

Федеральное государственное бюджетное образовательное учреждение
высшего образования
Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики

УТВЕРЖДАЮ
декан факультета вычислительной
математики и кибернетики


/И.А. Соколов /
«27» сентября 2022г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Современные методы распределенного хранения и обработки данных

Уровень высшего образования:

бакалавриат

Направление подготовки / специальность:

01.03.02 "Прикладная математика и информатика" (3++)

Направленность (профиль) ОПОП:

Искусственный интеллект и анализ данных

Форма обучения:

очная

Рабочая программа рассмотрена и утверждена
на заседании Ученого совета факультета ВМК
(протокол № 7 от 27 сентября 2022 года)

Москва 2022

Рабочая программа дисциплины (модуля) разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки 01.04.02 "Прикладная математика и информатика" утвержденного Приказом Министерства образования и науки РФ от 10 января 2018 г. N 13.

1. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОПОП ВО:

Настоящая дисциплина включена в учебный план по направлению 01.03.02 Прикладная математика и информатика, профиль Искусственный интеллект и анализ данных и входит в

Часть, формируемая участниками образовательных отношений (Дисциплины по выбору модуля "Математическое моделирование") Блока 1.

Дисциплина изучается на 3 курсе в 5 семестре.

2. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ:

Аналитика больших данных является одной из важных составляющих успеха таких титанов IT-индустрии, как Facebook, Amazon, Google, IBM, Microsoft. Данный курс посвящен изучению современных технологий анализа больших данных. Рассматриваются технологии и парадигмы OLAP, MapReduce, Hadoop, HDFS, YARN и пр.

3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ):

Планируемые результаты обучения по дисциплине (модулю)		
Содержание и код компетенции.	Индикатор (показатель) достижения компетенции	Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций
ОПК-1. Способен применять фундаментальные знания, полученные в области математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 – Обладает фундаментальными знаниями, полученными в области математических и (или) естественных наук ОПК-1.2 – Умеет использовать их в профессиональной деятельности ОПК-1.3 – Имеет навыки выбора методов решения задач профессиональной деятельности на основе теоретических знаний	Знать подходы, применяемые в индустрии для обработки и хранения больших массивов данных Уметь использовать современные инструменты, применяемые для работы с большими массивами данных Владеть навыками выбора инструментов для работы с большими массивами данных, релевантными практической задаче

4. Объем дисциплины (модуля) составляет 3 з.е., в том числе 54 академических часов, отведенных на контактную работу обучающихся с преподавателем, 54 академических часов на самостоятельную работу обучающихся.

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ), СТРУКТУРИРОВАННОЕ ПО ТЕМАМ (РАЗДЕЛАМ) С УКАЗАНИЕМ ОТВЕДЕННОГО НА НИХ КОЛИЧЕСТВА АКАДЕМИЧЕСКИХ ЧАСОВ И ВИДЫ УЧЕБНЫХ ЗАНЯТИЙ

5.1. Структура дисциплины (модуля) по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий (в строгом соответствии с учебным планом)

Наименование разделов и тем дисциплины (модуля), Форма промежуточной аттестации по дисциплине (модулю)	Номинальные трудозатраты обучающегося			Всего академических часов	Форма текущего контроля успеваемости* (наименование)
	Контактная работа (работа во взаимодействии с преподавателем) Виды контактной работы, академические часы		Самостоятельная работа обучающегося, академические часы		
	Занятия лекционного типа	Занятия семинарского типа			
Тема 1. Большие данные, аналитика и горизонтальная масштабируемость	4	2	6	12	решение задач
Тема 2. OLAP: основные понятия	2	2	4	8	решение задач
Тема 3. Введение и практические сценарии использования "больших данных"	2	2	4	8	решение задач
Тема 4. Введение в GFS и GoogleMapReduce	2	2	4	8	решение задач
Тема 5. Языки программирования для обработки больших данных	2	2	4	8	решение задач
Тема 6. Обзор средств Azure для обработки больших данных	2	2	4	8	решение задач
Тема 7. Примеры алгоритмов и теория сложности MapReduce	2	2	4	8	решение задач
Тема 8. HDFS и основы HadoopJava API	2	2	4	8	решение задач
Тема 9. HadoopMapReduce: управление и планирование вычислений	4	2	6	12	решение задач
Тема 10. HadoopStreaming	2	2	4	8	решение задач
Тема 11. Введение в ApacheSpark	2	2	4	8	решение задач
Тема 12. Системы хранения на базе HDFS: Pig, Hive, HBase, BigTable	4	2	6	12	решение задач
Промежуточная аттестация					экзамен

(экзамен)					
Итого	30	24	54	108	—

5.2. Содержание разделов (тем) дисциплины

№ п/п	Наименование разделов (тем) дисциплины	Содержание разделов (тем) дисциплин
1.	Тема 1. Большие данные, аналитика и горизонтальная масштабируемость	Большие данные, аналитика и горизонтальная масштабируемость
2.	Тема 2. OLAP: основные понятия	OLAP: основные понятия
3.	Тема 3. Введение и практические сценарии использования "больших данных"	Введение и практические сценарии использования "больших данных"
4.	Тема 4. Введение в GFS и GoogleMapReduce	Введение в GFS и GoogleMapReduce
5.	Тема 5. Языки программирования для обработки больших данных	Языки программирования для обработки больших данных
6.	Тема 6. Обзор средств Azure для обработки больших данных	Обзор средств Azure для обработки больших данных
7.	Тема 7. Примеры алгоритмов и теория сложности MapReduce	Примеры алгоритмов и теория сложности MapReduce
8.	Тема 8. HDFS и основы HadoopJava API	HDFS и основы HadoopJava API
9.	Тема 9. HadoopMapReduce: управление и планирование вычислений	HadoopMapReduce: управление и планирование вычислений
10.	Тема 10. HadoopStreaming	HadoopStreaming
11.	Тема 11. Введение в ApacheSpark	Введение в ApacheSpark
12.	Тема 12. Системы хранения на базе HDFS: Pig, Hive, HBase, BigTable	Системы хранения на базе HDFS: Pig, Hive, HBase, BigTable

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ (ФОС, ОЦЕНОЧНЫЕ И МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ) ДЛЯ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ).

Фонд оценочных средств приведен в отдельном документе

7. РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ:

7.1. Перечень основной и дополнительной литературы

Основная литература

1. Юре Лесковец, Ананд Раджараман, Джеффри Д. Ульман. Анализ больших наборов данных. — ДМК Пресс, 2016 — 498 с.
2. ЧакЛэм. Hadoop в действии. — ДМК Пресс, 2012 — 424 с.
3. Карау Х. и др. Изучаем Spark: молниеносный анализ данных //ДМК Пресс, М. – 2015.

Дополнительная литература

1. Том Уайт. Hadoop. Подробное руководство. — Питер, 2013 — 672 с.
2. SammerE. Hadoop operations. – " O'Reilly Media, Inc.", 2012.
3. Karau H. et al. Learning spark: lightning-fast big data analysis. – " O'Reilly Media, Inc.", 2015.
4. Pentreath N. Machine learning with spark. – Packt Publishing Ltd, 2015.
5. Karau H. Fast data processing with spark. – Packt Publishing Ltd, 2013.
6. MateiZaharia at al. Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing. NSDI (2012)
<http://usenix.org/system/files/conference/nsdi12/nsdi12-final138.pdf>
7. Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung “The Google File System”, 2003 <http://research.google.com/archive/gfs.html>
8. Jeffrey Dean and Sanjay Ghemawat “MapReduce: Simplified Data Processing on Large Clusters”, 2004 <http://research.google.com/archive/mapreduce.html>

7.2.Перечень лицензионного программного обеспечения, в том числе отечественного производства

При реализации дисциплины может быть использовано следующее программное обеспечение:

- Операционная система Windows
- Операционная система Debian Linux
- Программное обеспечение для подготовки слайдов лекций MS PowerPoint, MS Word
- Программное обеспечение для создания и просмотра pdf-документов Adobe Reader
- Издательская система LaTeX
- Язык программирования Python и среда разработки Jupiter Notebook (вместе с библиотеками numpy, scikit-learn, pandas)
- Язык программирования R и среда разработки R Studio
- Файловый архиватор 7z. Свободно-распространяемое ПО
- Браузеры Google Chrome, Mozilla Firefox. Свободно-распространяемое ПО
- Офисный пакет LibreOffice. Свободно-распространяемое ПО
- Visual Studio Community Интегрированная среда разработки ПО. Свободно-распространяемое ПО
- PyCharm Community Интегрированная среда разработки ПО. Свободно-распространяемое

ПО

- Anaconda Интегрированная среда разработки ПО. Свободно-распространяемое ПО

7.3. Перечень профессиональных баз данных и информационных справочных систем

1. <http://www.edu.ru> – портал Министерства образования и науки РФ
2. <http://www.ict.edu.ru> – система федеральных образовательных порталов «ИКТ в образовании»
3. <http://www.openet.ru> - Российский портал открытого образования
4. <http://www.mon.gov.ru> - Министерство образования и науки Российской Федерации
5. <http://www.fasi.gov.ru> - Федеральное агентство по науке и инновациям

7.4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Math-Net.Ru [Электронный ресурс] : общероссийский математический портал / Математический институт им. В. А. Стеклова РАН ; Российская академия наук, Отделение математических наук. - М. : [б. и.], 2010. - Загл. с титул. экрана. - Б. ц.
URL: <http://www.mathnet.ru>
2. Университетская библиотека Online [Электронный ресурс] : электронная библиотечная система / ООО "Директ-Медиа" . - М. : [б. и.], 2001. - Загл. с титул. экрана. - Б. ц. URL: www.biblioclub.ru
3. Универсальные базы данных EastView [Электронный ресурс] : информационный ресурс / EastViewInformationServices. - М. : [б. и.], 2012. - Загл. с титул. экрана. - Б. ц.
URL: www.ebiblioteka.ru
4. Научная электронная библиотека eLIBRARY.RU [Электронный ресурс] : информационный портал / ООО "РУНЭБ" ; Санкт-Петербургский государственный университет. - М. : [б. и.], 2005. - Загл. с титул. экрана. - Б. ц.
URL: www.eLibrary.ru

7.5. Описание материально-технического обеспечения.

Образовательная организация, ответственная за реализацию данной Программы, располагает соответствующей материально-технической базой, включая современную вычислительную технику, объединенную в локальную вычислительную сеть, имеющую выход в Интернет. Используются специализированные компьютерные классы, оснащенные современным оборудованием. Материальная база соответствует действующим санитарно-техническим нормам и обеспечивает проведение всех видов занятий (лекционных, практических, семинарских, лабораторных, дисциплинарной и междисциплинарной подготовки) и научно-исследовательской работы обучающихся, предусмотренных учебным планом.

8. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ

8.1. Формы и методы преподавания дисциплины

Используемые формы и методы обучения: лекции и лабораторные работы, самостоятельная работа студентов.

В процессе преподавания дисциплины преподаватель использует как классические формы и методы обучения (лекции и практические занятия), так и активные методы обучения.

При проведении лекционных занятий преподаватель использует аудиовизуальные, компьютерные и мультимедийные средства обучения, а также демонстрационные и наглядно-иллюстрационные (в том числе раздаточные) материалы.

Семинарские (практические) занятия по данной дисциплине проводятся с использованием компьютерного и мультимедийного оборудования, при необходимости - с привлечением полезных Интернет-ресурсов и пакетов прикладных программ.

8.2. Методические рекомендации преподавателю

Перед началом изучения дисциплины преподаватель должен ознакомить студентов с видами учебной и самостоятельной работы, перечнем литературы и интернет-ресурсов, формами

текущей и промежуточной аттестации, с критериями оценки качества знаний для итоговой оценки по дисциплине.

При проведении лекций, преподаватель:

- 1) формулирует тему и цель занятия;
- 2) излагает основные теоретические положения;
- 3) с помощью мультимедийного оборудования и/или под запись дает определения основных понятий, расчетных формул;
- 4) проводит примеры из отечественного и зарубежного опыта, дает текущие статистические данные для наглядного и образного представления изучаемого материала;
- 5) в конце занятия дает вопросы для самостоятельного изучения.

Во время выполнения заданий в учебной аудитории студент может консультироваться с преподавателем, определять наиболее эффективные методы решения поставленных задач. Если какая-то часть задания остается невыполненной, студент может продолжить её выполнение во время внеаудиторной самостоятельной работы.

Перед выполнением внеаудиторной самостоятельной работы преподаватель проводит инструктаж (консультацию) с определением цели задания, его содержания, сроков выполнения, основных требований к результатам работы, критериев оценки, форм контроля и перечня источников и литературы.

Для оценки полученных знаний и освоения учебного материала по каждому разделу и в целом по дисциплине преподаватель использует формы текущего, промежуточного и итогового контроля знаний обучающихся.

Для семинарских занятий

Подготовка к проведению занятий проводится регулярно. Организация преподавателем семинарских занятий должна удовлетворять следующим требованиям: количество занятий должно соответствовать учебному плану программы, содержание планов должно соответствовать программе, план занятий должен содержать перечень рассматриваемых вопросов.

Во время семинарских занятий используются словесные методы обучения, как беседа и дискуссия, что позволяет вовлекать в учебный процесс всех слушателей и стимулирует творческий потенциал обучающихся.

При подготовке семинарскому занятию преподавателю необходимо знать план его проведения, продумать формулировки и содержание учебных вопросов, выносимых на обсуждение.

В начале занятия преподаватель должен раскрыть теоретическую и практическую значимость темы занятия, определить порядок его проведения, время на обсуждение каждого учебного вопроса. В ходе занятия следует дать возможность выступить всем желающим и предложить выступить тем слушателям, которые проявляют пассивность.

Целесообразно, в ходе обсуждения учебных вопросов, задавать выступающим и аудитории дополнительные и уточняющие вопросы с целью выяснения их позиций по существу обсуждаемых проблем, а также поощрять выступление с места в виде кратких дополнений. На занятиях проводится отработка практических умений под контролем преподавателя

Для практических занятий

Подготовка преподавателя к проведению практического занятия начинается с изучения исходной документации и заканчивается оформлением плана проведения занятия.

На основе изучения исходной документации у преподавателя должно сложиться представление о целях и задачах практического занятия и о том объеме работ, который должен выполнить каждый обучающийся. Далее можно приступить к разработке содержания практического занятия. Для этого преподавателю (даже если он сам читает лекции по этому курсу) целесообразно вновь просмотреть содержание лекции с точки зрения предстоящего практического занятия. Необходимо выделить понятия, положения, закономерности, которые следует еще раз проиллюстрировать на конкретных задачах и упражнениях. Таким образом, производится отбор содержания, подлежащего усвоению.

Важнейшим элементом практического занятия является учебная задача (проблема), предлагаемая для решения. Преподаватель, подбирая примеры (задачи и логические задания) для практического занятия, должен представлять дидактическую цель: привитие каких навыков и умений применительно к каждой задаче установить, каких усилий от обучающихся она потребует, в чем должно проявиться творчество студентов при решении данной задачи.

Преподаватель должен проводить занятие так, чтобы на всем его протяжении студенты были заняты напряженной творческой работой, поисками правильных и точных решений, чтобы каждый получил возможность раскрыться, проявить свои способности. Поэтому при планировании занятия и разработке индивидуальных заданий преподавателю важно учитывать подготовку и интересы каждого студента. Педагог в этом случае выступает в роли консультанта, способного вовремя оказать необходимую помощь, не подавляя самостоятельности и инициативы студента.

8.3. Методические рекомендации студентам по организации самостоятельной работы.

Приступая к изучению новой учебной дисциплины, студенты должны ознакомиться с учебной программой, учебной, научной и методической литературой, имеющейся в библиотеке университета, встретиться с преподавателем, ведущим дисциплину, получить в библиотеке рекомендованные учебники и учебно-методические пособия, осуществить запись на соответствующий курс в среде электронного обучения университета.

Глубина усвоения дисциплины зависит от активной и систематической работы студента на лекциях и практических занятиях, а также в ходе самостоятельной работы, по изучению рекомендованной литературы.

На лекциях важно сосредоточить внимание на ее содержании. Это поможет лучше воспринимать учебный материал и уяснить взаимосвязь проблем по всей дисциплине. Основное содержание лекции целесообразнее записывать в тетради в виде ключевых фраз, понятий, тезисов, обобщений, схем, опорных выводов. Необходимо обращать внимание на термины, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации. Желательно оставлять в конспектах поля, на которых делать пометки из рекомендованной литературы, дополняющей материал прослушанной лекции, а также подчеркивающие особую важность тех или иных теоретических положений. С целью уяснения теоретических положений, разрешения спорных ситуаций необходимо задавать преподавателю уточняющие вопросы. Для закрепления содержания лекции в памяти, необходимо во время самостоятельной работы внимательно прочесть свой конспект и дополнить его записями из учебников и рекомендованной литературы. Конспектирование читаемых лекций и их последующая доработка способствует более глубокому усвоению знаний, и поэтому являются важной формой учебной деятельности студентов.

Методические указания для обучающихся по подготовке к семинарским занятиям

Для того чтобы семинарские занятия приносили максимальную пользу, необходимо помнить, что упражнение и решение задач проводятся по вычитанному на лекциях материалу и связаны, как правило, с детальным разбором отдельных вопросов лекционного курса. Следует подчеркнуть, что только после усвоения лекционного материала с определенной точки зрения (а именно с той, с которой он излагается на лекциях) он будет закрепляться на семинарских занятиях как в результате обсуждения и анализа лекционного материала, так и с помощью решения проблемных ситуаций, задач.

При этих условиях студент не только хорошо усвоит материал, но и научится применять его на практике, а также получит дополнительный стимул (и это очень важно) для активной проработки лекции.

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала вычислений составить краткий план решения проблемы (задачи). Решение проблемных задач или примеров следует

излагать подробно, вычисления располагать в строгом порядке, отделяя вспомогательные вычисления от основных. Решения при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Следует помнить, что решение каждой учебной задачи должно доводиться до окончательного логического ответа, которого требует условие, и по возможности с выводом. Полученный ответ следует проверить способами, вытекающими из существа данной задачи. Полезно также (если возможно) решать несколькими способами и сравнить полученные результаты. Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

При подготовке к семинарским занятиям следует использовать основную литературу из представленного списка, а также руководствоваться приведенными указаниями и рекомендациями. Для наиболее глубокого освоения дисциплины рекомендуется изучать литературу, обозначенную как «дополнительная» в представленном списке.

Методические указания для обучающихся по подготовке к практическим занятиям

Целью практических занятий по данной дисциплине является закрепление теоретических знаний, полученных при изучении дисциплины.

При подготовке к практическому занятию целесообразно выполнить следующие рекомендации: изучить основную литературу; ознакомиться с дополнительной литературой, новыми публикациями в периодических изданиях: журналах, газетах и т. д.; при необходимости доработать конспект лекций. При этом учесть рекомендации преподавателя и требования учебной программы.

При выполнении практических занятий основным методом обучения является самостоятельная работа студента под управлением преподавателя. На них пополняются теоретические знания студентов, их умение творчески мыслить, анализировать, обобщать изученный материал, проверяется отношение студентов к будущей профессиональной деятельности.

Оценка выполненной работы осуществляется преподавателем комплексно: по результатам выполнения заданий, устному сообщению и оформлению работы. После подведения итогов занятия студент обязан устранить недостатки, отмеченные преподавателем при оценке его работы.

Методические указания для самостоятельной работы обучающихся

Прочное усвоение и долговременное закрепление учебного материала невозможно без продуманной самостоятельной работы. Такая работа требует от студента значительных усилий, творчества и высокой организованности. В ходе самостоятельной работы студенты выполняют следующие задачи: дорабатывают лекции, изучают рекомендованную литературу, готовятся к практическим занятиям, к коллоквиуму, контрольным работам по отдельным темам дисциплины. При этом эффективность учебной деятельности студента во многом зависит от того, как он распорядился выделенным для самостоятельной работы бюджетом времени.

Результатом самостоятельной работы является прочное усвоение материалов по предмету согласно программы дисциплины. В итоге этой работы формируются профессиональные умения и компетенции, развивается творческий подход к решению возникших в ходе учебной деятельности проблемных задач, появляется самостоятельности мышления.

Решение задач

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала вычислений составить краткий план решения проблемы (задачи).

Решение проблемных задач или примеров следует излагать подробно, вычисления располагать в строгом порядке, отделяя вспомогательные вычисления от основных. Решения при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Следует помнить, что решение каждой учебной задачи должно доводиться до окончательного логического ответа, которого требует условие, и по возможности с выводом.

Полученный ответ следует проверить способами, вытекающими из существа данной задачи. Полезно также (если возможно) решать несколькими способами и сравнить полученные результаты.

Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

Задача — это цель, заданная в определенных условиях, решение задачи — процесс достижения поставленной цели, поиск необходимых для этого средств.

Алгоритм решения задач:

1. Внимательно прочитайте условие задания и уясните основной вопрос, представьте процессы и явления, описанные в условии.

2. Повторно прочтите условие для того, чтобы чётко представить основной вопрос, проблему, цель решения, заданные величины, опираясь на которые можно вести поиски решения.

3. Произведите краткую запись условия задания.

4. Если необходимо составьте таблицу, схему, рисунок или чертёж.

5. Определите метод решения задания, составьте план решения.

6. Запишите основные понятия, формулы, описывающие процессы, предложенные заданной системой.

7. Найдите решение в общем виде, выразив искомые величины через заданные.

9. Проверьте правильность решения задания.

10. Произведите оценку реальности полученного решения.

11. Запишите ответ.