

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

« 27 » декабря 20 22 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Дисциплина: Системы управления большими данными
(наименование)

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

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

Общая трудоёмкость: 180 (5)
(часы (ЗЕ))

Направление подготовки: 15.04.04 Автоматизация технологических процессов и производств
(код и наименование направления)

Направленность: Цифровые технологии проектирования систем управления и контроля авиационных двигателей и энергетических установок
(наименование образовательной программы)

1. Общие положения

1.1. Цели и задачи дисциплины

Формирование профессиональной компетенции в области разработки и использования систем обработки и анализа больших массивов данных.

В результате изучения дисциплины у студента формируются:

-представления о феномене больших данных, о научных и технических проблемах и возможностях, связанных с их появлением, о трендах в области технологий хранения и анализа больших данных

-знания причин возникновения тренда больших данных, процессов анализа больших данных, основных подходов к обработке больших массивов данных

-умения выбрать подходящую технологию хранения больших данных, умения работать с NoSQL СУБД

1.2. Изучаемые объекты дисциплины

методы анализа и хранения больших объемов данных, этапы жизненного цикла обработки больших данных, языки, наиболее приспособленные для обработки и аналитики больших данных, способы организации хранения и доступа к большим данным

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-2.4	ИД-1ПК-2.4	Знает: - основные тенденции накопления данных, свойства и особенности больших данных. - технологии, применяемые для хранения, извлечения, поиска и анализа больших данных	Знает основные требования к информационной безопасности в системах автоматизации и управления; современные программные средства и системы управления базами и банками данных, в том числе отечественного производства	Дифференцированный зачет

Компетенция	Индекс индикатора	Планируемые результаты обучения по дисциплине (знать, уметь, владеть)	Индикатор достижения компетенции, с которым соотнесены планируемые результаты обучения	Средства оценки
ПК-2.4	ИД-2ПК-2.4	Умеет: обрабатывать и анализировать большие объемы данных с помощью современных программных средств и систем управления базами и банками данных	Умеет использовать программные средства и системы управления базами и банками данных для решения задач в области профессиональной деятельности с учетом требований информационной безопасности	Отчёт по практическом у занятию
ПК-2.4	ИД-3ПК-2.4	Владеет: навыками использования современных инструментальных средства (в соответствии с профилем подготовки), способами хранения, поиска, извлечения и подготовки больших данных	Владеет навыками использования современных средств и систем управления базами и банками данных с учетом требований информационной безопасности	Защита лабораторной работы

3. Объем и виды учебной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		3	4
1. Проведение учебных занятий (включая проведение текущего контроля успеваемости) в форме:	72	36	36
1.1. Контактная аудиторная работа, из них:			
- лекции (Л)	36	18	18
- лабораторные работы (ЛР)			
- практические занятия, семинары и (или) другие виды занятий семинарского типа (ПЗ)	32	16	16
- контроль самостоятельной работы (КСР)	4	2	2
- контрольная работа			
1.2. Самостоятельная работа студентов (СРС)	108	54	54
2. Промежуточная аттестация			
Экзамен			
Дифференцированный зачет	9		9
Зачет	9	9	
Курсовой проект (КП)			
Курсовая работа (КР)			
Общая трудоемкость дисциплины	180	90	90

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				
Основные понятия и концепции, относящиеся к Большим данным (Big Data), подходы и определения	4	0	0	12
Определяющие характеристики больших данных (набор признаков VVV): объем (volume), скорость (velocity), многообразие (variety) Структурированные и неструктурированные данные Вертикальное и горизонтальное масштабирование Жизненный цикл данных				
Системы управления Большими данными	4	0	4	12
Классификация СУБД (по поддерживаемым моделям данных, по типам хранимой информации, по способу организации доступа, по архитектуре системы) Проблемы коллективного доступа к данным				
Реляционные СУБД применительно к Big Data	6	0	6	14
Понятие модели данных. Основные операции над данными. Ограничения целостности. Методология проектирования БД. Задачи инфологического, логического и физического этапов проектирования. Нормализация отношений. Запросы языка SQL - манипулирование данными				
Аналитика больших данных	4	0	6	16
Принципы и инструменты аналитики Задачи и компетенции аналитиков Big Data Прикладные инструменты анализа данных Big Data как рынок				
ИТОГО по 3-му семестру	18	0	16	54
4-й семестр				
NoSQL СУБД применительно к Big Data	4	0	4	18
СУБД «ключ — значение» Колоночные СУБД (семейство столбцов) Документоориентированные СУБД				
MongoDB	8	0	6	18
Организация данных в MongoDB, манипулирование данными, объекты и методы для работы с данными. Репликация и шардинг в MongoDB				
Хранилища данных, Аналитические платформы	6	0	6	18
Интерактивная аналитическая обработка OLAP (Online Analytical Processing) Характеристики BigData и хранилища данных				

Наименование разделов дисциплины с кратким содержанием	Объем аудиторных занятий по видам в часах			Объем внеаудиторных занятий по видам в часах
	Л	ЛР	ПЗ	СРС
ИТОГО по 4-му семестру	18	0	16	54
ИТОГО по дисциплине	36	0	32	108

Тематика примерных практических занятий

№ п.п.	Наименование темы практического (семинарского) занятия
1	Способы анализа данных. Получение первичных элементарных характеристик о наборах данных
2	Общие сведения о реляционной модели данных (РМД), основные термины, определения. Проектирование БД. Обзор нотаций описания БД.
3	Язык структурированных запросов (SQL). DDL DML
4	Обзор технологий No SQL, Сравнение технологий доступа к данным
5	Решение задач Data Mining. Задачи классификации, кластеризации: деревья решений, RandomForest

5. Организационно-педагогические условия

5.1. Образовательные технологии, используемые для формирования компетенций

<p>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.</p> <p>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.</p> <p>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.</p> <p>При проведении учебных занятий используются интерактивные лекции, групповые дискуссии, ролевые игры, тренинги и анализ ситуаций и имитационных моделей.</p>
--

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	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	Косякин С. И. Базы данных : учебное пособие. Пермь : Изд-во ПГТУ, 2007. 87 с.	3
3	Макшанов А. В., Журавлев А. Е., Тындыкарь Л. Н. Большие данные. Big Data. : учебник для вузов. Санкт-Петербург [и др.] : Лань, 2021. 184 с. 15,28 усл. печ. л.	1
2. Дополнительная литература		
2.1. Учебные и научные издания		
1	Шустова Л. И., Тараканов О. В. Базы данных : учебник. Москва : ИНФРА-М, 2016. 303 с. 19,0 усл. печ. л.	3
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	Большие данные. Big Data. Авторы: Макшанов А. В., Журавлев А. Е., Тындыкарь Л. Н. Санкт-Петербург : Лань, 2021	http://elib.pstu.ru/Record/RUPNRPUelib3327	сеть Интернет; свободный доступ

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

Вид ПО	Наименование ПО
Операционные системы	Windows 10 (подп. Azure Dev Tools for Teaching)
Офисные приложения.	Microsoft Office Professional 2007. лиц. 42661567
Среды разработки, тестирования и отладки	PostgreSQL (PostgreSQL License)
Среды разработки, тестирования и отладки	Язык R

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

Наименование	Ссылка на информационный ресурс
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
Лекция	Проектор, экран, ПК или ноутбук, маркерная доска, маркер	1
Практическое занятие	ПК с установленным ПО в комплекте: системный блок, монитор, клавиатура, мышь	10

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
«Системы управления большими данными»
Приложение к рабочей программе дисциплины

Направление подготовки:	15.04.04 Автоматизация технологических процессов и производств
Направленность (профиль) образовательной программы:	Цифровые технологии проектирования систем управления и контроля авиационных двигателей и энергетических установок
Квалификация выпускника:	«Магистр»
Выпускающая кафедра:	Микропроцессорных средств автоматизации
Форма обучения:	Очная

Курс: 2 **Семестр:** 3, 4

Трудоёмкость:

Кредитов по рабочему учебному плану:	5	ЗЕ
Часов по рабочему учебному плану:	180	ч.

Форма промежуточной аттестации:

Зачёт: 3 семестр

Диф. Зачёт: 4 семестр

Пермь 2022

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 двух семестров (3, 4 -го семестра учебного плана).

Предусмотрены аудиторские лекционные и практические занятия, а также самостоятельная работа студентов. В рамках освоения учебного материала дисциплины формируются компоненты дисциплинарных компетенций *знать, уметь, владеть*, указанные в РПД, и которые выступают в качестве контролируемых результатов обучения (табл. 1.1).

Контроль уровня усвоенных знаний, освоенных умений и приобретенных владений осуществляется в рамках текущего, промежуточного и итогового контроля при изучении теоретического материала, сдаче отчетов по практическим работам и 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		Рубежный		Итоговый	
	С	ТО	ОПЗ	Т/КР		Зачёт
Усвоенные знания						
ИД-1ПК-2.4 Знает: - основные тенденции накопления данных, свойства и особенности больших данных.	С	ТО	ТО			ТВ
ИД-1ПК-2.4 Знает: - технологии, применяемые для хранения, извлечения, поиска и анализа больших данных	С	ТО	ТО			ТВ
Освоенные умения						
ИД-2ПК-2.4 Умеет: обрабатывать и анализировать большие объемы данных с помощью современных программных средств и систем управления базами и банками данных			ОПЗ			ПЗ
Приобретенные владения						
ИД-3ПК-2.4 Владеет: навыками использования современных инструментальных средства (в соответствии с профилем подготовки), способами хранения, поиска, извлечения и подготовки больших данных			ОПЗ			ПЗ

С – собеседование по теме; ТО – коллоквиум (теоретический опрос); КЗ – кейс-задача (индивидуальное задание); ОПЗ – отчет по практическому занятию; Т/КР – рубежное тестирование (контрольная работа); ТВ – теоретический вопрос; ПЗ – практическое задание; КЗ – комплексное задание дифференцированного зачета.

Итоговой оценкой достижения результатов обучения по дисциплине являются промежуточные аттестация в виде зачета (3-й семестр) и диф. зачета (4-й семестр), проводимые с учётом 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 бланчного 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. Промежуточный контроль

Промежуточный контроль для комплексного оценивания усвоенных знаний, усвоенных умений и приобретенных владений (табл. 1.1) проводится в форме защиты отчетов по практическим работам.

2.2.1. Защита отчетов по практическим занятиям

Всего запланировано 5 практических занятий. Типовые темы практических занятий приведены в РПД.

Защита отчетов по практическим занятиям 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 отчетов по всем практическим 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 зачета. Зачет по дисциплине основывается на результатах выполнения предыдущих индивидуальных заданий студента 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, практические задания (ПЗ) для проверки усвоенных умений и комплексные задания (КЗ) для контроля уровня приобретенных владений всех заявленных компетенций.

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

2.3.3. Типовые вопросы и задания для зачета по дисциплине

Типовые вопросы для контроля усвоенных знаний:

1. Основные понятия и концепции, относящиеся к Большим данным (Big Data), подходы и определения

2. Определяющие характеристики больших данных (набор признаков VVV): объём (volume), скорость (velocity), многообразие (variety). Структурированные и неструктурированные данные. Вертикальное и горизонтальное масштабирование

3. Классификация СУБД (по поддерживаемым моделям данных, по типам хранимой информации, по способу организации доступа, по архитектуре системы). Проблемы коллективного доступа к данным

4. Реляционные СУБД. Понятие модели данных. Основные операции над данными. Ограничения целостности. Методология проектирования БД

5. NoSQL СУБД. СУБД «ключ — значение». Колоночные СУБД (семейство

столбцов). Документоориентированные СУБД

6. Задачи и методы извлечения знаний из больших данных (Data Mining). Инструменты Data Mining. Методы сбора и предобработки данных из различных источников

Типовые вопросы и практические задания для контроля освоенных умений:

7. Разработать модель данных для РСУБД.

8. Разработать программный код для эффективной обработки данных большого объема с использованием языков программирования;

9. Разработать структуры документов для хранения больших данных в MongoDB

10. Подготовить данные (например, результаты испытаний) к анализу, преобразовать их в заданные формы представления данных.

2.4. 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 зачете считается, что *полученная оценка за компонент проверяемой в билете компетенции обобщается на соответствующий компонент всех компетенций, формируемых в рамках данной учебной дисциплины.*

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 зачета используются типовые критерии, приведенные 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.

Перечень вопросов к зачету:

1. Основные методы критического анализа и основы системного подхода как общенаучного метода изучения больших данных
2. Анализ задач больших данных с использованием основ критического анализа и системного подхода
3. Поиск необходимой для решения поставленной задачи информации, критически оценивая надежность различных источников информации на основе больших данных
4. Жизненный цикл больших данных
5. Критерии сопоставления различных вариантов решения поставленной задачи создания больших данных (Data Generation/Data Capture)
6. Критический анализ собранной информации на соответствие ее условиям и критериям решения поставленной задачи при обслуживании больших данных (Data Maintenance)
7. Отличие фактов от мнений, интерпретаций и оценок при анализе собранной информации в процессе синтеза больших данных (Data Synthesis)
8. Вопросы безопасности больших данных.
9. Технологии хранения больших данных
10. Сопоставление и оценка различных вариантов решения поставленной задачи, определяя их достоинства и недостатки, при использовании больших данных (Data Usage)
11. Многомерное представление данных и многомерный куб. Представление данных в виде гиперкуба. Многомерные базы данных и OLAP-куб. OLAP-анализ данных

Примеры задач:

1. Ознакомьтесь с доступными способами обработки данных. Для предложенных преподавателем данных выполните консолидацию, трансформацию, визуализацию данных.
2. Выполните построение ассоциативных правил для предложенных преподавателем данных, используя различные параметры построения ассоциативных правил. Сравните полученные результаты. Опишите 4-5 ассоциативных правил, полученных в ходе выполнения работы.
3. Используя механизм кластеризации реализованный на алгоритме k-means, основываясь на данных предложенных преподавателем, решите задачу распределения данных на кластеры и выявления скрытых закономерностей. Проанализируйте получившуюся картину распределения.
4. Постройте прогноз для предложенных преподавателем данных с помощью линейной регрессии. Проанализируйте построенную с помощью линейной регрессии модель прогноза.