

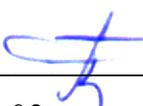
Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования



**Пермский национальный исследовательский
политехнический университет**

УТВЕРЖДАЮ

Проректор по образовательной
деятельности

 А.Б. Петроченков

« 03 » апреля 20 23 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Дисциплина: Библиотеки и фреймворки для кластерных технологий обработки
больших данных

(наименование)

Форма обучения: _____ очная
(очная/очно-заочная/заочная)

Уровень высшего образования: _____ магистратура
(бакалавриат/специалитет/магистратура)

Общая трудоёмкость: _____ 144 (4)
(часы (ЗЕ))

Направление подготовки: _____ 09.04.01 Информатика и вычислительная техника
(код и наименование направления)

Направленность: Информационные технологии интеллектуальной обработки
больших данных (Big Data)

(наименование образовательной программы)

1. Общие положения

1.1. Цели и задачи дисциплины

Формирование комплекса знаний, умений и навыков в области построения распределенных кластерных систем для обработки больших данных средствами современных языков программирования.

1.2. Изучаемые объекты дисциплины

Большие данные; распределенные кластерные системы; параллельные вычисления; язык программирования Python; модули и библиотеки научных вычислений; модули и библиотеки обработки данных.

1.3. Входные требования

Не предусмотрены

2. Планируемые результаты обучения по дисциплине

Компетенция	Индекс индикатора	Планируемые результаты обучения по дисциплине (знать, уметь, владеть)	Индикатор достижения компетенции, с которым соотнесены планируемые результаты обучения	Средства оценки
ПК-2.2	ИД-1ПК-2.2	Знает порядок работы и особенности библиотек и фреймворков, используемых для развертывания кластеров обработки больших данных	Знает порядок работы и особенности компьютерных программ и баз данных, используемых для мониторинга функционирования инфокоммуникационных систем и сервисов	Дифференцированный зачет
ПК-2.2	ИД-2ПК-2.2	Умеет разрабатывать целевую архитектуру вычислительных кластеров обработки больших данных и стратегию ее реализации	Умеет разрабатывать целевую архитектуру систем автоматизированного мониторинга и контроля функционирования инфокоммуникационных систем и сервисов и стратегию ее реализации	Отчёт по практическому занятию
ПК-2.2	ИД-3ПК-2.2	Владеет навыками поиска информации по инновационным и конкурентным библиотекам и фреймворкам для развертывания кластеров обработки больших данных	Владеет навыками поиска информации по инновационным и конкурентным системам автоматизированного мониторинга и контроля функционирования инфокоммуникационных систем и сервисов	Защита лабораторной работы

3. Объем и виды учебной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		3	
1. Проведение учебных занятий (включая проведение текущего контроля успеваемости) в форме:	54	54	
1.1. Контактная аудиторная работа, из них:			
- лекции (Л)	18	18	
- лабораторные работы (ЛР)	18	18	
- практические занятия, семинары и (или) другие виды занятий семинарского типа (ПЗ)	16	16	
- контроль самостоятельной работы (КСР)	2	2	
- контрольная работа			
1.2. Самостоятельная работа студентов (СРС)	90	90	
2. Промежуточная аттестация			
Экзамен			
Дифференцированный зачет	9	9	
Зачет			
Курсовой проект (КП)			
Курсовая работа (КР)			
Общая трудоемкость дисциплины	144	144	

4. Содержание дисциплины

Наименование разделов дисциплины с кратким содержанием	Объем аудиторных занятий по видам в часах			Объем внеаудиторных занятий по видам в часах
	Л	ЛР	ПЗ	СРС
3-й семестр				
Введение в кластерные технологии	4	2	0	10
Понятие параллельных вычислений. Понятие вычислительного кластера. ЦОД. Подходы к архитектуре распределенных приложений. DAG-топология. Поточковая обработка данных. Масштабируемость. Отказоустойчивость.				
Сбор и хранение больших данных	4	4	4	20
Понятие больших данных. Понятие «озера данных» (Data Lake). Мультимодальные большие данные. Распределенные хранилища. SQL- и NoSQL-подходы. Колоночные базы данных. Графовые базы данных. Распределенные файловые системы. HDFS. Cassandra. Greenplum. ClickHouse.				
Инструменты распределенного анализа больших данных	4	4	4	20
Задачи распределенного анализа. ETL-процедуры в кластере. Интеллектуальный анализ данных (data mining). Apache Spark. Hadoop. Zookeeper. Samza. Flink. Kafka. Elasticsearch.				

Наименование разделов дисциплины с кратким содержанием	Объем аудиторных занятий по видам в часах			Объем внеаудиторных занятий по видам в часах
	Л	ЛР	ПЗ	СРС
Использование кластерных технологий в машинном обучении	4	4	4	20
Задачи машинного обучения, требующие больших данных. Обучение ИНС на кластерах. CUDA-кластеры. PyTorch. Onyx. TensorFlow. Scikit-Learn. Большие языковые модели (LLMs). Трансформеры (BERT, GPT). Мультимодальные генеративные сети (DALL-E, Stable Diffusion, Whisper).				
Инструменты распределенного решения научных задач	2	4	4	20
Научоемкие вычислительные задачи. Понятие «добровольные вычисления» в контексте грид-вычислений. BOINC. World Community Grid. SETI@Home. Folding@Home. Rosetta@home. Преимущества и критика проектов.				
ИТОГО по 3-му семестру	18	18	16	90
ИТОГО по дисциплине	18	18	16	90

Тематика примерных практических занятий

№ п.п.	Наименование темы практического (семинарского) занятия
1	Сравнение распределенных баз данных и файловых систем
2	Обзор возможностей Apache Hadoop
3	Обзор возможностей Elasticsearch
4	Обзор проекта folding@home

Тематика примерных лабораторных работ

№ п.п.	Наименование темы лабораторной работы
1	Работа с Apache Spark
2	Работа с СУБД Greenplum
3	Разработка распределенной ETL-процедуры
4	Обучение готовой архитектуры ИНС на больших данных

5. Организационно-педагогические условия

5.1. Образовательные технологии, используемые для формирования компетенций

Проведение лекционных занятий по дисциплине основывается на активном методе обучения, при котором учащиеся не пассивные слушатели, а активные участники занятия, отвечающие на вопросы преподавателя. Вопросы преподавателя нацелены на активизацию процессов усвоения материала, а также на развитие логического мышления. Преподаватель заранее намечает список вопросов, стимулирующих ассоциативное мышление и установление связей с ранее освоенным материалом.

Практические занятия проводятся на основе реализации метода обучения действием: определяются проблемные области, формируются группы. При проведении практических занятий преследуются следующие цели: применение знаний отдельных дисциплин и креативных методов для решения проблем и принятия решений; отработка у обучающихся навыков командной работы, межличностных коммуникаций и развитие лидерских качеств; закрепление основ теоретических знаний.

Проведение лабораторных занятий основывается на интерактивном методе обучения, при котором обучающиеся взаимодействуют не только с преподавателем, но и друг с другом. При этом доминирует активность учащихся в процессе обучения. Место преподавателя в интерактивных занятиях сводится к направлению деятельности обучающихся на достижение целей занятия.

При проведении учебных занятий используются интерактивные лекции, групповые дискуссии, ролевые игры, тренинги и анализ ситуаций и имитационных моделей.

5.2. Методические указания для обучающихся по изучению дисциплины

При изучении дисциплины обучающимся целесообразно выполнять следующие рекомендации:

1. Изучение учебной дисциплины должно вестись систематически.
2. После изучения какого-либо раздела по учебнику или конспектным материалам рекомендуется по памяти воспроизвести основные термины, определения, понятия раздела.
3. Особое внимание следует уделить выполнению отчетов по практическим занятиям, лабораторным работам и индивидуальным комплексным заданиям на самостоятельную работу.
4. Вся тематика вопросов, изучаемых самостоятельно, задается на лекциях преподавателем. Им же даются источники (в первую очередь вновь изданные в периодической научной литературе) для более детального понимания вопросов, озвученных на лекции.

6. Перечень учебно-методического и информационного обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Печатная учебно-методическая литература

№ п/п	Библиографическое описание (автор, заглавие, вид издания, место, издательство, год издания, количество страниц)	Количество экземпляров в библиотеке
1. Основная литература		
1	Барский А. Б. Параллельные информационные технологии : учебное пособие. Москва : ИНТУИТ, 2007. 502 с.	4
2. Дополнительная литература		
2.1. Учебные и научные издания		

1	Дейтел П., Дейтел Х. Python: Искусственный интеллект, большие данные и облачные вычисления : пер. с англ. Санкт-Петербург [и др.] : Питер, 2020. 861 с. 69,660 усл. печ. л.	4
2	Доусон М. Програмируем на Python : пер. с англ. Санкт-Петербург [и др.] : Питер, 2021. 414 с. 33,540 усл. печ. л.	6
3	Мыльников Л. А. Статистические методы интеллектуального анализа данных. Санкт-Петербург : БХВ-Петербург, 2021. 240 с. 19,35 усл. печ. л.	14
2.2. Периодические издания		
	Не используется	
2.3. Нормативно-технические издания		
	Не используется	
3. Методические указания для студентов по освоению дисциплины		
	Не используется	
4. Учебно-методическое обеспечение самостоятельной работы студента		
	Не используется	

6.2. Электронная учебно-методическая литература

Вид литературы	Наименование разработки	Ссылка на информационный ресурс	Доступность (сеть Интернет / локальная сеть; авторизованный / свободный доступ)
Дополнительная литература	Основы работы с технологией параллельных вычислений CUDA : учебное пособие [Электронный ресурс]	https://e.lanbook.com/book/220436	локальная сеть; свободный доступ
Дополнительная литература	Сузи Р.А. Язык программирования Python : учебное пособие [Электронный ресурс]	http://www.iprbookshop.ru/52211.html	локальная сеть; свободный доступ
Дополнительная литература	Халеева Е.П. Анализ данных средствами языка R : учебное пособие [Электронный ресурс]	https://www.iprbookshop.ru/119442.html	локальная сеть; свободный доступ
Основная литература	Барский, А. Б. Параллельные информационные технологии : учебное пособие [Электронный ресурс]	https://www.iprbookshop.ru/97573.html	локальная сеть; свободный доступ
Основная литература	Гергель, В. П. Технологии построения и использования кластерных систем : учебное пособие [Электронный ресурс]	https://e.lanbook.com/book/100517	локальная сеть; свободный доступ
Основная литература	Протоodyаконов А.В. Алгоритмы Data Science и их практическая реализация на Python : учебное пособие [Электронный ресурс]	https://www.iprbookshop.ru/124000.html	локальная сеть; свободный доступ

6.3. Лицензионное и свободно распространяемое программное обеспечение, используемое при осуществлении образовательного процесса по дисциплине

Вид ПО	Наименование ПО
Операционные системы	Debian (GNU GPL)
Офисные приложения.	LibreOffice 6.2.4. OpenSource, бесплатен.
Системы управления проектами, исследованиями, разработкой, проектированием, моделированием и внедрением	Protege
Среды разработки, тестирования и отладки	PIP (The Python Package Installer) Free
Среды разработки, тестирования и отладки	PostgreSQL (PostgreSQL License)

6.4. Современные профессиональные базы данных и информационные справочные системы, используемые при осуществлении образовательного процесса по дисциплине

Наименование	Ссылка на информационный ресурс
База данных Elsevier "Freedom Collection"	https://www.elsevier.com/
База данных Scopus	https://www.scopus.com/
База данных научной электронной библиотеки (eLIBRARY.RU)	https://elibrary.ru/
Научная библиотека Пермского национального исследовательского политехнического университета	http://lib.pstu.ru/
Электронно-библиотечная система Лань	https://e.lanbook.com/
Электронно-библиотечная система IPRbooks	http://www.iprbookshop.ru/
Информационные ресурсы Сети КонсультантПлюс	http://www.consultant.ru/
Информационно-справочная система нормативно-технической документации "Техэксперт: нормы, правила, стандарты и законодательства России"	https://техэксперт.сайт/

7. Материально-техническое обеспечение образовательного процесса по дисциплине

Вид занятий	Наименование необходимого основного оборудования и технических средств обучения	Количество единиц
Лабораторная работа	ПК с предустановленным интерпретатором Python версии 3.8 или выше и средой разработки	20
Лекция	Мультимедийный проектор	1
Практическое занятие	ПК с предустановленным интерпретатором Python версии 3.8 или выше и средой разработки	20

8. Фонд оценочных средств дисциплины

Описан в отдельном документе

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Пермский национальный исследовательский политехнический
университет»

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

для проведения промежуточной аттестации обучающихся по дисциплине
«Методы и алгоритмы обработки слабоструктурированных данных»
Приложение к рабочей программе дисциплины

Направление подготовки:	09.04.01 Информатика и вычислительная техника
Направленность (профиль) образовательной программы:	Информационные технологии интеллектуальной обработки больших данных (Big Data)
Квалификация выпускника:	«Магистр»
Выпускающая кафедра:	Информационные технологии и автоматизированные системы
Форма обучения:	Очная

Курс: 2

Семестр: 3

Трудоёмкость:

Кредитов по рабочему учебному плану:	4	ЗЕ
Часов по рабочему учебному плану:	144	ч.

Форма промежуточной аттестации:

Дифференцированный зачет: 3 семестр

Пермь 2023 г.

Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине является частью (приложением) к рабочей программе дисциплины. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине разработан в соответствии с общей частью фонда оценочных средств для проведения промежуточной аттестации основной образовательной программы, которая устанавливает систему оценивания результатов промежуточной аттестации и критерии выставления оценок. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине устанавливает формы и процедуры текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине.

1. Перечень контролируемых результатов обучения по дисциплине, объекты оценивания и виды контроля

Согласно РПД, освоение учебного материала дисциплины запланировано в течение одного семестра (3-го семестра учебного плана) и разбито на 5 учебных модулей. В каждом модуле предусмотрены аудиторские лекционные, практические и лабораторные занятия, а также самостоятельная работа студентов. В рамках освоения учебного материала дисциплины формируются компоненты компетенций *знать, уметь, владеть*, указанные в РПД, которые выступают в качестве контролируемых результатов обучения по дисциплине (таблица 1.1).

Контроль уровня усвоенных знаний, освоенных умений и приобретенных владений осуществляется в рамках текущего, рубежного и промежуточного контроля при изучении теоретического материала, сдаче отчетов по лабораторным работам и дифференцированного зачета. Виды контроля сведены в таблицу 1.1.

Таблица 1.1. Перечень контролируемых результатов обучения по дисциплине

Контролируемые результаты обучения по дисциплине (ЗУВы)	Вид контроля					
	Текущий		Промежуточный /рубежный		Итоговый	
	С	ТО	ОЛР	Т/КР	Диф. зачет	
Усвоенные знания						
З.1 Знает порядок работы и особенности библиотек и фреймворков, используемых для развертывания кластеров обработки больших данных		ТО			ТЗ	
Освоенные умения						
У.1 Умеет разрабатывать целевую архитектуру вычислительных кластеров обработки больших данных и стратегию ее реализации			ОЛР1 ОЛР3		КЗ	
Приобретенные владения						
В.1 Владеет навыками поиска информации по инновационным и конкурентным библиотекам и фреймворкам для развертывания кластеров обработки больших данных			ОЛР2 ОЛР4		КЗ	

С – собеседование по теме; *ТО* – коллоквиум (теоретический опрос); *КЗ* – кейс-задача (индивидуальное задание); *ОЛР* – отчет по лабораторной работе; *Т/КР* – рубежное тестирование (контрольная работа); *ТВ* – теоретический вопрос; *ПЗ* – практическое задание; *КЗ* – комплексное задание экзамена.

Итоговой оценкой достижения результатов обучения по дисциплине является промежуточная аттестация в виде дифференцированного зачета,

проводимая с учетом результатов текущего и рубежного контроля.

2. Виды контроля, типовые контрольные задания и шкалы оценивания результатов обучения

Текущий контроль успеваемости имеет целью обеспечение максимальной эффективности учебного процесса, управление процессом формирования заданных компетенций обучаемых, повышение мотивации к учебе и предусматривает оценивание хода освоения дисциплины. В соответствии с Положением о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся по образовательным программам высшего образования – программам бакалавриата, специалитета и магистратуры в ПНИПУ предусмотрены следующие виды и периодичность текущего контроля успеваемости обучающихся:

- входной контроль, проверка исходного уровня подготовленности обучаемого и его соответствия предъявляемым требованиям для изучения данной дисциплины;

- текущий контроль усвоения материала (уровня освоения компонента «знать» заданных компетенций) на каждом групповом занятии и контроль посещаемости лекционных занятий;

- промежуточный и рубежный контроль освоения обучаемыми отдельных компонентов «знать», «уметь» заданных компетенций путем компьютерного или бланочного тестирования, контрольных опросов, контрольных работ (индивидуальных домашних заданий), защиты отчетов по лабораторным работам, рефератов, эссе и т.д.

Рубежный контроль по дисциплине проводится на следующей неделе после прохождения модуля дисциплины, а промежуточный – во время каждого контрольного мероприятия внутри модулей дисциплины;

- межсессионная аттестация, единовременное подведение итогов текущей успеваемости не менее одного раза в семестр по всем дисциплинам для каждого направления подготовки (специальности), курса, группы;

- контроль остаточных знаний.

2.1. Текущий контроль усвоения материала

Текущий контроль усвоения материала в форме собеседования или выборочного теоретического опроса студентов проводится по каждой теме. Результаты по 4-балльной шкале оценивания заносятся в книжку преподавателя и учитываются в виде интегральной оценки при проведении промежуточной аттестации.

2.2. Рубежный (промежуточный) контроль

Рубежный (промежуточный) контроль для комплексного оценивания усвоенных знаний, усвоенных умений и приобретенных владений (таблица 1.1) проводится в форме защиты лабораторных работ и выполнения кейс-задач.

2.2.1. Защита лабораторных работ

Всего запланировано 4 лабораторных работы. Типовые темы лабораторных работ приведены в РПД.

Защита лабораторной работы проводится индивидуально каждым студентом

или группой студентов. Типовые шкала и критерии оценки приведены в общей части ФОС образовательной программы.

2.3. Промежуточная аттестация (итоговый контроль)

Допуск к промежуточной аттестации осуществляется по результатам текущего и рубежного контроля. Условиями допуска являются успешная сдача всех лабораторных работ и положительная интегральная оценка по результатам текущего и рубежного контроля.

2.3.1. Процедура промежуточной аттестации без дополнительного аттестационного испытания

Промежуточная аттестация проводится в форме дифференцированного зачета. Зачет по дисциплине основывается на результатах выполнения лабораторных работ и участия в семинарах по данной дисциплине.

Критерии выведения итоговой оценки за компоненты компетенций при проведении промежуточной аттестации в виде зачета приведены в общей части ФОС образовательной программы.

2.3.2. Процедура промежуточной аттестации с проведением аттестационного испытания

В отдельных случаях (например, в случае переаттестации дисциплины) промежуточная аттестация в виде зачета по дисциплине может проводиться с проведением аттестационного испытания по билетам. Билет содержит теоретические вопросы (ТВ) для проверки усвоенных знаний и комплексные задания (КЗ) для контроля уровня усвоенных умений и приобретенных владений всех заявленных компетенций.

Билет формируется таким образом, чтобы в него попали вопросы и практические задания, контролирующие уровень сформированности всех заявленных компетенций.

2.4.2.1. Типовые вопросы и задания для зачета по дисциплине Типовые вопросы для контроля усвоенных знаний:

1. Понятие параллельных вычислений.
2. Понятие вычислительного кластера.
3. ЦОД.
4. Подходы к архитектуре распределенных приложений.
5. DAG-топология.
6. Поточковая обработка данных.
7. Понятие больших данных.
8. Понятие «озера данных» (Data Lake).
9. Распределенные хранилища.
10. Колоночные базы данных.
11. Графовые базы данных.
12. Распределенные файловые системы.
13. ETL-процедуры в кластере.
14. Обзор возможностей Apache Spark.

15. Обзор возможностей Apache Hadoop.
16. Обзор возможностей Apache Kafka.
17. Обзор возможностей Elasticsearch.
18. Задачи машинного обучения, требующие больших данных.
19. Наукоемкие вычислительные задачи.
20. Понятие «добровольные вычисления» в контексте грид-вычислений.

Типовые комплексные задания для контроля освоенных умений и приобретенных владений:

1. Описать процедуру преобразования pandas DataFrame в Apache Spark DataFrame
2. Дана схема DataFrame: `schema='a long, b double, c string, d date, e timestamp'`. Привести примеры исходных данных, подходящих под указанную схему.
3. Описать реализацию распределенной ETL-процедуры (ETL pipeline).

2.3.2.2. Шкалы оценивания результатов обучения на зачете

Оценка результатов обучения по дисциплине в форме уровня сформированности компонентов *знать, уметь, владеть* заявленных компетенций проводится по 4-х балльной шкале оценивания.

Типовые шкала и критерии оценки результатов обучения при сдаче зачета для компонентов *знать, уметь и владеть* приведены в общей части ФОС образовательной программы.

3. Критерии оценивания уровня сформированности компонентов и компетенций

3.1. Оценка уровня сформированности компонентов компетенций

При оценке уровня сформированности компетенций в рамках выборочного контроля при экзамене считается, что *полученная оценка за компонент проверяемой в билете компетенции обобщается на соответствующий компонент всех компетенций, формируемых в рамках данной учебной дисциплины.*

Типовые критерии и шкалы оценивания уровня сформированности компонентов компетенций приведены в общей части ФОС образовательной программы.

3.2. Оценка уровня сформированности компетенций

Общая оценка уровня сформированности всех компетенций проводится путем агрегирования оценок, полученных студентом за каждый компонент формируемых компетенций, с учетом результатов текущего и рубежного контроля в виде интегральной оценки по 4-х балльной шкале. Все результаты контроля заносятся в оценочный лист и заполняются преподавателем по итогам промежуточной аттестации.

Форма оценочного листа и требования к его заполнению приведены в общей части ФОС образовательной программы.

При формировании итоговой оценки промежуточной аттестации в виде экзамена используются типовые критерии, приведенные в общей части ФОС образовательной программы.