

Минобрнауки России

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Оренбургский государственный университет»

Кафедра прикладной математики

РАБОЧАЯ ПРОГРАММА

ДИСЦИПЛИНЫ

«Б1.Д.В.Э.1.2 Компьютерные технологии обработки больших массивов данных»

Уровень высшего образования

МАГИСТРАТУРА

Направление подготовки

01.04.02 Прикладная математика и информатика
(код и наименование направления подготовки)

Оптимизация и оптимальное управление
(наименование направленности (профиля) образовательной программы)

Квалификация

Магистр

Форма обучения

Очная

Год набора 2019

Рабочая программа рассмотрена и утверждена на заседании кафедры

Кафедра прикладной математики

наименование кафедры

протокол № 6 от "6" февраля 2019 г.

Заведующий кафедрой

Кафедра прикладной математики

наименование кафедры


подпись

И.П. Болодурина

расшифровка подписи

Исполнители:

должность


подпись

Д.И. Парфёнов

расшифровка подписи

должность

подпись


расшифровка подписи

И.П. Болодурина

СОГЛАСОВАНО:

Председатель методической комиссии по направлению подготовки

01.04.02 Прикладная математика и информатика

код наименование

личная подпись



И.П. Болодурина

расшифровка подписи

Научный руководитель магистерской программы



личная подпись

А.Н. Машаков

расшифровка подписи

Заведующий отделом комплектования научной библиотеки

личная подпись



Н.Н. Грицай

расшифровка подписи



Уполномоченный по качеству факультета

личная подпись



М.В. Крюкова

расшифровка подписи

№ регистрации _____

© Парфёнов Д.И., 2019

© Болодурина И.П., 2019

© ОГУ, 2019

1 Цели и задачи освоения дисциплины

Цель (цели) освоения дисциплины:

формирование 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) изучить технологии хранения, обработки и анализа больших 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.

2 Место дисциплины в структуре образовательной программы

Дисциплина относится к дисциплинам (модулям) по выбору вариативной части блока Д «Дисциплины (модули)»

Пререквизиты дисциплины: *Отсутствуют*

Постреквизиты дисциплины: *Отсутствуют*

3 Требования к результатам обучения по дисциплине

Процесс изучения дисциплины направлен на формирование следующих результатов обучения

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1-В-1 Имеет представление о принципах сбора, отбора и обобщения информации для анализа проблемных ситуаций УК-1-В-3 Имеет практический опыт работы с информационными объектами и сетью Интернет, опыт научного поиска, опыт библиографического разыскания, разработки научного исследования	<u>Знать:</u> - принципы обработки больших данных в распределенных вычислительных системах; <u>Уметь:</u> - настраивать распределенные базы данных. <u>Владеть:</u> навыками администрирования распределенных вычислительных систем.
ПК*-2 Способен осуществлять выбор источников информации,	ПК*-2-В-1 Имеет представление о методах, приемах, алгоритмах и способах сбора, обработки и интерпретации данных	<u>Знать:</u> - знать архитектуры вычислительных систем

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
планировать аналитические работы, определять необходимые технические средства для обработки данных	современных научных исследований, необходимых для формирования выводов по соответствующим научным исследованиям ПК*-2-В-2 Осуществляет поиск, отбор и систематизацию информации для определения альтернативных вариантов стратегических решений в проблемной ситуации и обоснования выбора оптимальной стратегии ПК*-2-В-3 Демонстрирует навыки работы с Интернет-ресурсами и иными источниками информации, собирает необходимые для аналитической деятельности данные, определяет технические средства для обработки данных ПК*-2-В-4 Демонстрирует навыки планирования аналитических работ, обработки информации проведенных экспериментов и анализа полученных данных	для обработки больших данных. Уметь: - разрабатывать алгоритмы и методы для обработки больших данных. Владеть: - инструментами анализа данных с использованием современных языков запросов.
ПК*-3 Способен разрабатывать и анализировать концептуальные и теоретические модели решаемых научных проблем и задач	ПК*-3-В-1 Имеет представление об основных известных математических моделях практических задач в области проводимой научно-исследовательской работы ПК*-3-В-3 Разрабатывает методы, алгоритмы, математические модели по тематике проводимого научно-исследовательского направления ПК*-3-В-4 Разрабатывает концептуальную модель прикладной области, выбирает инструментальные средства и технологии проектирования информационных систем; проводит формализацию и реализацию решения прикладных задач	Знать: методы, алгоритмы, математические модели искусственных нейронных сетей применяемых в анализе данных Уметь: разрабатывать концептуальную модель прикладной области, выбирать методы анализа данных, а также адаптировать базовые методы к решению прикладных задач Владеть: навыками формализации и анализа данных

4 Структура и содержание дисциплины

4.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 3 зачетные единицы (108 академических часов).

Вид работы	Трудоемкость, академических часов
------------	-----------------------------------

	1 семестр	всего
Общая трудоёмкость	108	108
Контактная работа:	34,25	34,25
Лекции (Л)	18	18
Практические занятия (ПЗ)	16	16
Промежуточная аттестация (зачет, экзамен)	0,25	0,25
Самостоятельная работа: - выполнение индивидуального творческого задания (ИТЗ); - выполнение расчетно-графического задания (РГЗ); - написание реферата (Р); - написание эссе (Э); - самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий); - подготовка к практическим занятиям; - подготовка к коллоквиумам; - подготовка к рубежному контролю и т.п.)	73,75	73,75
Вид итогового контроля (зачет, экзамен, дифференцированный зачет)	зачет	

Разделы дисциплины, изучаемые в 1 семестре

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
1	Большие данные. Основы систем больших данных	28	4	4		20
2	Архитектура системы обработки больших данных	26	4	4		18
3	Параллельные алгоритмы для работы с данными	26	4	4		18
4	Программные платформы и системы для больших данных	28	6	4		18
	Итого:	108	18	16		74
	Всего:	108	18	16		74

4.2 Содержание разделов дисциплины

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.

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).

Раздел № 3 Параллельные алгоритмы для работы с данными

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

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. Состав и возможности программного комплекса Apache Hadoop. Языки

4.3 Практические занятия (семинары)

№ занятия	№ раздела	Тема	Кол-во часов
1	1	Выбор инструментария для работы с BigData	4
2	2	Визуализация данных	4
3	3	Разработка запросов на выборку информации на основе map-reduce	4
4	4	Изучение и конфигурирование программного комплекса Apache Hadoop. Размещения набора данных по заданной тематике.	4
		Итого:	16

5 Учебно-методическое обеспечение дисциплины

5.1 Основная литература

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>

5.2 Дополнительная литература

1. Форман, Д. Много цифр. Анализ больших данных при помощи Excel / Форман Д.; Пер. с англ. Соколовой А. - М.: Альпина Пабли., 2016. - 461 с. - Режим доступа: <http://znanium.com/catalog/product/551044>
2. Интеллектуальные системы и нечеткая логика: учебник / В.П. Корячко, М.А. Бакулева, В.И. Орешков. - М.: КУРС, 2017. - 352 с. - Режим доступа: <http://znanium.com/catalog/product/882796>
3. Кухаренко, Б.Г. Интеллектуальные системы и технологии / Б.Г. Кухаренко; Министерство транспорта Российской Федерации, Московская государственная академия водного транспорта. – М.: Альтаир: МГАВТ, 2015. - 115 с. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=429758>

5.3 Периодические издания

- 1 Информационные технологии : журнал // Информационные технологии с ежемесячным приложением. - М. : Агентство "Роспечать".
- 2 Открытые системы. СУБД: журнал. – М.: Агентство "Роспечать".
- 3 Информационные технологии: журнал. - М.: Агентство "Роспечать".
- 4 IT-News/ Новости информационных технологий: газета. - М.: Агентство "Роспечать".

5.4 Интернет-ресурсы

<https://openedu.ru/course/misis/ABD/> - «Открытое образование», Каталог курсов, MOOK: «Введение в инженерию больших данных»;

5.5 Программное обеспечение, профессиональные базы данных и информационные справочные системы

1. Операционная система Microsoft Windows
2. Apache Hadoop - свободно распространяемый набор утилит, библиотек и фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов.
3. Open Office/LibreOffice - свободный офисный пакет программ, включающий в себя текстовый и табличный редакторы, редактор презентаций и другие офисные приложения.
4. Яндекс.Браузер - браузер, созданный компанией «Яндекс» на основе движка (бесплатная версия) Режим доступа: <https://browser.yandex.ru>.
5. Nature Publishing Group [Электронный ресурс] : реферативная база данных. - Режим доступа: <http://www.nature.com/siteindex/index.html>, в локальной сети ОГУ.
6. SCOPUS [Электронный ресурс] : реферативная база данных / компания Elsevier. – Режим доступа: <https://www.scopus.com/>, в локальной сети ОГУ.
7. Большая советская энциклопедия [Электронный ресурс]: универсальная справочная энциклопедия международного уровня. – Режим доступа: <https://bigenc.ru/> в локальной сети ОГУ.

6 Материально-техническое обеспечение дисциплины

Учебные аудитории для проведения занятий лекционного типа, семинарского типа, для проведения групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

Аудитории оснащены комплектами ученической мебели, техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Помещение для самостоятельной работы обучающихся оснащены компьютерной техникой, подключенной к сети "Интернет", и обеспечением доступа в электронную информационно-образовательную среду ОГУ.