

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт прикладной математики и компьютерных наук

УТВЕРЖДЕНО:
Директор
А. В. Замятин

Рабочая программа дисциплины

Технологии высокопроизводительной обработки больших данных

по направлению подготовки / специальности

10.05.01 Компьютерная безопасность

Направленность (профиль) подготовки/ специализация:
Анализ безопасности компьютерных систем

Форма обучения
Очная

Квалификация
Специалист по защите информации

Год приема
2025

СОГЛАСОВАНО:
Руководитель ОП
В.Н. Тренькаев

Председатель УМК
С.П. Сущенко

Томск – 2025

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-1 Способен оценивать роль информации, информационных технологий и информационной безопасности в современном обществе, их значение для обеспечения объективных потребностей личности, общества и государства.

ОПК-14 Способен проектировать базы данных, администрировать системы управления базами данных в соответствии с требованиями по защите информации.

ОПК-2 Способен применять программные средства системного и прикладного назначений, в том числе отечественного производства, для решения задач профессиональной деятельности.

ОПК-3 Способен на основании совокупности математических методов разрабатывать, обосновывать и реализовывать процедуры решения задач профессиональной деятельности.

ПК-2 Способен разрабатывать требования к программно-аппаратным средствам защиты информации компьютерных систем и сетей.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-1.1 Учитывает современные тенденции развития информационных технологий в своей профессиональной деятельности

ИОПК-14.1 Понимает модели и структуры данных, физические модели баз данных, принципы организации и методы проектирования баз данных, языки и системы программирования баз данных

ИОПК-2.3 Формулирует предложения по применению программных средств системного и прикладного назначений, в том числе отечественного производства, используемых для решения задач профессиональной деятельности

ИОПК-3.2 Осуществляет применение основных понятий, фактов, концепций, принципов математики и информатики для решения задач профессиональной деятельности

ИПК-2.2 Разрабатывает математические модели, реализуемые в средствах защиты информации

ИПК-2.3 Проводит исследования с целью нахождения наиболее целесообразных практических решений по обеспечению защиты информации

2. Задачи освоения дисциплины

– обучить студентов различным методам организации высокопроизводительных вычислений;

– сформировать у студентов навыки использования различных программных инструментов для организации параллельных вычислений.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к Блоку 1 «Дисциплины (модули)».

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор. Дисциплина входит в «Модуль «Введение в искусственный интеллект».

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Восьмой семестр, экзамен

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: «Введение в интеллектуальный анализ данных».

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

-лекции: 16 ч.

-практические занятия: 32 ч.

в том числе практическая подготовка: 32 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Актуальность, базовая терминология и тенденции развития. Параллелизм компьютерных вычислений.

Краткое содержание.

Понятие высокопроизводительных вычислений. Способы повышения производительности вычислительной системы. Понятие параллельных вычислений. Причины, порождающие вычислительный параллелизм. Мультипроцессирование. Классификация уровней параллелизма, предложенная П. Треливеном. Векторная обработка данных. Многофункциональная обработка данных. Конвейер команд.

Тема 2. Архитектура высокопроизводительных вычислительных систем. Классификация вычислительных систем.

Краткое содержание.

Классификация вычислительных систем Флинна. Системы с одним потоком команд и одним потоком данных. Системы с одним потоком команд и множеством потоков данных. Системы с множеством потоков команд и одним потоком данных. Системы с множеством потоков команд и множеством потоков данных.

Классификация вычислительных систем с множеством потоков команд и множеством потоков данных. Системы с разделяемой памятью. Системы с распределённой памятью. Вычислительные системы с централизованной общей памятью. Вычислительные системы с распределённой общей памятью. Симметричная многопроцессорная архитектура. Архитектура CC-NUMA. Архитектура COMA. Архитектура NCC-NUMA. Вычислители с массовым параллелизмом. Кластеры, их типы.

Тема 3. Грид-системы. Облачные технологии

Краткое содержание.

Понятие грид-систем. Свойства грид-систем. Класс задач, эффективно решаемых с помощью грид-систем. Классификация грид-систем с точки зрения выделения вычислительных ресурсов. Примеры грид-систем. Понятие облачных технологий и вычислений. Свойства облачных технологий. Классификация облачных сервисов по типу ресурса, к которому предоставляется доступ. Классификация облачных сервисов по модели развёртывания. Технология обработки данных MapReduce. Сеть доставки содержимого. Обнаружение ресурсов в грид-системах. Соответствие виртуальной топологии и аппаратной архитектуры. Справедливое распределение ресурсов между приложениями.

Тема 4. Распределённые файловые системы.

Понятие распределённых файловых систем. Свойства распределённых файловых систем. Примеры распределённых файловых систем.

Тема 5. Общие вычисления на видеокарте: архитектура современных видеокарт и программные интерфейсы.

Понятие общих вычислений на видеокарте. Архитектура современных видеокарт в сравнении с архитектурой центрального процессора. Сравнение производительности современных видеокарт и центральных процессоров. Графический конвейер.

Программные интерфейсы для доступа к вычислительным ресурсам видеокарты, их классификация. Технология Nvidia CUDA, её архитектура и типы памяти. Способы оптимизации программ, написанных с применением технологии Nvidia CUDA.

Тема 6. Программирование для высокопроизводительных вычислений. Методология проектирования параллельных алгоритмов.

Проблемы параллельного программирования. Методология организации параллельных вычислений для SIMD архитектуры. Методология организации параллельных вычислений для MIMD архитектуры. Показатели качества параллельных методов.

Библиотеки для обмена сообщениями. MPI (Message Passing Interface). OpenMP (Open Multi-Processing). Пример программы с использованием OpenMP. Шаблоны эффективной работы с памятью. Балансировка нагрузки в гетерогенных кластерах.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, опросов по лекционному материалу, контроля выполнения практических работ, и фиксируется в форме контрольной точки не менее одного раза в семестр.

Практическая подготовка оценивается по результатам выполненных практических работ.

Оценочные материалы текущего контроля размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен в восьмом семестре проводится в письменно-устной форме. Экзаменационный билет содержит теоретический вопрос. Продолжительность экзамена 1,5 часа. Допуском до экзамена является выполнение двух практических работ. При выполнении менее трёх работ максимальная оценка на экзамене – 4.

Оценка «отлично» ставится, если студент полноценно ответил на вопрос билета, а также на дополнительный вопрос, требующий аналитического сопоставления знаний, полученных при изучении различных тем данной дисциплины. Оценка «хорошо» ставится, если студент полноценно ответил на вопрос билета, но не ответил на дополнительный вопрос. Оценка «удовлетворительно» ставится, если студент частично ответил на вопрос билета.

Оценочные материалы для проведения промежуточной аттестации размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в системе электронного обучения «IDO» - <https://lms.tsu.ru/course/view.php?id=28933>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

в) Методические указания по проведению лабораторных работ.

г) Методические указания по организации самостоятельной работы студентов.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

– S. Srinivasan. Cloud Computing Basics electronic resource / New York, NY : Springer New York : Imprint: Springer, 2014. – 400 с.

– A Ohri. R for Cloud Computing electronic resource: An Approach for Data Scientists. – New York, NY : Springer New York : Imprint: Springer, 2014. – 336 с.

– Zaigham Mahmood. Cloud Computing: Methods and Practical Approaches, – London: Springer London: Imprint: Springer, 2013. – 259 с.

б) дополнительная литература:

– Christoph Fehling, Frank Leymann, Ralph Retter, Walter Schupeck, Peter Arbitter. Cloud Computing Patterns electronic resource: Fundamentals to Design, Build, and Manage Cloud Applications - Vienna: Springer Vienna: Imprint: Springer, 2014. - 280 с.

– Xiaolin Li, Judy Qiu. Cloud Computing for Data-Intensive Applications electronic resource – New York, NY: Springer New York: Imprint: Springer, 2014. – 226 с.

– Zaigham Mahmood. Cloud Computing electronic resource: Challenges, Limitations and R&D Solutions – New York, NY: Springer New York: Imprint: Springer, 2014. – 321 с.

в) ресурсы сети Интернет:

– MapReduce Tutorial: https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html

– HDFS Architecture Guide: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

– Microsoft Visual Studio Community Edition 2013 (C++), библиотека классов, реализующая технологию MapReduce (например, <https://github.com/cdmh/mapreduce>). Все используемые программные продукты являются свободно распространяемыми.

б) информационные справочные системы:

– Электронный каталог Научной библиотеки	ТГУ	–
http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system		

– Электронная библиотека (репозиторий)	ТГУ	–
http://vital.lib.tsu.ru/vital/access/manager/Index		

– ЭБС Лань – <http://e.lanbook.com/>

– ЭБС Консультант студента – <http://www.studentlibrary.ru/>

– Образовательная платформа Юпайт – <https://urait.ru/>

– ЭБС ZNANIUM.com – <https://znanium.com/>

– ЭБС IPRbooks – <http://www.iprbookshop.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

Для проведения лабораторных работ требуются компьютеры, оснащённые дискретными видеокартами фирмы Nvidia с поддержкой технологии CUDA.

15. Информация о разработчиках

Дружинин Денис Вячеславович, канд. техн. наук, доцент кафедры теоретических основ информатики.