств по данной дисциплине.

6.2. Промежуточная аттестация

Форма промежуточной аттестации по дисциплине – экзамен (6 семестр) для очной формы обучения, экзамен (7 семестр) для очно-заочной формы обучения, экзамен (4 год обучения) для заочной формы обучения

Форма проведения экзамена: *письменно по билетам*

Перечень вопросов для подготовки к экзамену:

ОПК-4.1

1. Понятие “Большие данные”
2. Виды массивов данных
3. Базовые принципы обработки больших данных
4. Технологии обработки больших данных: NoSQL
5. Технологии обработки больших данных: MapReduce
6. Технологии обработки больших данных: Hadoop
7. Технологии обработки больших данных: R
8. Реляционные системы управления базами данных
9. Прогнозирование. Виды прогнозов
10. Определение больших данных, ключевые характеристики.
11. Основные виды данных
12. Парадигма MapReduce. Принцип работы. Сильные и слабые стороны. Области применения
13. “Жизненный цикл” проекта по аналитике больших данных
14. Научные проблемы больших данных
15. Общие методы анализа социальных и медиа процессов

Экзамен оценивается по четырехбалльной шкале: «отлично» / «хорошо» / «удовлетворительно» / «неудовлетворительно».

Критерии оценивания:

«Отлично» - студент дал ответ на два вопроса, допускаются недочеты

«Хорошо» - студент дал ответ на два вопроса, допустил не грубые ошибки

«Удовлетворительно» - студент дал ответ на один из вопросов, или допустил грубые ошибки в одном из ответов

«Неудовлетворительно» - студент не дал ответ на вопросы или допустил грубые ошибки в обоих вопросах

7. Методические рекомендации для обучающихся по освоению дисциплины

Методические рекомендации ко всем видам аудиторных занятий, а также методические рекомендации по организации самостоятельной работы, в том числе по подготовке к текущему контролю и промежуточной аттестации представлены в Методических рекомендациях для обучающихся по освоению дисциплины «Введение в Big Data».

8. Учебно-методическое и информационное обеспечение дисциплины

8.1. Перечень основной и дополнительной учебной литературы

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.

Дополнительная литература

1. Радченко, И. А. Технологии и инфраструктура Big Data : учебное пособие / И. А. Радченко, И. Н. Николаев. — Санкт-Петербург : НИУ ИТМО, 2018. — 52 с.

2. MySQL 8 для больших данных / Ш. Чаллавала, Д. Лакхатария, Ч. Мехта, К. Пател ; перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2018. — 226 с.

3. Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с.

8.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

1. <https://ru.pdfdrive.com/big-data-e171375954.html>

8.3. Перечень программного обеспечения

1. Microsoft Office

8.4. Перечень информационных справочных систем

1. Elib RSHU

2. ЭБС "Лань"

8.5. Перечень профессиональных баз данных

1. Электронно-библиотечная система elibrary;

2. data.gov.ru

3. data.worldbank.org

9. Материально-техническое обеспечение дисциплины

Лекционные занятия

- Лекционная аудитория

Лабораторные занятия

- Аудитория, оснащенная персональными компьютерами или мультимедийным оборудованием (ауд.101,104,108,23 и кв.14 2-го корпуса РГГМУ)

Самостоятельная работа

- Читальный зал библиотеки, читальный зал Российской Национальной Библиотеки

10. Особенности освоения дисциплины для инвалидов и лиц с ограниченными возможностями здоровья

Обучение обучающихся с ограниченными возможностями здоровья при необходимости осуществляется на основе адаптированной рабочей программы с использованием специальных методов обучения и дидактических материалов, составленных с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся (обучающегося).

При определении формы проведения занятий с обучающимся-инвалидом учитываются рекомендации, содержащиеся в индивидуальной программе реабилитации инвалида, относительно рекомендованных условий и видов труда.

При необходимости для обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья создаются специальные рабочие места с учетом нарушенных функций и ограничений жизнедеятельности.