

Документ подписан простой электронной подписью
Информация об органе
ФИО: Клочков Юрий Сергеевич
Должность: и.о. ректора
Дата подписания: 10.04.2024 14:34:37
Уникальный программный ключ:
4e7c4ea90328ec8e65c5d8058549a2538d7400d1

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

_____ 2023г.
«_____»_____

РАБОЧАЯ ПРОГРАММА

дисциплины:	Методы обработки больших данных
направление подготовки:	45.03.04 Интеллектуальные системы в гуманитарной сфере
направленность (профиль):	Разработка и программирование интеллектуальных систем
форма обучения:	очная

Рабочая программа рассмотрена
на заседании кафедры бизнес – информатики и математики

1. Цели и задачи освоения дисциплины

Цель дисциплины заключается в ознакомлении обучающихся с существующими методами и системами обработки больших данных, их областями применимости, преимуществами и недостатками.

Задачи дисциплины:

- знакомство с основными понятиями и алгоритмами обработки больших данных, особенностями адаптации алгоритмов на графах и алгоритмов машинного обучения к возможностям обработки больших объемов данных;
- формирование умений преобразовывать задачу обработки больших объемов данных к задаче статической, пакетной (batch) или потоковой (streaming) обработки, применять алгоритмы машинного обучения и алгоритмы на графах для анализа больших данных;
- развитие умений вычислять необходимые ресурсы для сборки вычислительного кластера, запускать задачи на вычислительном кластере и отлаживать программы в локальной файловой системе;
- формирование навыков культурной постановки, анализа и решения задач, возникающих при работе с большими данными;
- развитие у обучающихся творческого и интеллектуального потенциала.

2. Место дисциплины в структуре ОПОП ВО

Дисциплина относится к блоку факультативных дисциплин.

Необходимыми условиями для освоения дисциплины являются:

- знание основных структур данных, технологий и методов программирования; основных понятий параллельных и распределенных вычислений;
- умение разрабатывать алгоритмы решения задач и записывать их на языке программирования;
- владение методами машинного обучения, навыками использования компьютерных технологий и средств обработки информации.

3. Результаты обучения по дисциплине

Процесс изучения дисциплины направлен на формирование следующих компетенций:

Таблица 3.1

Код и наименование компетенции	Код и наименование индикатора достижения компетенции (ИДК)	Код и наименование результата обучения по дисциплине
ОПК-1. Способен применять в профессиональной деятельности методы математического анализа, логики и моделирования, теоретического и экспериментального исследования в информатике, лингвистике и гуманитарных науках	ОПК-1.2 Владеет навыками теоретического и экспериментального исследования объектов профессиональной деятельности	Знать: (З1) математические модели обработки больших данных
		Уметь: (У1) применять математические модели и алгоритмы для обработки больших данных
		Владеть: (В1) практическими навыками исследования в области применения методов, алгоритмов и технологий обработки больших данных
ОПК-2. Способен к профессиональному росту и совершенствованию в области гуманитарных, социальных и лингвистических наук, а также в сфере техники и технологии	ОПК-2.2 Использует аппаратно-программные средства, информационно-коммуникационные технологии получения знаний в профессиональной	Знать: (З2) архитектуру распределенных файловых систем
		Уметь: (У2) находить информацию по большим данным и методам их обработки с использованием современных справочных и

информатики	деятельности, в области гуманитарных, социальных и лингвистических наук	библиотечных системам и системам дистанционного образования
		Владеть: (В2) практическими навыками поиска информации по большим данным и методам их обработки с использованием современных справочных и библиотечных системам и системам дистанционного образования
ОПК-5. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-5.2 Понимает принципы работы современных информационно-коммуникационных и интеллектуальных технологии, ориентируясь на задачи профессиональной деятельности	Знать: (З3) основные требования и стандарты в области описания и документирования методов обработки больших данных
		Уметь: (У3) оценивать функциональные возможности систем и технологий для обработки больших данных на основании технической документации
		Владеть: (В3) практическим опытом изучения систем обработки больших данных

4. Объем дисциплины

Общий объем дисциплины составляет 1 зачетную единицу, 36 часов.

Таблица 4.1

Форма обучения	Курс/ семестр	Аудиторные занятия / контактная работа, час.			Самостоятельная работа, час.	Контроль, час.	Форма промежуточной аттестации
		Лекции	Практические занятия	Лабораторные занятия			
Очная	4/7	-	-	14	22	-	Зачет

5. Структура и содержание дисциплины

5.1. Структура дисциплины.

очная форма обучения (ОФО)

Таблица 5.1.1

№ п/п	Структура дисциплины		Аудиторные занятия, час.			СРС, час.	Всего, час.	Код ИДК	Оценочные средства
	Номер раздела	Наименование раздела	Л.	Пр.	Лаб.				
1	1	Введение. Распределенные файловые системы	-	-	2	3	5	ОПК-1.2 ОПК-2.2 ОПК-5.2	Контрольная работа №1, Проектное задание
2	2	Модель вычислений MapReduce	-	-	2	3	5	ОПК-1.2 ОПК-2.2 ОПК-5.2	Контрольная работа №1, Проектное задание
3	3	SQL over BigData. Hive	-	-	2	3	5		Контрольная работа №1, Проектное задание
4	4	Beyond MapReduce. Spark	-	-	2	3	5		Контрольная работа №1, Проектное задание

5	5	Машинное обучение на больших данных	-	-	2	3	5		Контрольная работа №2, Проектное задание
6	6	Потоковая обработка данных	-	-	2	3	5	ОПК-1.2 ОПК-2.2 ОПК-5.2	Контрольная работа №2, Проектное задание
7	7	Key-value хранилища в больших данных	-	-	2	4	6	ОПК-1.2 ОПК-2.2 ОПК-5.2	Контрольная работа №2, Проектное задание
	Зачет		-	-	-	-	-	ОПК-1.2 ОПК-2.2 ОПК-5.2	Вопросы для зачета
Итого:			-	-	14	22	36		

5.2. Содержание дисциплины.

5.2.1. Содержание разделов дисциплины (дидактические единицы).

Раздел 1. «Введение. Распределенные файловые системы». Понятие «большие данные». Постановка задачи обработки и хранения больших данных. Примеры применения больших данных в IT индустрии. Архитектура распределенных файловых систем.

Раздел 2. «Модель вычислений MapReduce». Математическая модель парадигмы MapReduce. Hadoop MapReduce. Расширения модели. Часто применяемые техники в обработке данных.

Раздел 3. «SQL over BigData. Hive». Hive: мотивация, языковая модель. Hive: расширенные возможности.

Раздел 4. «Beyond MapReduce. Spark». Недостатки MapReduce. Spark. Spark SQL. Обработка графов при помощи Spark. Оптимизация Spark.

Раздел 5. «Машинное обучение на больших данных». Алгоритмы для работы с большими данными. API для обучения алгоритмов на больших данных.

Раздел 6. «Потоковая обработка данных». Обработка больших данных в режиме реального времени. Модель парадигмы Kafka.

Раздел 7. «Key-value хранилища в больших данных». HBase. Cassandra.

5.2.2. Содержание дисциплины/модуля по видам учебных занятий.

Лекционные занятия

Лекционные занятия учебным планом не предусмотрены.

Практические занятия

Практические занятия учебным планом не предусмотрены.

Лабораторные работы

Таблица 5.2.1

№ п/п	Номер раздела дисциплины	Объем, час.			Тема лабораторной работы
		ОФО	ЗФО	ОЗФО	
1	1	2	-	-	Введение. Распределенные файловые системы
2	2	2	-	-	Модель вычислений MapReduce
3	3	2	-	-	SQL over BigData. Hive
4	4	2	-	-	Beyond MapReduce. Spark
5	5	2	-	-	Машинное обучение на больших данных
6	6	2	-	-	Потоковая обработка данных
7	7	2	-	-	Key-value хранилища в больших данных
Итого:		14	-	-	

Самостоятельная работа студента

Таблица 5.2.2

№ п/п	Номер раздела дисциплины	Объем, час.			Тема	Вид СРС
		ОФО	ЗФО	ОЗФО		
1	1	3	-	-	Введение. Распределенные файловые системы	Изучение рекомендуемой литературы, подготовка к лабораторной работе, к контрольной работе
2	2	3	-	-	Модель вычислений MapReduce	Изучение рекомендуемой литературы, подготовка к лабораторной работе, к контрольной работе
3	3	3	-	-	SQL over BigData. Hive	Изучение рекомендуемой литературы, подготовка к лабораторной работе, к контрольной работе
4	4	3	-	-	Beyond MapReduce. Spark	Изучение рекомендуемой литературы, подготовка к лабораторной работе, к контрольной работе
5	5	3	-	-	Машинное обучение на больших данных	Изучение рекомендуемой литературы, подготовка к лабораторной работе, к контрольной работе
6	6	3	--	-	Потоковая обработка данных	Изучение рекомендуемой литературы, подготовка к лабораторной работе, к контрольной работе
7	7	4	-	-	Key-value хранилища в больших данных	Изучение рекомендуемой литературы, подготовка к лабораторной работе, к контрольной работе
Итого:		22	-	-		

5.2.3. Преподавание дисциплины ведется с применением следующих видов образовательных технологий:

- визуализация учебного материала в PowerPoint в диалоговом режиме (лекционные занятия);
- работа в малых группах (лабораторные работы);
- индивидуальная работа (лабораторные работы).

6. Тематика курсовых работ/проектов

Курсовые работы/проекты учебным планом не предусмотрены.

7. Контрольные работы

Контрольные работы учебным планом не предусмотрены.

8. Оценка результатов освоения дисциплины

8.1. Критерии оценивания степени полноты и качества освоения компетенций в соответствии с планируемыми результатами обучения приведены в Приложении 1.

8.2. Рейтинговая система оценивания степени полноты и качества освоения компетенций обучающихся очной формы обучения представлена в таблице 8.1.

Таблица 8.1

№ п/п	Виды мероприятий в рамках текущего контроля	Количество баллов
1 текущая аттестация		
1	Контрольная работа №1	25
	ИТОГО за первую текущую аттестацию	25
2 текущая аттестация		
2	Контрольная работа №2	15
3	Защита проектного задания	60
	ИТОГО за вторую текущую аттестацию	75
	ВСЕГО	100

9. Учебно-методическое и информационное обеспечение дисциплины/модуля

9.1. Перечень рекомендуемой литературы представлен в Приложении 2.

9.2. Современные профессиональные базы данных и информационные справочные системы:

- Электронный каталог/Электронная библиотека ТИУ <http://webirbis.tsogu.ru/>
- Цифровой образовательный ресурс – библиотечная система IPR SMART — <https://www.iprbookshop.ru/>
- Электронно-библиотечная система «Консультант студента» www.studentlibrary.ru
- Электронно-библиотечная система «Лань» <https://e.lanbook.com>
- Образовательная платформа ЮРАЙТ www.urait.ru
- Научная электронная библиотека ELIBRARY.RU <http://www.elibrary.ru>
- Национальная электронная библиотека (НЭБ)
- Библиотеки нефтяных вузов России :
- Электронная нефтегазовая библиотека РГУ нефти и газа им. Губкина <http://elib.gubkin.ru/>,
- Электронная библиотека Уфимского государственного нефтяного технического университета <http://bibl.rusoil.net/> ,
- Библиотечно-информационный комплекс Ухтинского государственного технического университета УГТУ <http://lib.ugtu.net/books>
- Электронная справочная система нормативно-технической документации «Технорматив»
- ЭКБСОН- информационная система доступа к электронным каталогам библиотек сферы образования и науки.

9.3. Лицензионное и свободно распространяемое программное обеспечение, в т.ч. отечественного производства:

- Microsoft Windows,
- Microsoft Office Professional Plus,
- Visual Studio Code,
- Visual Studio Community.

10. Материально-техническое обеспечение дисциплины/модуля

Помещения для проведения всех видов работы, предусмотренных учебным планом, укомплектованы необходимым оборудованием и техническими средствами обучения.

Таблица 10.1

Обеспеченность материально-технических условий реализации ОПОП ВО

№ п/п	Наименование учебных предметов, курсов, дисциплин (модулей), практики, иных видов учебной деятельности, предусмотренных учебным планом образовательной программы	Наименование помещений для проведения всех видов учебной деятельности, предусмотренной учебным планом, в том числе помещения для самостоятельной работы, с указанием перечня основного оборудования, учебно-наглядных пособий	Адрес (местоположение) помещений для проведения всех видов учебной деятельности, предусмотренной учебным планом (в случае реализации образовательной программы в сетевой форме дополнительно указывается наименование организации, с которой заключен договор)
1	2	3	4
	Методы обработки больших данных	Лабораторные занятия: Учебная аудитория для проведения лабораторных работ; текущего контроля и промежуточной аттестации Учебная мебель: столы, стулья, доска аудиторная. Компьютер в комплекте – 1 шт., проектор – 1 шт., проекционный экран – 1 шт.	625039, г. Тюмень, ул. Мельникайте, д. 70

11. Методические указания по организации СРС

11.1. Методические указания по подготовке к лабораторным занятиям.

Лабораторные занятия способствуют углублённому изучению дисциплины и служат основной формой подведения итогов самостоятельной работы студентов. Основная цель лабораторных занятий заключается не только углубить и закрепить теоретические знания, но и сформировать практические компетенции, необходимые будущим специалистам.

На лабораторных занятиях приветствуется активное участие в обсуждении конкретных ситуаций, способность на основе полученных знаний находить наиболее эффективные решения поставленных проблем, уметь находить полезный дополнительный материал по тематике занятий.

Студенту рекомендуется следующая схема подготовки к занятию:

1. Изучить рекомендованную литературу;
2. При затруднениях сформулировать вопросы к преподавателю

3. После выполнения лабораторной работы оформит отчет и подготовиться к защите.

11.2. Методические указания по организации самостоятельной работы.

Самостоятельная работа является одной из важнейших форм изучения любой дисциплины. Она позволяет систематизировать и углубить теоретические знания, закрепить умения и навыки, способствует развитию умений пользоваться научной и учебно-методической литературой. Познавательная деятельность в процессе самостоятельной работы требует от студента высокого уровня активности и самоорганизованности.

В учебном процессе выделяют два вида самостоятельной работы: аудиторная и внеаудиторная.

Аудиторная самостоятельная работа по дисциплине выполняется на учебных занятиях под непосредственным руководством преподавателя и по его заданию.

Внеаудиторная самостоятельная работа студентов представляет собой логическое продолжение аудиторных занятий. Затраты времени на выполнение этой работы регламентируются рабочим учебным планом. Режим работы выбирает сам обучающийся в зависимости от своих способностей и конкретных условий.

Самостоятельная работа может осуществляться индивидуально или группами студентов в зависимости от цели, объема, конкретной тематики самостоятельной работы, уровня сложности, уровня умений студентов.

Самостоятельная работа включает в себя работу с конспектом лекций, изучение и конспектирование рекомендуемой литературы, подготовка мультимедиа-сообщений/докладов, подготовка реферата, тестирование, решение задач и упражнений по образцу, решение вариативных задач, выполнение чертежей, схем, расчетов (графических работ), решение ситуационных (профессиональных) задач, подготовка к деловым играм, проектирование и моделирование разных видов и компонентов профессиональной деятельности, научно-исследовательскую работу и др.

Контроль результатов внеаудиторной самостоятельной работы студентов может осуществляться в пределах времени, отведенного на обязательные учебные занятия по дисциплине и внеаудиторную самостоятельную работу студентов по дисциплине, может проходить в письменной, устной или смешанной форме.

Планируемые результаты обучения для формирования компетенции и критерии их оценивания

Дисциплина Методы обработки больших данных

Код, направление подготовки 45.03.04 Интеллектуальные системы в гуманитарной сфере

Направленность (профиль) Разработка и программирование интеллектуальных систем

Код компетенции	Код и наименование результата обучения по дисциплине	Критерии оценивания результатов обучения			
		1-2	3	4	5
ОПК-1	Знать: (31) математические модели обработки больших данных	Не знает математические модели обработки больших данных	Демонстрирует знание отдельных математических моделей обработки больших данных	Демонстрирует достаточное знание математических моделей обработки больших данных	Демонстрирует исчерпывающее знание математических моделей обработки больших данных
	Уметь: (У1) применять математические модели и алгоритмы для обработки больших данных	Не умеет применять математические модели и алгоритмы для обработки больших данных	Демонстрирует умение применять отдельные математические модели и алгоритмы для обработки больших данных	Демонстрирует достаточные умения применять математические модели и алгоритмы для обработки больших данных	Демонстрирует исчерпывающие умения применять математические модели и алгоритмы для обработки больших данных
	Владеть: (В1) практическими навыками исследования в области применения методов, алгоритмов и технологий обработки больших данных	Не имеет практических навыков исследования в области применения методов, алгоритмов и технологий обработки больших данных	Демонстрирует практические навыки исследования в области применения методов, алгоритмов и технологий обработки больших данных, но допускает существенные ошибки	Демонстрирует практические навыки исследования в области применения методов, алгоритмов и технологий обработки больших данных, но допускает незначительные ошибки	Демонстрирует практические навыки исследования в области применения методов, алгоритмов и технологий обработки больших данных
ОПК-2	Знать: (32) архитектуру распределенных файловых систем	Не знает архитектуру распределенных файловых систем	Демонстрирует знание основных компонентов распределенных файловых систем, но не знает основные проблемы в работе распределенных систем, виды отказов узлов, связи между узлами	Демонстрирует достаточные знания в области архитектуры распределенных файловых систем	Демонстрирует исчерпывающие знания в области архитектуры распределенных файловых систем

Код компетенции	Код и наименование результата обучения по дисциплине	Критерии оценивания результатов обучения			
		1-2	3	4	5
	Уметь: (У2) находить информацию по большим данным и методам их обработки с использованием современных справочных и библиотечных системам и системам дистанционного образования	Не умеет находить информацию по большим данным и методам их обработки с использованием современных справочных и библиотечных системам и системам дистанционного образования	Демонстрирует умение находить информацию по большим данным и методам их обработки, но не способен обобщать, анализировать, систематизировать и представлять результаты исследований	Демонстрирует достаточные умения поиска, анализа, обобщения и представления информации по большим данным и методам их обработки	Демонстрирует исчерпывающие умения поиска, анализа, обобщения и представления информации по большим данным и методам их обработки
	Владеть: (В2) практическими навыками поиска информации по большим данным и методам их обработки с использованием современных справочных и библиотечных системам и системам дистанционного образования	Не имеет практических навыков поиска информации по большим данным и методам их обработки с использованием современных справочных и библиотечных систем и систем дистанционного образования	Демонстрирует практические навыки поиска информации по большим данным и методам их обработки с использованием современных справочных и библиотечных систем и систем дистанционного образования, но не владеет навыками анализа, обобщения и систематизации	Демонстрирует достаточные практические навыки поиска информации по большим данным и методам их обработки с использованием современных справочных и библиотечных систем и систем дистанционного образования	Демонстрирует исчерпывающие практические навыки поиска информации по большим данным и методам их обработки с использованием современных справочных и библиотечных систем и систем дистанционного образования
ОПК-5	Знать: (З3) основные требования и стандарты в области описания и документирования методов обработки больших данных	Не знает основные требования и стандарты в области описания и документирования методов обработки больших данных	Демонстрирует знание отдельных требований и стандартов в области описания обработки больших данных, но не знает стандарты в области их документирования	Демонстрирует достаточные знания основных требований и стандартов в области описания и документирования обработки больших данных	Демонстрирует исчерпывающие знания основных требований и стандартов в области описания и документирования обработки больших данных

Код компетенции	Код и наименование результата обучения по дисциплине	Критерии оценивания результатов обучения			
		1-2	3	4	5
	Уметь: (У3) оценивать функциональные возможности систем и технологий для обработки больших данных на основании технической документации	Не умеет оценивать функциональные возможности систем и технологий для обработки больших данных на основании технической документации	Демонстрирует умение оценивать функциональные возможности систем и технологий для обработки больших данных на основании технической документации, но допускает существенные ошибки	Демонстрирует умение оценивать функциональные возможности систем и технологий для обработки больших данных на основании технической документации, но допускает при этом несущественные ошибки	Демонстрирует исчерпывающие умения оценивать функциональные возможности систем и технологий для обработки больших данных на основании технической документации
	Владеть: (В3) практическим опытом изучения систем обработки больших данных	Не имеет практического опыта изучения систем обработки больших данных	Демонстрирует навыки изучения систем обработки больших данных, но при этом часто требует консультаций и помощи	Демонстрирует достаточные навыки изучения систем обработки больших данных	Имеет положительный практический опыт изучения систем обработки больших данных

КАРТА
обеспеченности дисциплины учебной и учебно-методической литературой

Дисциплина Методы обработки больших данных

Код, направление подготовки 45.03.04 Интеллектуальные системы в гуманитарной сфере

Направленность (профиль) Разработка и программирование интеллектуальных систем

№ п/п	Название учебного, учебно-методического издания, автор, издательство, вид издания, год издания	Количество экземпляров в БИК	Контингент обучающихся, использующих указанную литературу	Обеспеченность обучающихся литературой, %	Наличие электронного варианта в ЭБС (+/-)
1	Паттерсон, Д. Глубокое обучение с точки зрения практика / Д. Паттерсон, А. Гибсон. — Москва : ДМК Пресс, 2018. — 418 с. — ISBN 978-5-97060-481-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/116122	ЭР*	30	100	+
2	Шарден, Б. Крупномасштабное машинное обучение вместе с Python : учебное пособие / Б. Шарден, Л. Массарон, А. Боскетти ; перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2018. — 358 с. — ISBN 978-5-97060-506-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/105836	ЭР*	30	100	+

ЭР* – электронный ресурс для автор. пользователей доступен через Электронный каталог/Электронную библиотеку ТИУ <http://webirbis.tsogu.ru/>