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: 170737
Подписал: заместитель директора академии Паринов Денис Владимирович
Дата: 02.09.2021

1. Общие сведения о дисциплине (модуле).

Цель– дисциплина предназначена для того, чтобы дать знания, умения и основные навыки, позволяющие создавать высокопроизводительные реализации известных методов вычислительной математики, анализа и обработки данных. Целью освоения дисциплины является – освоение базовых знаний в области архитектуры современных многопроцессорных вычислительных систем параллельной обработки информации, технологий организации параллельных вычислений на многопроцессорных вычислительных комплексах с распределенной или общей оперативной памятью.

В результате освоения дисциплины студент должен:

2. Планируемые результаты обучения по дисциплине (модулю).

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

ОПК-3 - Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности;

ОПК-4 - Способен участвовать в разработке стандартов, норм и правил, а также технической документации, связанной с профессиональной деятельностью;

ОПК-9 - Способен осваивать методики использования программных средств для решения практических задач;

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

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

Знать:

Управление эффективностью ресурсного обеспечения ИТ проекта

Уметь:

Управление взаимоотношениями с заинтересованными лицами в процессе обеспечения техническими, технологическими информационно-коммуникационными ресурсами

Владеть:

навыками практического применения технологии анализа Big Data

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 4 з.е. (144 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	Сем. №5
Контактная работа при проведении учебных занятий (всего):	80	80
В том числе:		
Занятия лекционного типа	32	32
Занятия семинарского типа	48	48

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 64 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	Введение в понятия высокопроизводи-тельных вычислений. Основные направ-ления развития высокопроизводительных компьютеров. Важность проблематики параллельных вычислений Пути достижения параллелизма. Векторная и конвейерная обработка данных. Многопроцессорная и многомашинная, параллельная обработка данных. Закон Мура, сдерживающие факторы наращивания количества тран-зисторов на кристалле и частоты процес-соров. Привлекательность подхода па-раллельной обработки данных Сдерживающие факторы повсеместного внедрения параллельных вычислений Ведомственные, национальные и другие программы, направленные на развитие параллельных вычислений в России. Необходимость изучения дисциплины параллельного программирования. Пере-чень критических задач, решение кото-рых без использования параллельных вычислений затруднено или вовсе невоз-можно.
2	Классификация микропроцессорныз ВС Системы с распределенной, общей памятью, примеры систем. Массивно-параллельные системы (MPP). Симметричные мультипроцессорные системы (SMP). Параллельные векторные системы (PVP). Системы с неоднородным доступом к памяти (Numa) Компьютерные кластеры – специализированные и полнофункцио-нальные. История возникновения компь-ютерных кластеров–проект Beowulf. Ме-та-компьютинг. Классификация Флинна, Шора и т.д. Организация межпроцессор-ных связей – коммуникационные тополо-гии. Примеры сетевых решений для со-здания кластерных систем
3	Основные принципы организации параллельной обработки данных: модели, методы и технологии параллельного программирования Функциональный параллелизм, паралле-лизм по данным. Парадигма master-slave. Парадигма SPMD. Парадигма конвейеризации. Парадигма «разделяй и властвуй». Спекулятивный параллелизм. Важность выбора технологии для реализации алго-ритма Модель обмена сообщениями – MPI. Модель общей памяти – OpenMP. Концепция виртуальной, разделяемой памяти – Linda. Российские разработки – Т-система, система DVM. Проблемы создания средства автоматического распараллеливания программ
4	Параллельное программирование с ис-пользованием интерфейса передачи со-общений MPI Библиотека MPI Библиотека MPI. Модель SIMD. Инициа-лизация и завершение MPI-приложения. Точечные обмены данными между про-цессами MPI-программы. Режимы буфе-ризации. Проблема deadlock’ов. Коллек-тивные взаимодействия процессов в MPI. Управление группами и коммуникаторами в MPI
5	Параллельное программирование на системах с общей памятью (OpenMP) Введение в OpenMP Стандарты программирования для систем с разделяемой памятью. Создание многопоточных приложений. Использование многопоточности при программировании для многоядерных платформ. Синхронизация данных между ветвями в параллельной про-грамме. Директивы языка OpenMP
6	Параллельное программирование многоядерных GPU. Кластеры из GPU и супер-компьютеры на гибридной схеме Существующие многоядерные системы. GPU. Использование OpenMP и MPI тех-нологий совместно с CUDA. Степень па-раллелизма численного алгоритма. Закон Амдала. Параллельный алгоритм реше-ния СЛАУ

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Введение в сети 1. Планирование архитектуры 2. Подготовка документации для архитектуры 3. Подключение к оборудовани. cisco и настройка сети 4. Статическая маршрутизация 5. Динамическая маршрутизация и VPN
2	Высокопроизводительные вычисления. Классификация микропроцессорных ВС 1. Высокопроизводительные вычисления 2. Классификация ВС
3	Параллельное программирование. Параллельная обработка данных 1. Параллельное программирование 2. Параллельная обраборка данных
4	Работа с vCenter Server 1.Понятие vCenter 2. Установка и развертывание vCenter
5	Развертывание среды для обработки данных при помощи Hadoop 1. Обработка данных в Hadoop 2. Распределенная файловая система Hadoop 3. Разработка приложений
6	Составление спецификаций оборудования для работы с высокопроизводительными вычислениями 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	Выполнение курсовой работы.
5	Подготовка к промежуточной аттестации.
6	Подготовка к текущему контролю.

4.4. Примерный перечень тем курсовых работ

1. Анализ данных с использованием алгоритмов кластеризации
2. Кластеризация данных с помощью нечетких отношений
3. Метрики, применяемые в Data mining
4. Основные стандарты Data mining

5. Направления использования эволюционных алгоритмов анализа данных

6. Анализ данных с использованием генетических алгоритмов

7. Применение методов Data mining для решения практических задач

8. Технология Knowledge Discovery in Databases (KDD)

9. Характеристики промышленных инструментальных средств Data mining

10. Использование реляционной модели построения хранилищ данных (ROLAP)

11. Использование многомерного подхода в построении хранилищ данных (MOLAP)

12. Использование гибридных (HOLAP) и виртуальных хранилищ данных

13. Технологии и методы оценки качества, очистки и предобработки анализируемых данных

14. Технология практического применения сэмплинга (sampling)

15. Сущность и направления использования аффинитивного анализа данных

16. Подходы к решению задач поиска ассоциативных правил

17. Анализ данных с использованием сети Кохонена (Kohonen network)

18. Анализ данных с использованием самоорганизующихся карт Кохонена (Self organizing map)

19. Технология анализа данных с применением регрессионных моделей

20. Технология построение и оценка значимости простой регрессионной модели

21. Характеристика алгоритмов построения деревьев решений

22. Подготовка управленческих решений на основе метода деревьев решений

23. Принципы построения и направления практического применения нейросетевых моделей

24. Подходы к анализу данных на базе ансамблей моделей

25. Применение моделей анализа временных рядов

26. Технологии обогащения данных

27. Технологии упрощения деревьев решений

28. Алгоритмы обучения нейронных сетей

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.	https://e.lanbook.com/book/107901
2	Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных Флах П.	https://e.lanbook.com/book/69955
3	Математические методы распознавания образов Местецкий Л.М	https://e.lanbook.com/book/100634

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

http://smarty.php.net
http://yandex.ru/cgi/vesna
http://www.virtech.ru
http://www.oracle.sun.com

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

ОС Windows
Microsoft Office;
ASP.NET;
Silverlight 4

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

Компьютер преподавателя
Intel Core i7-9700 / Asus PRIME H310M-R R2.0 / 2x8GB / SSD 250Gb / DVDRW
Компьютеры студентов
Intel Core i9-9900 / B365M Pro4 / 2x16GB / SSD 512Gb /
экран для проектора, маркерная доска,

Проектор Optoma W340UST,

9. Форма промежуточной аттестации:

Курсовая работа в 5 семестре.
Экзамен в 5 семестре.

10. Оценочные материалы.

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

Авторы:

старший преподаватель Академии
"Высшая инженерная школа" И.В. Зенковский

Согласовано:

Заместитель директора академии Д.В. Паринов

Председатель учебно-методической
комиссии Д.В. Паринов