

МИНОБРНАУКИ РОССИИ

**Федеральное государственное автономное образовательное
учреждение высшего образования "Пермский
государственный национальный исследовательский
университет"**

Физико-математический институт

**Авторы-составители: Русакова Ольга Леонидовна
Ромашкина Татьяна Витальевна**

Рабочая программа дисциплины

БОЛЬШИЕ ДАННЫЕ

Код УМК 102174

Утверждено
Протокол №1
от «19» июня 2024 г.

Пермь, 2024

1. Наименование дисциплины

Большие данные

2. Место дисциплины в структуре образовательной программы

Дисциплина входит в обязательную часть Блока « Б.1 » образовательной программы по направлениям подготовки (специальностям):

Направление: **01.03.04** Прикладная математика

направленность Интеллектуальный анализ данных, программирование и искусственный интеллект

3. Планируемые результаты обучения по дисциплине

В результате освоения дисциплины **Большие данные** у обучающегося должны быть сформированы следующие компетенции:

01.03.04 Прикладная математика (направленность : Интеллектуальный анализ данных, программирование и искусственный интеллект)

ПК.4 Способен решать профессиональные задачи, возникающие в области искусственного интеллекта

Индикаторы

ПК.4.3 Разрабатывает системы анализа больших данных

4. Объем и содержание дисциплины

Направления подготовки	01.03.04 Прикладная математика (направленность: Интеллектуальный анализ данных, программирование и искусственный интеллект)
форма обучения	очная
№№ семестров, выделенных для изучения дисциплины	4
Объем дисциплины (з.е.)	3
Объем дисциплины (ак.час.)	108
Контактная работа с преподавателем (ак.час.), в том числе:	42
Проведение лекционных занятий	14
Проведение практических занятий, семинаров	28
Самостоятельная работа (ак.час.)	66
Формы текущего контроля	Итоговое контрольное мероприятие (1) Письменное контрольное мероприятие (2)
Формы промежуточной аттестации	Зачет (4 семестр)

5. Аннотированное описание содержания разделов и тем дисциплины

Введение в курс

Общая информация о структуре курса, изучаемых технологиях и методах

Поколения платформ данных

Три поколения платформ Больших данных: Хранилища данных, Озера данных, Поточковые системы. Основные принципы их построения и отличия друг от друга

Распределенная обработка данных

Основные понятия распределенных систем. CAP теорема. Требования к распределенным системам со стороны задач обработки данных. Распределенное обучение на GPU. Концепция MapReduce

Федеративное обучение

Технология федеративного обучения. Назначения и возможности. Классификация системы федеративного обучения. Типы распределенных данных. Основные фреймворки федеративного обучения. Основные алгоритмы.

Хранение Больших данных

Структурированные, полуструктурированные и неструктурированные данные. ACID требования. Способы горизонтального масштабирования. SQL, noSQL и newSQL системы хранения

Обработка потоковых данных

Потоковые данные. Системы обработки потоковых сообщений. Лямбда архитектура для обработки потоковых данных. Гамма архитектура для обработки потоковых данных

Алгоритмы анализа Больших данных

Особенности алгоритмов обработки Больших данных. Принципы построения алгоритмов обработки Больших данных. Основные алгоритмы обработки Больших данных

6. Методические указания для обучающихся по освоению дисциплины

Освоение дисциплины требует систематического изучения всех тем в той последовательности, в какой они указаны в рабочей программе.

Основными видами учебной работы являются аудиторские занятия. Их цель - расширить базовые знания обучающихся по осваиваемой дисциплине и систему теоретических ориентиров для последующего более глубокого освоения программного материала в ходе самостоятельной работы. Обучающемуся важно помнить, что контактная работа с преподавателем эффективно помогает ему овладеть программным материалом благодаря расстановке необходимых акцентов и удержанию внимания интонационными модуляциями голоса, а также подключением аудио-визуального механизма восприятия информации.

Самостоятельная работа преследует следующие цели:

- закрепление и совершенствование теоретических знаний, полученных на лекционных занятиях;
- формирование навыков подготовки текстовой составляющей информации учебного и научного назначения для размещения в различных информационных системах;
- совершенствование навыков поиска научных публикаций и образовательных ресурсов, размещенных в сети Интернет;
- самоконтроль освоения программного материала.

Обучающемуся необходимо помнить, что результаты самостоятельной работы контролируются преподавателем во время проведения мероприятий текущего контроля и учитываются при промежуточной аттестации.

Обучающимся с ОВЗ и инвалидов предоставляется возможность выбора форм проведения мероприятий текущего контроля, альтернативных формам, предусмотренным рабочей программой дисциплины. Предусматривается возможность увеличения в пределах 1 академического часа времени, отводимого на выполнение контрольных мероприятий.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации.

При проведении текущего контроля применяются оценочные средства, обеспечивающие передачу информации, от обучающегося к преподавателю, с учетом психофизиологических особенностей здоровья обучающихся.

7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

При самостоятельной работе обучающимся следует использовать:

- конспекты лекций;
- литературу из перечня основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля);
- текст лекций на электронных носителях;
- ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины;
- лицензионное и свободно распространяемое программное обеспечение из перечня информационных технологий, используемых при осуществлении образовательного процесса по дисциплине;
- методические указания для обучающихся по освоению дисциплины.

8. Перечень основной и дополнительной учебной литературы

Основная:

1. Интеллектуальный анализ распределенных данных на базе облачных вычислений [Текст] / [М.С. Куприянов [и др.], 2011. 147 с.
2. Интеллектуальный анализ данных в распределенных системах [Текст] : [монография] / [М. С. Куприянов, И. И. Холод, З. А. Каршиев, И. А. Голубев], 2012. 108, [1] с.

Дополнительная:

1. Методы и модели анализа данных : OLAP и Data Mining [Текст] : учеб. пособие по специальности 071900 информ. системы и технологии” направления 654700 ”Информ. системы” / А. А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод, 2004. 336 с.
2. Технологии анализа данных [Текст] : Data Mining, Visual Mining, Text Mining, OLAP : учеб. пособие по специальности 071900 ”информац. системы и технологии” направления 654700 ”Информационные системы” / А.А. Барсегян [и др.], 2007. VIII, 375 с.

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

<http://www.psu.ru/elektronnye-resursy-dlya-psu> Электронные ресурсы для ПГНИУ

<http://window.edu.ru> Единое окно доступа к образовательным ресурсам

<http://www.mathnet.ru> Общероссийский математический портал

<http://www.psu.ru/elektronnye-resursy-dlya-psu> Электронные ресурсы для ПГНИУ

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Образовательный процесс по дисциплине **Большие данные** предполагает использование следующего программного обеспечения и информационных справочных систем:

Образовательный процесс по дисциплине предполагает использование следующего программного обеспечения и информационных справочных систем:

- доступ в режиме on-line в Электронную библиотечную систему (ЭБС);

- доступ в электронную информационно-образовательную среду университета.

Необходимое лицензионное и (или) свободно распространяемое программное обеспечение:

- приложение позволяющее просматривать и воспроизводить медиаконтент PDF-файлов «Adobe Acrobat Reader DC»;

- офисный пакет приложений «LibreOffice»;

При освоении материала и выполнении заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ (student.psu.ru).

При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).

система LMS Moodle (<http://e-learn.psu.ru/>), которая поддерживает возможность использования текстовых материалов и презентаций, аудио- и видеоконтент, а так же тесты, проверяемые задания, задания для совместной работы.

система тестирования Indigo (<https://indigotech.ru/>).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для лекционных занятий требуется аудитория, оснащенная: специализированной мебелью, презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для проведения семинарских (практических) занятий - Компьютерный класс, оборудованный: специализированной мебелью, меловой и (или) маркерной доской, компьютерами, проектором, экраном, специализированным программным обеспечением.

Для групповых (индивидуальных) консультаций, текущего контроля и промежуточной аттестации аудитория, оснащенная: специализированной мебелью, ноутбуком/компьютером, меловой (и) или маркерной доской, проектором, экраном.

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютера с доступом к локальной и глобальной компьютерным сетям.

2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.

5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет Libreoffice.

Справочно-правовая система «КонсультантПлюс»

**Фонды оценочных средств для аттестации по дисциплине
Большие данные**

**Планируемые результаты обучения по дисциплине для формирования компетенции.
Индикаторы и критерии их оценивания**

ПК.4

Способен решать профессиональные задачи, возникающие в области искусственного интеллекта

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
<p>ПК.4.3 Разрабатывает системы анализа больших данных</p>	<p>Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем больших данных</p>	<p align="center">Неудовлетворител теоретическое содержание курса не освоено, необходимые практически навыки и умения не сформированы, выполненные учебные задания содержат грубые ошибки, дополнительная самостоятельная работа над курсом не приведет к существенному повышению качества выполнения учебных заданий</p> <p align="center">Удовлетворительн теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки и умения работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий содержат ошибки</p> <p align="center">Хорошо теоретическое содержание курса освоено полностью, без пробелов, некоторые практические навыки и умения сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками</p>

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		<p style="text-align: center;">Отлично</p> <p>теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки и умения сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено количеством баллов, близким к максимальному</p>

Оценочные средства текущего контроля и промежуточной аттестации

Схема доставки : Базовая

Вид мероприятия промежуточной аттестации : Зачет

Способ проведения мероприятия промежуточной аттестации : Оценка по дисциплине в рамках промежуточной аттестации определяется на основе баллов, набранных обучающимся на контрольных мероприятиях, проводимых в течение учебного периода.

Максимальное количество баллов : 100

Конвертация баллов в отметки

«отлично» - от 81 до 100

«хорошо» - от 61 до 80

«удовлетворительно» - от 50 до 60

«неудовлетворительно» / «незачтено» менее 50 балла

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ПК.4.3 Разрабатывает системы анализа больших данных	Лабораторная работа 1. Сбор данных с удалённого ресурса. Письменное контрольное мероприятие	Знать математические методы решения основных задач анализа БД больших размеров;
ПК.4.3 Разрабатывает системы анализа больших данных	Лабораторная работа 1. Анализ данных с помощью БД Письменное контрольное мероприятие	Уметь проводить препроцессинг данных; анализировать возможности аналитических платформ для решения конкретных задач обработки данных
ПК.4.3 Разрабатывает системы анализа больших данных	Итоговое контрольное мероприятие. Зачет. Итоговое контрольное мероприятие	Проверяются знания основных понятий, методов разработки, изученных в курсе, и умение применять их на практике.

Спецификация мероприятий текущего контроля

Лабораторная работа 1. Сбор данных с удалённого ресурса.

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **13**

Показатели оценивания	Баллы
Работа выполнена полностью. Нет ошибок в программе. Возможно наличие небольшого отклонения от ожидаемого результата, не являющиеся следствием	25

незнания или непонимания учебного материала. Обучающийся показал полный объем знаний, умений в освоении, пройденных тем и применение их на практике.	
Работа выполнена полностью, но применён неэффективный метод решения. Допущена одна ошибка или два-три недочёта.	19
Допущены более одной ошибки или более двух-трех недочётов.	13
Работа выполнена неверно	менее 13

Лабораторная работа 1. Анализ данных с помощью БД

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **13**

Показатели оценивания	Баллы
Работа выполнена полностью. Нет ошибок в программе. Возможно наличие небольшого отклонения от ожидаемого результата, не являющиеся следствием незнания или непонимания учебного материала. Обучающийся показал полный объем знаний, умений в освоении, пройденных тем и применение их на практике.	25
Работа выполнена полностью, но применён неэффективный метод решения. Допущена одна ошибка или два-три недочёта.	19
Допущены более одной ошибки или более двух-трех недочётов.	13
Работа выполнена неверно	менее 13

Итоговое контрольное мероприятие. Зачет.

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **40**

Проходной балл: **17**

Показатели оценивания	Баллы
Вопрос 1 Способы установки количества потока в библиотеке OpenMP.	10
Вопрос 2 Статистические агрегатные функции в СУБД	10
Вопрос 3 Запрос в тестовую БД	10
Вопрос 4 Запрос в тестовую БД	10