

**МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ**  
**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ**  
**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**  
**«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»**  
**(РУТ (МИИТ))**



Рабочая программа дисциплины (модуля),  
как компонент образовательной программы  
высшего образования - программы бакалавриата  
по направлению подготовки  
09.03.01 Информатика и вычислительная техника,  
утвержденной первым проректором РУТ (МИИТ)  
Тимониным В.С.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

**Методы анализа и обработки больших данных**

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Направленность (профиль): IT-сервисы и технологии обработки данных на транспорте

Форма обучения: Очная

Рабочая программа дисциплины (модуля) в виде электронного документа выгружена из единой корпоративной информационной системы управления университетом и соответствует оригиналу

Простая электронная подпись, выданная РУТ (МИИТ)  
ID подписи: 937226  
Подписал: руководитель образовательной программы  
Проневич Ольга Борисовна  
Дата: 14.06.2024

## 1. Общие сведения о дисциплине (модуле).

Целью освоение дисциплины (модуля) является приобретения практически навыков и умений, позволяющих создавать высокопроизводительные реализации известных методов вычислительной математики, анализа и обработки больших данных.

Задачами освоения дисциплины (модуля) являются:

освоение базовых знаний в области архитектуры современных многопроцессорных вычислительных систем параллельной обработки информации,

освоений технологий организации параллельных вычислений на многопроцессорных вычислительных комплексах с распределенной или общей оперативной памятью.

## 2. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

**ПК-1** - Способен анализировать большие данные с использованием существующей в организации методологической и технологической инфраструктуры;

**ПК-7** - Способен к организации процессов разработки программного обеспечения .

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

### **Знать:**

программное обеспечение, необходимое для работы с большими данными,

необходимые элементы инфраструктуры для обработки больших данных.

### **Уметь:**

Обрабатывать данные в Hadoop

Работать с распределенными файловыми системами в Hadoop

Настраивать кластеры в Hadoop

Составлять спецификации оборудования для различных задач

### **Владеть:**

Инструментами настройки сетей

Знаниями о высокопроизводительных вычислениях

Навыками работы со статистическими параметрами вычислений  
Технологиями параллельной обработки данных  
Навыками работы с vCenter Server  
Инструментами работы с Hadoop  
Техниками настройки кластеров в Hadoop

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 3 з.е. (108 академических часа(ов)).

3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Тип учебных занятий	Количество часов	
	Всего	Семестр №5
Контактная работа при проведении учебных занятий (всего):	64	64
В том числе:		
Занятия лекционного типа	32	32
Занятия семинарского типа	32	32

3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 44 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

4. Содержание дисциплины (модуля).

4.1. Занятия лекционного типа.

№ п/п	Тематика лекционных занятий / краткое содержание
1	<p>Тема 1. Большие данные. История развития методов анализа и обработки</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- эволюция понятия больших данных</li> <li>- хранилища данных</li> <li>- история развития методов обработки больших данных</li> <li>- отличия методов анализа от методов обработки больших данных</li> </ul>
2	<p>Тема 2. Жизненный цикл данных и метаданные</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- элементы жизненного цикла и методы управления жизненным циклом</li> <li>- метаданные</li> </ul>
3	<p>Тема 3. Архитектура систем обработки больших данных</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- прием данных</li> <li>- сбор данных</li> <li>- анализ данных</li> <li>- представления результатов</li> <li>- витрины данных</li> </ul>
4	<p>Тема 4. Введение в понятия высокопроизводительных вычислений.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Важность проблематики параллельных вычислений</li> <li>- Пути достижения параллелизма. Векторная и конвейерная обработка данных.</li> <li>- Многопроцессорная и многомашинная, параллельная обработка данных.</li> <li>- Закон Мура, сдерживающие факторы наращивания количества транзисторов на кристалле и частоты процессоров. Привлекательность подхода параллельной обработки данных</li> </ul>
5	<p>Тема 5. Основные направления развития высокопроизводительных компьютеров.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Сдерживающие факторы повсеместного внедрения параллельных вычислений</li> <li>- Ведомственные, национальные и другие программы, направленные на развитие параллельных вычислений в России.</li> <li>- Необходимость изучения дисциплины параллельного программирования. Перечень критических задач, решение которых без использования параллельных вычислений затруднено или вовсе невозможно.</li> </ul>
6	<p>Тема 6. Задачи параллельной обработки данных</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- программные структуры алгоритмов параллельной обработки данных</li> <li>- инструменты параллельной обработки данных</li> <li>- конвейерно-параллельной обработки интегрированных потоков данных</li> </ul>
7	<p>Тема 7. Классификация вычислительных сетей</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Системы с распределенной, общей памятью, примеры систем.</li> <li>- Массивно-параллельные системы (MPP). Симметричные мультипроцессорные системы (SMP). - Параллельные векторные системы (PVP).</li> <li>- Системы с неоднородным доступом к памяти (Numa)</li> <li>- Компьютерные кластеры – специализированные и полнофункциональные. История возникновения компьютерных кластеров – проект Beowulf. Метакомпьютинг. Классификация Флинна, Шора и т.д.</li> <li>- Организация межпроцессорных связей – коммуникационные топологии.</li> <li>- Примеры сетевых решений для создания кластерных систем</li> </ul>
8	<p>Тема 8. Основные принципы организации параллельной обработки данных: модели, методы и технологии параллельного программирования</p> <p>Рассматриваемые вопросы:</p>

№ п/п	Тематика лекционных занятий / краткое содержание
	<p>-Функциональный параллелизм, параллелизм по данным.</p> <p>-Парадигма master-slave. Парадигма SPMD. Парадигма конвейеризации. Парадигма «разделяй и властвуй». Спекулятивный параллелизм. Важность выбора технологии для реализации алгоритма</p> <p>-Модель обмена сообщениями – MPI.</p> <p>-Модель общей памяти – OpenMP. Концепция виртуальной, разделяемой памяти – Linda.</p> <p>Российские разработки – Т-система, система DVM. Проблемы создания средства автоматического распараллеливания программ</p>
9	<p>Тема 9. Параллельное программирование с использованием интерфейса передачи сообщений MPI</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>-Библиотека MPI.</li> <li>- Модель SIMD. Инициализация и завершение MPI-приложения.</li> </ul> <p>Точечные обмены данными между процессами MPI-программы. Режимы буферизации.</p> <p>Проблема deadlock'ов. Коллективные взаимодействия процессов в MPI. Управление группами и коммутаторами в MPI</p>
10	<p>Тема 10. Параллельное программирование на системах с общей памятью (OpenMP)</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>-Введение в OpenMP</li> <li>-Стандарты программирования для систем с разделяемой памятью.</li> <li>-Создание многопоточных приложений. Использование многопоточности при программировании для многоядерных платформ.</li> <li>-Синхронизация данных между ветвями в параллельной программе. Директивы языка OpenMP</li> </ul>
11	<p>Тема 11. Параллельное программирование многоядерных GPU. Кластеры из GPU и суперкомпьютеры на гибридной схеме</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>-Существующие многоядерные системы.</li> <li>-GPU.</li> <li>-Использование OpenMP и MPI технологий совместно с CUDA.</li> <li>-Степень параллелизма численного алгоритма. Закон Амдала. Параллельный алгоритм решения СЛАУ</li> </ul>
12	<p>Тема 12. Программные платформы и системы для больших данных</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- системы управления потоками данных</li> <li>- системы хранения больших данных</li> <li>- платформы больших данных</li> <li>- обработка больших данных в реальном времени</li> </ul>

#### 4.2. Занятия семинарского типа.

##### Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	<p>Тема 1. Изучение технологий Hadoop и MapReduce</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Hadoop</li> <li>- MapReduce</li> </ul>
2	<p>Тема 2. Анализ больших массивов данных инструментами python</p> <p>Рассматриваемые вопросы:</p>

№ п/п	Тематика практических занятий/краткое содержание
	- библиотеки анализа и загрузки больших данных - фреймворки Python с параллельной обработкой данных
3	Тема 3. Высокопроизводительные вычисления с фреймворком Apache Spark Рассматриваемые вопросы: - знакомство фреймворком - знакомство с архитектурой системы - реализации концепции MapReduce - разработки программ с использованием Apache Spark
4	Тема 4. Введение в высокопроизводительные серверы на python Рассматриваемые вопросы: - работа с распределенными системами - создание кода для работы на GRU
5	Тема 5. Параллельное программирование. Параллельная обработка данных Рассматриваемые вопросы: 1. Параллельное программирование 2. Параллельная обработка данных
6	Тема 6. Работа с vCenter Server Рассматриваемые вопросы: 1. Понятие vCenter 2. Установка и развертывание vCenter
7	Тема 7. Развертывание среды для обработки данных при помощи Hadoop Рассматриваемые вопросы: 1. Обработка данных в Hadoop 2. Распределенная файловая система Hadoop 3. Разработка приложений MapReduce 4. Настройка кластера Hadoop
8	Тема 8. Составление спецификаций оборудования для работы с высокопроизводительными вычислениями Рассматриваемые вопросы: 1. Составление спецификаций оборудования для работы с высокопроизводительными вычислениями

#### 4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Работа с учебной литературой
2	Участие в онлайн-конференциях и мастер-классах
3	Поиск алгоритмов обработки данных в открытых источниках
4	Подготовка к промежуточной аттестации.
5	Подготовка к текущему контролю.
6	Подготовка к промежуточной аттестации.
7	Подготовка к текущему контролю.

#### 4.4. Примерный перечень тем курсовых работ

1. Робот по сбору и передаче данных о лицензиях по разработке, производству, испытанию и ремонту авиационной техники.
2. Робот по сбору и передаче данных о лицензиях по перевозкам железнодорожным транспортом опасных грузов.
3. Робот по сбору и передаче данных о лицензиях по перевозкам железнодорожным транспортом пассажиров.
4. Робот по сбору и передаче данных о лицензиях по погрузочно-разгрузочной деятельности применительно к опасным грузам на железнодорожном транспорте.
5. Робот по сбору и передаче данных о лицензиях по перевозкам внутренним водным транспортом, морским транспортом опасных грузов.
6. Робот по сбору и передаче данных о лицензиях по перевозкам внутренним водным транспортом, морским транспортом пассажиров.
7. Робот по сбору и передаче данных о лицензиях по погрузочно-разгрузочной деятельности применительно к опасным грузам на внутреннем водном транспорте, в морских портах.
8. Робот по сбору и передаче данных о лицензиях по осуществлению буксировок морским транспортом.
9. Робот по сбору и передаче данных о лицензиях на деятельность по перевозкам пассажиров и иных лиц автобусами.
10. Робот по сбору и передаче данных из реестра российских перевозчиков, допущенных к осуществлению международных автомобильных перевозок.
11. Робот по сбору и передаче данных из реестра транспортных средств, включенных в реестр лицензий на деятельность по перевозкам пассажиров и иных лиц автобусами.
12. Робот по сбору и передаче данных из реестра удостоверений об утверждении курсов подготовки водителей автотранспортных средств, перевозящих опасные грузы.
13. Робот по сбору и передаче данных о лицензиях на образовательную деятельность.
14. Робот по сбору и передаче данных о лицензиях на медицинскую деятельность
15. Робот по сбору и передаче данных о лицензиях на фармацевтическую деятельность
16. Робот по сбору и передаче данных о лицензиях на техническое обслуживание медицинских изделий

17. Робот по сбору и передаче данных о лицензиях на производство биомедицинских клеточных продуктов

18. Робот по сбору и передаче данных о лицензиях по обработке (переработке) лома и отходов драгоценных металлов

19. Робот по сбору и передаче данных о лицензиях по скупке у физических лиц ювелирных и других изделий из драгоценных металлов и драгоценных камней лома таких изделий

20. Робот по сбору и передаче данных о лицензиях на деятельность по технической защите конфиденциальной информации

21. Робот по сбору и передаче данных о лицензиях на деятельность по разработке и производству средств защиты конфиденциальной информации

22. Робот по сбору и передаче данных из реестра сертифицированных средств защиты информации

23. Робот по сбору и передаче данных из реестра аккредитованных органов по сертификации

24. Робот по сбору и передаче данных из реестра аккредитованных испытательных лабораторий

25. Робот по сбору и передаче данных из реестра промышленной продукции, произведенной на территории Российской Федерации

26. Робот по сбору и передаче данных о лицензиях на деятельность по тушению пожаров в населенных пунктах, на производственных объектах и объектах инфраструктуры, по тушению лесных пожаров

27. Робот по сбору и передаче данных о лицензиях на деятельность по монтажу, техническому обслуживанию и ремонту средств обеспечения пожарной безопасности зданий и сооружений

28. Робот по сбору и передаче данных о лицензиях на космическую деятельность

29. Робот по сбору и передаче данных о лицензиях по управлению многоквартирными домами

30. Робот по сбору и передаче данных из реестра субъектов страхового дела

31. Робот по сбору и передаче данных из реестра саморегулируемых организаций в сфере финансового рынка

32. Робот по сбору и передаче данных из реестра лицензий на осуществление деятельности в области гидрометеорологии и смежных с ней областях

33. Робот по сбору и передаче данных из реестра лицензий на осуществление работ по активному воздействию на гидрометеорологические и геофизические процессы и явления



34. Сравнение цен товаров в 2 интернет-магазинах (М.Видео и ДНС)

35. Произвольная тема по выбору студента (с обязательным предварительным согласованием)

5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

№ п/п	Библиографическое описание	Место доступа
1	Глубокое обучение Гудфеллоу Я., Бенджио И., Курвилль А.	<a href="https://e.lanbook.com/book/107901">https://e.lanbook.com/book/107901</a>
2	Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных Флах П.	<a href="https://e.lanbook.com/book/69955">https://e.lanbook.com/book/69955</a>
3	Математические методы распознавания образов Местецкий Л.М	<a href="https://e.lanbook.com/book/100634">https://e.lanbook.com/book/100634</a>
4	Габдуллин, Н. М. Развитие человеческого капитала и цифровой экономики в регионах России: факторный и кластерный анализ : монография / Н. М. Габдуллин. — Казань : КФУ, 2019. — 268 с. — ISBN 978-5-00130-291-9. — Текст : электронный // Лань : электронно-библиотечная система	<a href="https://e.lanbook.com/book/173018">https://e.lanbook.com/book/173018</a>
5	Гитис, Л. Х. Статистическая классификация и кластерный анализ / Л. Х. Гитис. — Москва : Горная книга, 2003. — 157 с. — ISBN 5-7418-0010-6. — Текст : электронный // Лань : электронно-библиотечная система.	<a href="https://e.lanbook.com/book/3493">https://e.lanbook.com/book/3493</a>
6	Гласснер, Э. Глубокое обучение без математики. Том 2. Практика : руководство / Э. Гласснер ; перевод с английского В. А. Яроцкого. — Москва : ДМК Пресс, 2020. — 610 с. — ISBN 978-5-97060-767-1. — Текст : электронный // Лань : электронно-библиотечная система	<a href="https://e.lanbook.com/book/131710">https://e.lanbook.com/book/131710</a>
7	Кук, Д. Машинное обучение с использованием библиотеки H2O / Д. Кук ; перевод с английского А. Б. Огурцова. — Москва : ДМК Пресс, 2018. — 250 с. — ISBN 978-5-97060-508-0. — Текст : электронный // Лань : электронно-библиотечная система.	<a href="https://e.lanbook.com/book/97353">https://e.lanbook.com/book/97353</a>
8	Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают	<a href="https://e.lanbook.com/book/69955">https://e.lanbook.com/book/69955</a>

	знания из данных / П. Флах. — Москва : ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060-273-7. — Текст : электронный // Лань : электронно-библиотечная система.	
9	Целых, А. Н. Современные технологии противодействия финансовым преступлениям : учебное пособие / А. Н. Целых. — Ростов-на-Дону : ЮФУ, 2019. — 119 с. — ISBN 978-5-9275-3286-5. — Текст : электронный // Лань : электронно-библиотечная система	<a href="https://e.lanbook.com/book/141063">https://e.lanbook.com/book/141063</a>
10	Шалев-Шварц, Ш. Идеи машинного обучения : учебное пособие / Ш. Шалев-Шварц, Бен-Давид Ш. ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2019. — 436 с. — ISBN 978-5-97060-673-5. — Текст : электронный // Лань : электронно-библиотечная система.	<a href="https://e.lanbook.com/book/131686">https://e.lanbook.com/book/131686</a>
11	Габдуллин, Н. М. Развитие человеческого капитала и цифровой экономики в регионах России: факторный и кластерный анализ : монография / Н. М. Габдуллин. — Казань : КФУ, 2019. — 268 с. — ISBN 978-5-00130-291-9	<a href="https://e.lanbook.com/book/173018">https://e.lanbook.com/book/173018</a>
12	Гласнер, Э. Глубокое обучение без математики. Том 2. Практика : руководство / Э. Гласнер ; перевод с английского В. А. Яроцкого. — Москва : ДМК Пресс, 2020. — 610 с. — ISBN 978-5-97060-767-1	<a href="https://e.lanbook.com/book/131710">https://e.lanbook.com/book/131710</a>
13	Кук, Д. Машинное обучение с использованием библиотеки H2O / Д. Кук ; перевод с английского А. Б. Огурцова. — Москва : ДМК Пресс, 2018. — 250 с. — ISBN 978-5-97060-508-0	<a href="https://e.lanbook.com/book/97353">https://e.lanbook.com/book/97353</a>
14	Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. — Москва : ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060-273-7	<a href="https://e.lanbook.com/book/69955">https://e.lanbook.com/book/69955</a>
15	Шалев-Шварц, Ш. Идеи машинного обучения : учебное пособие / Ш. Шалев-Шварц, Бен-Давид Ш. ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2019. — 436 с. — ISBN 978-5-97060-673-5	<a href="https://e.lanbook.com/book/131686">https://e.lanbook.com/book/131686</a>
16	Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль ; перевод с английского А. А. Слинкина. — 2-е изд. — Москва : ДМК Пресс, 2018. — 652 с. — ISBN 978-5-97060-618-6	<a href="https://e.lanbook.com/book/107901">https://e.lanbook.com/book/107901</a>

6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

<https://habr.com/ru> - база знаний в виде статей, обзоров

<https://journal.tinkoff.ru/short/ai-for-all/> - база данных нейронных сетей

<https://vc.ru/services/916617-luchshie-neyroseti-bolshaya-podborka-iz-top-200-ii-generatorov-po-kategoriyam> - база данных нейронных сетей

<https://github.com/abalmumcu/bert-rest-api> - профессиональная платформа для командой работы над проектов (нейронная сеть bert)

<http://library.miit.ru/> - электронно-библиотечная система Научно-технической библиотеки МИИТ

<https://proglib.io/p/raspoznavanie-obektov-s-pomoshchyu-yolo-v3-na-tensorflow-2-0-2020-11-08> - профессиональная библиотека программистов

[https://yandex.cloud/ru/blog/posts/2022/12/andrey-berger-and-yandex-cloud?utm\\_referrer=https%3A%2F%2Fyandex.ru%2F](https://yandex.cloud/ru/blog/posts/2022/12/andrey-berger-and-yandex-cloud?utm_referrer=https%3A%2F%2Fyandex.ru%2F) – библиотека профессиональных статей разработчиков Яндекс

<https://yandex.cloud/ru/blog> - библиотека профессиональных статей разработчиков Яндекс

<https://tproger.ru/translations/opencv-python-guide> - библиотека основных команд OpenCV

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

Microsoft Office 2010

VMware Workstation

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю).

1 учебный класс

Компьютер преподавателя

Компьютеры студентов

Монитор

Проектор

Экран для проектора

Маркерная доска

9. Форма промежуточной аттестации:

Зачет в 5 семестре.

#### 10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

И.В. Зенковский

Согласовано:

Директор

Б.В. Игольников

Руководитель образовательной  
программы

О.Б. Проневич

Председатель учебно-методической  
комиссии

Д.В. Паринов