

# БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Проректор по учебной  
работе ВГУ



А.Л. Толстик

Регистрационный № УД-1047/баз.

## Технологии и компьютерные системы обработки больших объемов информации

Учебная программа учреждения высшего образования  
по учебной дисциплине для специальности второй ступени высшего  
образования (магистратуры) с углубленной подготовкой специалиста

1-31 81 09

Алгоритмы и системы обработки больших объемов  
информации

2014 г.

**СОСТАВИТЕЛИ:**

**В.М. Котов**, заведующий дискретной математики и алгоритмики Белорусского государственного университета, доктор физ.-матем. наук, профессор;

**Н.А. Лиходед**, профессор кафедры вычислительной математики Белорусского государственного университета, доктор физ.-матем. наук, профессор;

**Е.П. Соболевская**, доцент кафедры дискретной математики и алгоритмики Белорусского государственного университета, кандидат физико-математических наук, доцент;

**А.А. Толстик**, ассистент кафедры вычислительной математики Белорусского государственного университета

**РЕЦЕНЗЕНТЫ:**

С. В. Баханович – ведущий научный сотрудник отдела вычислительной математики Института математики НАН Беларуси, кандидат физико-математических наук

А.О. Сикорский – директор ООО «ЯндексБел»

**РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:**

Кафедрой дискретной математики и алгоритмики Белорусского государственного университета  
(протокол № 13 от 16 апреля 2014 г.).

Научно-методическим советом по прикладной математике и информатике Учебно-методического объединения по естественнонаучному образованию  
(протокол № 7 от 22 апреля 2014г.).

Ответственный за редакцию: Е.П. Соболевская

Ответственный за выпуск: Е.П. Соболевская

## Пояснительная записка

Учебная программа по учебной дисциплине «Технологии и компьютерные системы обработки больших объемов информации» разработана в соответствии с типовыми учебными планами и образовательными стандартами второй ступени высшего образования (магистратуры) с углубленной подготовкой специалиста по специальности 1-31 81 09 «Алгоритмы и системы обработки больших объемов информации».

Учебная дисциплина «Технологии и компьютерные системы обработки больших объемов информации» знакомит магистрантов с основными тенденциями развития современных технологий и принципами построения компьютерных систем, предназначенных для обработки больших объемов информации.

Основой для изучения учебной дисциплины являются следующие учебные дисциплины первой ступени высшего образования: «Программирование», «Операционные системы», «Алгоритмы и структуры данных», «теория алгоритмов», «Архитектура компьютеров», «Компьютерные сети», «Модели данных и системы управления базами данных» и «Исследование операций».

**Цель преподавания учебной дисциплины «Технологии и компьютерные системы обработки больших объемов информации»:** создание базы для программной реализации и интегрирования в системы хранения и обработки данных алгоритмов решения прикладных задач, возникающих при обработке больших объемов информации и формирование у магистрантов умения использовать технологии и компьютерные системы распределенных вычислений при решении задач, связанных с обработкой и анализом данных.

При изложении материала учебной дисциплины важно показать возможности использования современных систем хранения данных при решении прикладных задач обработки и анализа больших объемов информации, возникающих в различных областях науки, техники, экономики и др. Целесообразно также показать важность эффективной параллельной и распределенной обработки данных.

**Основные задачи,** решаемые при изучении учебной дисциплины «Технологии и компьютерные системы обработки больших объемов информации»:

- изучение различных компьютерных систем, предназначенных для обработки больших объемов информации;
- изучение особенностей реализации алгоритмов, использующих различные виды памяти компьютера;
- использование кластерных компьютерных систем при решении математических и прикладных задач.

В результате изучения учебной дисциплины магистрант должен:

**знать:**

- тенденции развития технологий и компьютерных систем обработки больших объемов информации;
- теоретические основы параллельных вычислений и многопоточного программирования;
- современные технологии распределенных вычислений, хранения и обработки больших объемов данных, их преимущества, недостатки и области применения;
- принципы построения распределенных систем.

**уметь:**

- реализовывать и интегрировать разработанные алгоритмы в специализированные компьютерные системы;
- использовать технологии и компьютерные системы распределенных вычислений при решении задач, связанных с обработкой и анализом данных;
- принимать решения о выборе технологии с точки зрения ее оптимальности для решения поставленной задачи.

**владеть:**

- современными компьютерными технологиями обработки больших объемов данных;
- навыками выбора и обоснования методов и инструментов решения задач, требующих параллельного и распределенного программирования;
- навыками реализации и использования параллельных и распределенных программ.

Учебная программа рассчитана на 172 часа, из них 68 аудиторных часов, в том числе 34 лекционных часов и 34 часа лабораторных занятий.

Рекомендуемая форма текущей аттестации – экзамен, зачет.

### Примерный тематический план

№ темы	Количество аудиторных часов	
	Лекции	Практические занятия
1. Введение	2	
<b>Раздел I. Системы хранения данных</b>		
2. Введение в системы хранения данных. Среда систем хранения данных	2	2
3. Архитектура систем хранения данных	2	2
<b>Раздел II. Многопоточное программирование</b>		
4. Одновременные системы. Процессы и потоки	2	2
5. Синхронизация, Модель памяти и низкоуровневые примитивы синхронизации	2	2
<b>Раздел III. Параллельные вычисления и технологии параллельного программирования</b>		
6. Основы параллельного программирования на системах с общей памятью. Технология OpenMP	6	6
7. Основы параллельного программирования на системах с распределенной памятью. Технология MPI	6	6
<b>Раздел IV. Модель вычислений MapReduce и платформа Apache Hadoop</b>		
8. Модель вычислений MapReduce. Область применения и примеры задач	4	4
9. Платформа Apache Hadoop. Инструментарий для работы с Hadoop	4	4
10. Приемы и стратегии реализации MapReduce-программ	4	6
<b>Всего часов</b>	<b>68</b>	

## Содержание учебного материала

### **1. Введение**

Различные типы компьютерных систем обработки и хранения больших объемов информации, их преимущества, недостатки и области применения.

### **Раздел I. Системы хранения данных**

#### **2. Введение в системы хранения данных. Среда систем хранения данных**

Данные и информация, типы данных, эволюция систем хранения данных, большие данные. Среда систем хранения данных, основные элементы. Виртуализация приложений и серверов.

#### **3. Архитектура систем хранения данных**

Основные компоненты. Управление кешированием. Защита от сбоя. Классы систем хранения данных.

### **Раздел II. Многопоточное программирование**

#### **4. Одновременные системы. Процессы и потоки**

Одновременные системы. Способы реализации одновременных систем, процессы и потоки, программный инструментарий.

#### **5. Синхронизация. Модель памяти и низкоуровневые примитивы синхронизации**

Основы многопоточного программирования. Типичные ошибки многопоточного программирования, Взаимные исключения и условная синхронизация. Модель памяти и низкоуровневые примитивы синхронизации.

### **Раздел III. Параллельные вычисления**

#### **6. Основы параллельного программирования на системах с общей памятью. Технология OpenMP**

Системы с общей памятью. Технология OpenMP. Типовые модели программирования и шаблоны. Преимущества и недостатки. Область применения и примеры задач

#### **7. Основы параллельного программирования на системах с распределенной памятью. Технология MPI**

Системы с распределенной памятью. Технология MPI. Типовые модели программирования и шаблоны. Преимущества и недостатки. Область применения и примеры задач

## **Раздел IV. Модель вычислений MapReduce и платформа Apache Hadoop**

### **8. Модель вычислений MapReduce. Область применения и примеры задач**

Модель вычислений MapReduce. Принципы параллельной реализации вычислений. Область применения и примеры задач. Ограничения модели MapReduce, расширения и альтернативные подходы.

### **9. Платформа Apache Hadoop. Инструментарий для работы с Hadoop**

Платформа Apache Hadoop. Интерфейсы прикладного программирования и реализация программ для Hadoop. Высокоуровневые языки и инструментарий для работы с Hadoop.

### **10. Приемы и стратегии реализации MapReduce-программ**

Приемы и стратегии реализации MapReduce-программ. Практические примеры использования MapReduce.

## **Информационно-методическая часть**

### **Литература**

#### ***Основная***

1. *Воеводин, В.В.* Параллельные вычисления / В. В. Воеводин, Вл. В. Воеводин. СПб. : БХВ-Петербург, 2002. – 608 с.
2. В., Богачев К.Ю. Основы параллельного программирования. М.: БИНОМ. Лаборатория знаний, 2010. 342 с.
3. Антонов А.С. Параллельное программирование с использованием технологии MPI. М.: Изд-во МГУ, 2004. 71 с.
4. Шпаковский Г.И., Серикова Н.В. Программирование для многопроцессорных систем в стандарте MPI. Минск: БГУ, 2002 г. 323 с.

#### ***Дополнительная***

5. Гергель В.П., Стронгин Р.Г. Основы параллельных вычислений для многопроцессорных вычислительных систем. Нижний Новгород: Изд-во ННГУ им. Н.И. Лобачевского, 2003. 184 с.
6. *Кормен, Т.* Алгоритмы : построение и анализ/ Т. Кормен, Ч. Лейзерсон, Р. Ривест, К. Штайн. М. : Вильямс, 2005. 1296 с.
7. *Лиходед Н. А.* Методы распараллеливания гнезд циклов: курс лекций. - Минск : БГУ, 2008. – 100 с.

### **Диагностика компетенций магистранта**

Условия для самостоятельной работы магистрантов, в частности, для развития навыков самоконтроля, способствующих интенсификации учебного процесса, обеспечиваются наличием и полной доступностью электронных (и бумажных) учебно-методических пособий по основным разделам дисциплин, а также доступность использования суперкомпьютера.

Текущий контроль усвоения знаний рекомендуется осуществлять в виде проверки индивидуальных заданий, проведения коллоквиумов или промежуточных экзаменов.