

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Проректор по учебной работе
и образовательным инновациям

О.Н. Здрок

«30» *июня* 2020 г.

Регистрационный № УД 8450/уч.

СИСТЕМЫ ПАРАЛЛЕЛЬНОЙ И ПОТОКОВОЙ ОБРАБОТКИ ДАННЫХ

Учебная программа учреждения высшего образования по учебной
дисциплине для специальности

1-31 80 03 Математика и компьютерные науки

профилизация

Веб-программирование и интернет-технологии

2020 г.

Учебная программа составлена на основе образовательного стандарта высшего образования ОСВО 1-31 80 03-2019 и учебных планов G31-033/уч., G31з-034/уч. от 11.04.2019

СОСТАВИТЕЛЬ:

Гуревский Алексей Николаевич, старший преподаватель кафедры веб-технологий и компьютерного моделирования Белорусского государственного университета.

РЕЦЕНЗЕНТ:

Ковалева Ирина Львовна, кандидат технических наук, доцент кафедры программного обеспечения информационных систем и технологий Белорусского национального технического университета.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой Веб-технологий и компьютерного моделирования (протокол №9 от 20.05.2020 г.);

Научно-методическим советом Белорусского государственного университета (протокол №5 от 17.06.2020 г.)

Заведующий кафедрой _____ В.М. Волков

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Целью дисциплины «Системы параллельной и потоковой обработки данных» является ознакомление студентов магистратуры с теоретическими основами параллельной обработки данных и инструментами, доступными для потоковой обработки данных.

Задачи учебной дисциплины:

- формирование у магистрантов способностей самостоятельно разрабатывать алгоритмы решения задач и их анализировать;
- развивать и использовать инструментальные средства, информационные среды, автоматизированные системы;
- использовать математические и компьютерные методы исследований при анализе современных естественнонаучных, экономических, социально-политических процессов;
- приобретение способностей самостоятельно расширять компьютерные математические знания с дальнейшим их использованием при анализе математических моделей широкого круга прикладных задач.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием (магистра): учебная дисциплина имеет прикладную направленность и готовит магистрантов к освоению и реализации методов потоковой и параллельной обработки данных.

Учебная дисциплина «Системы параллельной и потоковой обработки данных» относится к модулю «Анализ данных и методы искусственного интеллекта» компонента учреждения высшего образования и адресована студентам 2-го года обучения в магистратуре по специальности 1-31 80 03 Математика и компьютерные науки, профилизация «Веб-программирование и интернет-технологии».

Изучение дисциплины «Системы параллельной и потоковой обработки данных» базируется на знаниях, полученных при изучении дисциплины «Машинное обучение».

Требования к компетенциям

Освоение учебной дисциплины «Системы параллельной и потоковой обработки данных» должно обеспечить формирование следующих *специализированных компетенций*:

СК-2. Быть способным эффективно использовать инструменты анализа и алгоритмы обработки данных.

В результате изучения дисциплины студент магистратуры должен:

знать:

- классификацию систем параллельной обработки данных;
- концепции работы систем потоковой обработки данных
- особенности работы фреймворков для потоковой обработки данных

уметь:

- проводить первичную обработку данных;
- применять известные системы потоковой обработки данных
- оценивать качество предложенного решения;

владеть:

- основами технологий параллельного программирования
- способами автоматизации деятельности на основе анализа данных и моделирования, без явного предварительного написания алгоритмов действий.

Структура учебной дисциплины

В соответствии с учебными планами специальности на изучение дисциплины отводится:

Профилизация, форма обучения	Срок обучения, лет	Курс	Семестр	Зачетные единицы	Экзамен, семестр	Зачет, семестр	Всего часов	В том числе ауд.	Из них	
									лекций	лабораторных занятий
Веб-программирование и интернет технологии, дневная	1.8	2	3	6		3	198	72	36	36
Веб-программирование и интернет технологии, заочная	2	2	3	6		3	198	16	8	8

Трудоемкость учебной дисциплины составляет 6 зачетных единиц.

Форма текущей аттестации по учебной дисциплине – дневная и заочная формы обучения – зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. Введение. Большие задачи и суперкомпьютеры. Суперкомпьютеры в компьютерном мире. Последовательная, параллельная и конвейерная обработка данных. Закон Амдала и его следствия. Последовательные и параллельные алгоритмы.

Тема 2. Архитектура параллельных вычислительных систем. Компьютеры с общей и распределенной памятью, принципы организации, сходства и различия, достоинства и недостатки. Скалярные и векторные команды. Векторизация программ.

Тема 3. Технологии параллельного программирования. Схемы вычислительного процесса для параллельных программ: SPMD, Мастер/Рабочие. Эффективность, продуктивность, переносимость. Технологии MPI, OpenMP, Linda

Тема 4. Производительность параллельных вычислительных систем. Методы оценки производительности. Введение единого числового параметра, Mflops, MIPS. Пиковая и реальная производительность компьютеров.

Тема 5. Введение в теорию анализа структуры программ и алгоритмов. Графовые модели программ. Построение графа алгоритма для линейного класса программ. Виды параллелизма: конечный, массовый, координатный, скошенный.

Тема 6. Введение в потоковую обработку данных. Системы реального времени и потоковые системы. Безопасность потоковых систем. Масштабирование.

Тема 7. Получение данных от клиентов. Паттерны взаимодействия. Отказоустойчивость.

Тема 8. Транспортировка данных. Основные концепции. Безопасность. Отказоустойчивость.

Тема 9. Анализ потоковых данных. Архитектуры распределенной обработки потоков. Скользящие и прыгающие окна. Методы обобщения.

Тема 10. Сохранение результатов анализа данных. Долговременные хранилища. Хранение данных в памяти

Тема 11. Применение Apache Kafka для потоковой обработки данных. Установка и настройка Kafka. Анализ данных с помощью Storm и Kafka. Хранилище данных в памяти.

Тема 12. Применение Apache Spark для потоковой обработки данных. Установка и настройка Spark. Spark Streaming.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ
дневная форма получения образования

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УРС	Литература	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное			
1	Введение. Большие задачи и суперкомпьютеры	2						[1] о, [1] д	Опрос
2	Архитектура параллельных вычислительных систем	2			2			[1] о, [1] д	Опрос, защита лабораторной работы
3	Технологии параллельного программирования	4			4			[1] о, [1] д	Опрос, защита лабораторной работы
4	Производительность параллельных вычислительных систем	2			2			[1] о, [1] д	Опрос, защита лабораторной работы
5	Введение в теорию анализа структуры программ и алгоритмов	2			2			[1] о, [1] д	Опрос, защита лабораторной работы
6	Введение в потоковую обработку данных	2						[2-3] о [2] д	Опрос

7	Получение данных от клиентов	2			2			[2-3] о [2] д	Опрос, защита лабораторной работы
8	Транспортировка данных	2			2			[2-3] о [2] д	Опрос, защита лабораторной работы
9	Анализ потоковых данных	4			4			[2-3] о [2] д	Опрос, защита лабораторной работы
10	Сохранение результатов анализа данных	2			2			[2-3] о [2] д	Опрос, защита лабораторной работы
11	Применение Apache Kafka для потоковой обработки данных	6			8			[2-3] о [2] д	Опрос, защита лабораторной работы, творческое задание
12	Применение Apache Spark для потоковой обработки данных	6			8			[2-3] о [2] д	Опрос, защита лабораторной работы, творческое задание
	ВСЕГО	36			36				Зачет

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ
заочная форма получения образования

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УРС	Литература	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное			
1	Введение. Большие задачи и суперкомпьютеры.	2			2		[1] о [1] д	Опрос, защита лабораторной работы	
2	Архитектура параллельных вычислительных систем.								
3	Технологии параллельного программирования.								
4	Производительность параллельных вычислительных систем.								
5	Введение в теорию анализа структуры программ и алгоритмов								
6	Введение в потоковую обработку данных.	2			2		[2-3] о [2] д	Опрос, защита лабораторной работы	

7	Получение данных от клиентов.								
8	Транспортировка данных.								
9	Анализ потоковых данных.								
10	Сохранение результатов анализа данных								
11	Применение Apache Kafka для потоковой обработки данных.	4			4			[2-3] о [2] д	Опрос, защита лабораторной работы, творческое задание
12	Применение Apache Spark для потоковой обработки данных								
ВСЕГО		8			8				Зачет

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Параллельная обработка данных [Электронный ресурс]. – Режим доступа: <https://parallel.ru/vvv>. – Дата доступа: 15.05.2020.
2. Пселтис, Э. Поточковая обработка данных / Э. Пселтис. – М. : ДМК Пресс, 2018. – 218 с.
3. Карау, Х. Изучаем Spark. Молниеносный анализ данных / Х. Карау. – М. : ДМК Пресс, 2015. – 304 с.

Перечень дополнительной литературы

1. Воеводин, В. Параллельные вычисления / В. Воеводин. – СПб.: БХВ-Петербург, 2002. – 608 с.
2. Нархид, Н. Apache Kafka. Поточковая обработка и анализ данных / Н. Нархид. – СПб. : Питер, 2019. – 320 с.

ПЕРЕЧЕНЬ РЕКОМЕНДУЕМЫХ СРЕДСТВ ДИАГНОСТИКИ И МЕТОДИКА ФОРМИРОВАНИЯ ИТОГОВОЙ ОЦЕНКИ

Диагностика результатов учебной деятельности по дисциплине «Системы параллельной и потоковой обработки данных» проводится, как правило, во время аудиторных занятий. Для диагностики используются устная и устно-письменная формы:

- экспресс-опрос на аудиторных занятиях;
- защита отчетов по заданиям для лабораторных работ;

Контрольные мероприятия проводятся в соответствии с учебно-методической картой дисциплины.

Формой текущей аттестации по дисциплине «Системы параллельной и потоковой обработки данных» учебным планом предусмотрен зачет.

При формировании итоговой оценки используется рейтинговая оценка знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний и текущей аттестации студентов по дисциплине.

Весовые коэффициенты, определяющие вклад текущего контроля знаний и текущей аттестации в рейтинговую оценку:

- отчеты по лабораторным работам – 50 %;
- выполнение творческого задания – 50 %.

Примерная тематика лабораторных занятий

1. Векторизация программ (2 ч).
2. Технологии MPI, OpenMP, Linda (4 ч).
3. Методы оценки производительности (2 ч).
4. Построение графа алгоритма для линейного класса программ (2 ч).
5. Типичные паттерны взаимодействия при получении данных от клиента (2 ч).
6. Применение базовых концепция транспортировки данных (2 ч).
7. Алгоритмы анализа потоковых данных (4 ч).
8. Хранилища данных (2 ч).
9. Использование Apache Kafka на реальном примере (6 ч).
10. Использование Apache Spark на реальном примере (6 ч).

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используется **метод проектного обучения**, который предполагает:

- способ организации учебной деятельности студентов, развивающий актуальные для учебной и профессиональной деятельности навыки планирования, самоорганизации, сотрудничества и предполагающий создание собственного продукта;

- приобретение навыков для решения исследовательских, творческих, социальных, предпринимательских и коммуникационных задач.

При организации образовательного процесса используются **методы и приемы развития критического мышления**, которые представляют собой систему, формирующую навыки работы с информацией в процессе чтения и письма; понимания информации как отправного, а не конечного пункта критического мышления.

При организации образовательного процесса используется **метод группового обучения**, который представляет собой форму организации учебно-познавательной деятельности обучающихся, предполагающую функционирование разных типов малых групп, работающих как над общими, так и специфическими учебными заданиями.

Методические рекомендации по организации самостоятельной работы обучающихся

При изучении учебной дисциплины рекомендуется использовать следующие формы самостоятельной работы:

- поиск (подбор) и обзор литературы и электронных источников по индивидуально заданной проблеме курса;
- научно-исследовательские работы;
- подготовка к участию в конференциях и конкурсах.

**ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ С ДРУГИМИ
ДИСЦИПЛИНАМИ СПЕЦИАЛЬНОСТИ**

<p>Название учебной дисциплины, с которой требуется согласование</p>	<p>Название кафедры</p>	<p>Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине</p>	<p>Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)</p>
<p>Приложения компьютерного моделирования</p>	<p>Кафедра Веб-технологий и компьютерного моделирования</p>	<p>Изменений нет</p>	<p>Оставить содержание учебной программы без изменений (протокол №9 от 20.05.2020 г.)</p>

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры *Веб-технологий и компьютерного моделирования* (протокол № __ от __202__ г.)

Заведующий кафедрой

докт. физ.-мат. наук, доцент

В.М. Волков

УТВЕРЖДАЮ

Декан факультета

докт. физ.-мат. наук, доцент

С.М. Босяков