МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ФГБОУ ВО «ИГУ»
Кафедра естественнонаучных дисциплин

УТВЕРЖДАЮ

ДЕКАН В.К. КАРНАУХОВА

“21” марта 2018г.

Рабочая программа дисциплины

Наименование дисциплины	**Б1.В.ДВ.10.1 Параллельные вычисления и большие данные** *(индекс дисциплины по учебному плану, наименование дисциплины)*
Направление подготовки	**09.03.03 "Прикладная информатика"** *(код, наименование направления подготовки)*
Тип образовательной программы	**Прикладной бакалавриат** *(академический бакалавриат, прикладной бакалавриат)*
Направленность (профиль) подготовки	**Прикладная информатика в дизайне** **Разработка программного обеспечения** *(наименование профиля)*
Квалификация (степень) выпускника	**Бакалавр**
Форма обучения	**Очная, очно-заочная** *(очная, очно-заочная)*

Одобрено УМС факультета сервиса и рекламы:
Протокол № 8
от «14» марта 2018 г.

Председатель В.К.Карнаухова

Рекомендовано кафедрой естественнонаучных дисциплин:
Протокол № 9
от «06» марта 2018г.

Зав.кафедрой А.Г. Балахчи

Иркутск 2018 г.

Содержание

	стр.
1. Цели и задачи дисциплины (модуля)	3
2. Место дисциплины (модуля) в структуре ОПОП.	3
3. Требования к результатам освоения дисциплины (модуля)	3
4.Объем дисциплины (модуля) и виды учебной работы	4
5.Содержание дисциплины (модуля)	
5.1 Содержание разделов и тем дисциплины (модуля)	5
5.2 Разделы дисциплины (модуля) и междисциплинарные связи с обеспечиваемыми (последующими) дисциплинами (модулями)	6
5.3 Разделы и темы дисциплин (модулей) и виды занятий	6
6. Перечень семинарских, практических занятий, лабораторных работ, план самостоятельной работы студентов, методические указания по организации самостоятельной работы студентов	7
7. Примерная тематика курсовых работ (проектов) (при наличии)	12
8. Учебно-методическое и информационное обеспечение дисциплины (модуля):	12
а) основная литература;	12
б) дополнительная литература;	12
в) программное обеспечение;	12
г) базы данных, поисково-справочные и информационные системы	12
9. Материально-техническое обеспечение дисциплины (модуля)	13
10. Образовательные технологии	13
11. Оценочные средства (ОС)	13

1. Цели и задачи дисциплины.

Целью дисциплины является формирование у обучающихся комплекс теоретических знаний и практических навыков по работе с большими данными. Знания, полученные в результате освоения дисциплины, помогут в понимании принципов работы различного программного обеспечения в этой области, а также помогут при сборе, разработке методов анализа и самом анализе структурированной или неструктурированной информации огромных объемов

Задачи:

- получение начальных знаний о предмете Big Data
- получение начальных знаний и умений по получению Big Data
- получение начальных знаний и умений по простейшей обработке Big Data
- получение начальных знаний и умений по созданию систем обработки Big Data и использованию BigData в системах принятия решений
- получение начальных знаний и умений по созданию систем обработки Big Data и использованию BigData в системах реального времени

2. Место дисциплины в структуре ОПОП

Дисциплина «Параллельные вычисления и большие данные» является одним из базовых курсов, формирующих профессиональную направленность обучающегося. Для успешного освоения дисциплины у студентов должен быть сформированы базовый уровень ряда дескрипторов по следующим компетенциям:

ПК-2 способность разрабатывать, внедрять и адаптировать прикладное программное обеспечение

ПК-3 способность проектировать ИС в соответствии с профилем подготовки по видам обеспечения

ПК-8 способность программировать приложения и создавать программные прототипы решения прикладных задач

ПК-15 способность осуществлять тестирование компонентов информационных систем по заданным сценариям

ПК-23 способностью применять системный подход и математические методы в формализации решения прикладных задач;

Компетенции, приобретенные при освоении дисциплины, могут быть использованы в практической и научно-исследовательской деятельности в задачах, связанных с обработкой больших объемов различной информации, а также в качестве базы при последующем более детальном ознакомлении с различными вопросами анализа Больших Данных.Результатом освоения дисциплины может стать выполнение общего контрольного работы, связанной с тематикой курса.

3. Требования к результатам освоения дисциплины (модуля):

Процесс изучения дисциплины направлен на формирование следующих компетенций:

ПК-2 способность разрабатывать, внедрять и адаптировать прикладное программное обеспечение

ПК-3 способность проектировать ИС в соответствии с профилем подготовки по видам обеспечения

ПК-8 способность программировать приложения и создавать программные прототипы решения прикладных задач

ПК-15 способность осуществлять тестирование компонентов информационных систем по

заданным сценариям

ПК-23 способностью применять системный подход и математические методы в формализации решения прикладных задач;

ПК-24 способностью готовить обзоры научной литературы и электронных информационно-образовательных ресурсов для профессиональной деятельности

В результате изучения дисциплины студент должен

знать:

З1 ПК-2. Принципы организации и содержание этапов процесса разработки, внедрения и адаптации программного обеспечения;

З2 ПК-2. Возможности современных и перспективных средств разработки программного обеспечения;

З3 ПК-2. Технологии программирования;

З4 ПК-2. Слои программных компонентов;

З5 ПК-2. Технологические и технико-эксплуатационные характеристики архитектур развертывания программных компонентов;

З6 ПК-2. Компоненты программно-технических архитектур, существующие приложения и интерфейсы взаимодействия с ними;

З1 ПК-3. основные понятия проектирования информационных систем по видам обеспечения;

З2 ПК-3. принципы, методологию проектирования и использования профильных ИС по видам обеспечения;

З3 ПК-3. основные технологии проектирования информационных систем, включая каноническую, типовую и CASE технологию. языки формализации функциональных спецификаций;

З4 ПК-3. методы и средства организации и управления проектом ИС на всех стадиях жизненного цикла.

З1 ПК-8. Методы структурного и объектно-ориентированного программирования;

З2 ПК-8. Методы и приемы формализации и алгоритмизации поставленных задач;

З3 ПК-8. Нотации и программные продукты для графического отображения алгоритмов;

З4 ПК-8. Алгоритмы решения типовых задач, области и способы их применения;

З5 ПК-8. Синтаксис выбранного компьютерного языка: программирования или разметки, особенности написания кода на этом языке;

З6 ПК-8. Стандартные библиотеки выбранного языка программирования;

З7 ПК-8. Особенности выбранной среды программирования и средства системы управления базами данных;

З8 ПК-8. Инструментарий для создания и актуализации исходных текстов кода;

З14 ПК-8. Современные компиляторы, отладчики и оптимизаторы программного кода.

З.1 ПК-15. Архитектура, устройство и функционирование вычислительных систем

З.2 ПК-15. Современные методики тестирования разрабатываемых ИС

З1 ПК-23. Теоретические системные основы для формализации экономических проблемных ситуаций;

З2 ПК-23. Принципы, методы математического моделирования;

З3 ПК-23. Этапы формализации прикладных задач с использованием методов экономико-математического моделирования;

З4 ПК-23. Закономерности построения, функционирования и развития систем целеобразования.

З1 ПК-24. принцип нелинейного структурирования информации; закономерности и принципы развития научного знания;

З2 ПК-23. Структуру информационных ресурсов Интернета, предоставляемые услуги, виды доступа к информации;

З3 ПК-23. Методы и технологии поиска и обработки информации средствами Интернет.

уметь:

У1 ПК-8. Использовать методы и приемы формализации и алгоритмизации поставленных задач;

У2 ПК-8. Использовать методы структурного и объектно-ориентированного программирования;

У4 ПК-8. Применять стандартные алгоритмы в соответствующих областях;

У5 ПК-8. Применять выбранные языки программирования/разметки для написания кода;

У6 ПК-8. Использовать выбранную среду программирования и средства системы управления базами данных;

У8 ПК-8. Применять инструментарий для создания и актуализации исходных текстов кода;

У10 ПК-8. Выявлять ошибки в коде ПО;

У11 ПК-8. Применять методы и приемы отладки программного кода;

У13 ПК-8. Применять современные компиляторы, отладчики и оптимизаторы программного кода.

У1 ПК-15. Кодировать на языках программирования

У2 ПК-15. Тестировать ИС с использованием тест-планов (сценариев)

У1 ПК-23. Проводить системный анализ прикладной области;

применять математические методы для формализации и решения прикладных задач;

У2 ПК-23. Строить модели экономических процессов, исследовать их и вырабатывать рекомендации по их практическому применению;

У3 ПК-23. Использовать для анализа проблемной ситуации методы и принципы системного подхода, соответствующие методы измерений и оценки информационных ресурсов в конкретной предметной области;

У4 ПК-23. Обрабатывать статистическую информацию

У1 ПК-24. работать с большими объемами информации;

писать тексты литературным, техническим и рекламным языком;

У2 ПК-24. реферировать, аннотировать и модифицировать тексты на русском и английском языках;

У3 ПК-24. работать с агрегаторами новостей, электронными подписками, социальными сетями, форумами;

владеть:

В1 ПК-2. Навыками разработки, внедрения и адаптации ПО;

В2 ПК-2. Навыками работы в современной программно-технической среде в различных операционных системах.

В1 ПК-8. Навыками формализации и алгоритмизации поставленных задач;

В2 ПК-8. Навыками написания кода с использованием языков программирования/разметки, определения и манипулирования данными;

В3 ПК-8. Навыками оформления кода разметки и программирования в соответствии с установленными требованиями;

В4 ПК-8. Навыками работы с системой контроля версий; навыками проверки и отладки кода программы или разметки;

В1 ПК-15. Навыками проведения интеграционного и модульного тестирования ИС на основе тест-планов (сценариев) в соответствии с трудовым заданием

В1 ПК-23 навыками работы с инструментами системного анализа;

В2 ПК-23 навыками построения, исследования экономико-математических моделей предметной области, а также практического применения широко используемых в экономике прикладных математических моделей для решения экономических задач..

В1 ПК-24. Достаточным словарным запасом, грамматической базой для чтения и

написания обзоров на русском и английском языках;

В2 ПК-23. Различными методами поиска информации в Интернет (по ключевым словам, с помощью каталогов);

В3 ПК-24. Навыками подготовки обзоров научной литературы и электронных информационно-образовательных ресурсов.

4. Объем дисциплины (модуля) и виды учебной работы (разделяется по формам обучения)

Вид учебной работы	Всего часов / зачетных единиц	Семестры			
		3	4	5	6
Аудиторные занятия (всего)	**54**		54		
В том числе:					
Лекции	**18**		18		
Практические занятия (ПЗ)	**36**		36		
Семинары (С)					
Лабораторные работы (ЛР)					
Самостоятельная работа (всего)	**90**		90		
В том числе:					
Курсовой проект (работа)					
Расчетно-графические работы					
Реферат (при наличии)					
Другие виды самостоятельной работы					
Творческая работа (интерактивные мультимедийные проекты)					
Самоподготовка и выполнение домашних заданий	90		90		
Контроль					
Вид промежуточной аттестации (зачет, экзамен)			зачО		
Контактная работа (всего)	**63**		63		
Общая трудоемкость часы зачетные единицы	**144** **4**		144 4		

Очно-заочное

Вид учебной работы	Всего часов / зачетных единиц	Семестры			
		3	4	5	6
Аудиторные занятия (всего)	24		24		
В том числе:					

Лекции	6		6		
Практические занятия (ПЗ)	18		18		
Семинары (С)					
Лабораторные работы (ЛР)					
Самостоятельная работа (всего)	**48**		48		
В том числе:					
Курсовой проект (работа)					
Расчетно-графические работы					
Реферат (при наличии)					
Другие виды самостоятельной работы					
Творческая работа (интерактивные мультимедийные проекты)					
Самоподготовка и выполнение домашних заданий	48		48		
Контроль					
Вид промежуточной аттестации (зачет, экзамен)	ЗаО		ЗаО		
Контактная работа (всего)	**29**		29		
Общая трудоемкость часы зачетные единицы	**72** **2**		72 2		

5. Содержание дисциплины (модуля)

5.1. Содержание разделов и тем дисциплины (модуля). Все разделы и темы нумеруются

Раздел 1. Введение

Тема 1. Проверка базовых знаний и вводная лекция

Введение в Big Data и машинное обучение. Объем данных, скорость данных, различность данных, качество данных, значимость данных. Получение данных. Хранение данных. Доступ к данным. Обработка данных. Проверка гипотез и выявление скрытых зависимостей. Параллельные и облачные вычисления. Машинное обучение. Искусственные нейронные сети. Входной тест.

Тема 2. Введение в Питон

Элементы языка Питон: базовые операторы, структуры и функции. Работа с файлами. Работа с модулями. Создание модулей.

Раздел 2. Дата майнинг

Тема 3. Протоколы и форматы интернета

Структура сервер-клиент. Работа с сокетами. Telnet. Организация протокола HTTP, команды, авторизация. Организация протокола FTP, команды, авторизация. Шифрование. Организация протокола SSH, авторизация. Протокол HTTPS.

Тема 4. Протоколы и форматы баз данных

Структура реляционной базы данных на примере MySQL. SQL - команды. Создание базы данных. Добавление, удаление и редактирование записей. Поиск и выбор записей. Объединения таблиц. Индексы.

Тема 5. Подготовка к анализу текстов

Команды и элементы языка regexp. Шаблоны. Контекстный поиск. Контекстный поиск и замена. Многовариантный поиск. Переменные regexp. Особенности использования на Питоне. Ассоциированные массивы - реализация и использование. Задача сбора статистики.

Тема 6. Подготовка к анализу соотношений

Графы, отношения и пары. Ребра и вершины. Ориентированные и неориентированные графы. Маршруты, пути и циклы. Связность графа. Деревья. Отображение и анализ графов.

Раздел 3. Предварительная обработка больших данных

Тема 7. Принятие простых решений

Статистические распределения, вероятность. Условная вероятность. Функция распределения и плотность распределения. Интегральная оценка вероятности события. Принятие решения на основе статистических данных. Байесовский подход. Проверка статистических гипотез. Ошибки первого и второго рода.

Тема 8. Выделение явных закономерностей

Корреляция. Регрессионная зависимость. Линейная и нелинейная регрессия. Метод наименьших квадратов и случаи его сведения к системе линейных уравнений. Метод максимального правдоподобия. Периодичность процессов и Фурье-анализ.

Тема 9. Выявление скрытых закономерностей.

Метод наименьших квадратов и случаи его сведения к анализу матриц. Матрицы: определители и ранги, собственные числа и собственные вектора. Уменьшение числа параметров. Линейно-зависимые и линейно-независимые параметры. Метод главных компонент. Матричный анализ. Сингулярное разложение.

Раздел 4. Экспертные системы

Тема 10. Линейные модели и деревья принятия решений.

Таблицы принятия решений. Деревья принятия решений. Листья и ветви. Обучение (построение) дерева. Условия ветвления. Отсечение ветвей. Глубина дерева и остановка.

Тема 11. Нейронные сети.

Искусственный нейрон. Искусственные нейронные сети. Входные, скрытые и выходные нейроны. Типы и структура нейронных сетей. Сети прямого распространения и рекуррентные сети. Архитектуры нейронных сетей. Обучение с учителем, без учителя и с подкреплением. Генетические алгоритмы.

Раздел 5. Технологии работы с большими объемами входных данных

Тема 12. Введение в нереляционные базы данных (NoSQL)

Языки разметки. XML. Структура языка XML. Элементы языка XML. JSON и XML - сходства и отличия. Способы генерации JSON документов. Технологии, использующие JSON-документы.

Тема 13. Технологии и реализации NoSQL

Серверные приложения. Хранение и доступ к большим данным. MongoDB. Стурктура базы данных. Команды MongoDB. Создание записей. Изменение записей. Удаление записей. Выбор записей.

Тема 14. Технологии реального времени в BigData

Параллельные вычисления и обработка данных в режиме реального времени. Кластеры и OpenMPI. Параллельные вычисления на OpenMPI. Элементы программы и управления процессами. GPU и CUDA. Параллельные вычисления на CUDA. MapReduce. Операции Map и Reduce. Реализация алгоритма сбора статистики на примере MapReduce для MongoDb.

Распределенные системы. Hadoop. Распределенные файловые системы. HDFS. Распределенные базы данных. HBASE.

Тема 15. Финальный проект.

Разработка собственной системы анализа BigData на доступных наборах больших данных (например http://www.datasciencecentral.com/profiles/blogs/big-data-sets-available-for-free; https://catalog.data.gov/dataset).

5.2 Разделы дисциплины и междисциплинарные связи с обеспечиваемыми (последующими) дисциплинами

Пока нет

5.3. Разделы и темы дисциплин (модулей)и виды занятий

№ п/п	Наименование раздела (темы) Наименование темы	Виды занятий в часах					
		Лекц.	Практ зан.	Семин	Лаб.	СРС	Всего
	Раздел 1. Введение	**2**	**3**			**8**	**13**
1.	Тема 1. Проверка базовых знаний и вводная лекция	1	1			3	**5**
2.	Тема 2. Введение в Питон	1	2			5	**8**
3.	**Раздел 2. Дата майнинг**	**4**	**8**			**20**	**32**
4.	Тема 3. Протоколы и форматы интернета	1	2			5	**8**
5.	Тема 4. Протоколы и форматы баз данных	1	2			5	**8**
6.	Тема 5. Подготовка к анализу текстов	1	2			5	**8**
7.	Тема 6. Подготовка к анализу соотношений	1	2			5	**8**
8.	**Раздел 3. Предварительная обработка больших данных**	**3**	7			**18**	**28**
9.	Тема 7. Принятие простых решений	1	2			5	**8**
10.	Тема 8. Выделение явных закономерностей	1	2			5	**8**
11.	Тема 9. Выявление скрытых закономерностей.	1	3			8	**12**
12.	**Раздел 4. Экспертные системы**	**4**	**6**			**16**	**26**
13.	Тема 10. Линейные модели и деревья принятия решений.	1	2			6	**9**
14.	Тема 11. Нейронные сети.	3	4			10	**17**
15.	**Раздел 5. Технологии работы с большими объемами входных данных**	**5**	**12**			**28**	**45**
16.	Тема 12. Введение в нереляционные базы данных (NoSQL)	1	2			5	**8**
17.	Тема 13. Технологии и	1	4			9	**14**

	реализации NoSQL						
18.	Тема 14.Технологии реального времени в BigData	2	4			9	**15**
19.	Тема 15. Финальный проект	1	2			5	**8**
20.	ИТОГО	**18**	**36**			**90**	**144**

6. Перечень семинарских, практических занятий и лабораторных работ

№ п/п	№ раздела и темы дисциплины (модуля)	Наименование семинаров, практических и лабораторных работ	Трудоемкость (час.)	Оценоч-ные средства	Формируемые компетенции
1	**2**	**3**	**4**	**5**	**6**
1.	Р1.Т1	Тема 1.Проверка базовых знаний и вводная лекция	1	ПЗ	ПК-2, ПК-8, ПК-15, ПК-24
2.	Р1.Т2.	Тема 2.Введение в Питон	2	ПЗ	ПК-2, ПК-8
3.	Р2.Т3.	Тема 3. Протоколы и форматы интернета	2	ПЗ	ПК-2, ПК-8, ПК-15, ПК-23
4.	Р2.Т4.	Тема 4. Протоколы и форматы баз данных	2	ПЗ	ПК-2, ПК-8, ПК-15, ПК-23
5.	Р2.Т5.	Тема 5. Подготовка к анализу текстов	2	ПЗ	ПК-2, ПК-8, ПК-23, ПК-24
6.	Р2.Т6.	Тема 6. Подготовка к анализу соотношений	2	ПЗ	ПК-2, ПК-8, ПК-24
7.	Р3.Т7.	Тема 7.Принятие простых решений	2	ПЗ	ПК-2, ПК-8, ПК-23
8.	Р3.Т8.	Тема 8. Выделение явных закономерностей	2	ПЗ	ПК-2, ПК-8, ПК-23
9.	Р3.Т9.	Тема 9. Выявление скрытых закономерностей.	3	ПЗ	ПК-2, ПК-8, ПК-23
10.	Р4.Т10.	Тема 10. Линейные модели и деревья принятия решений.	2	ПЗ	ПК-2, ПК-8 ПК-12, ПК-23
11.	Р4.Т11.	Тема 11. Нейронные сети.	4	ПЗ	ПК-2, ПК-8, ПК-23
12.	Р5.Т12.	Тема 12. Введение в нереляционные базы данных (NoSQL)	2	ПЗ	ПК-2, ПК-8, ПК-11
13.	Р5.Т13.	Тема 13. Технологии и реализации NoSQL	4	ПЗ	ПК-2, ПК-8, ПК-15
14.	Р5.Т14.	Тема 14.Технологии реального времени в BigData	4	ПЗ	ПК-2, ПК-8 ПК-15 ПК-23, ПК-24
15.	Р5.Т15.	Тема 15.Финальный проект	2	ПЗ	ПК-2, ПК-8

					ПК-15 ПК-23, ПК-24

6.1. План самостоятельной работы студентов

Самостоятельная работа студентов предполагается в виде:

- изучения отдельных вопросов тематического плана дисциплины;
- подготовка к практическим занятиям;
- выполнение домашних заданий из практических работ;
- подготовка к экзамену;
- выполнение проекта;
- подготовка к зачету.

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомен-дуемая литера-тура	Коли-чество часов
1.	Р1.Т1	Выполнение практических заданий	Домашнее задание к практической работе 1. Ответить на вопросы.	СИР [1-3]	3
2.	Р1.Т2.	Выполнение практических заданий	Домашнее задание к практической работе 2. Ответить на вопросы.	СИР [1-3]	5
3.	Р2.Т3.	Выполнение практических заданий	Домашнее задание к практической работе 3. Ответить на вопросы.	СИР [1-3]	5
4.	Р2.Т4.	Выполнение практических заданий	Домашнее задание к практической работе 4.	СИР [1-3]	5
5.	Р2.Т5.	Выполнение практических заданий	Домашнее задание к практической работе 5. Ответить на вопросы.	СИР [1-3]	5
6.	Р2.Т6.	Выполнение практических заданий	Домашнее задание к практической работе 6. Ответить на вопросы.	СИР [1-3]	5
7.	Р3.Т7.	Выполнение практических заданий	Домашнее задание к практической работе 7. Ответить на вопросы.	СИР [1-3]	5
8.	Р3.Т8.	Выполнение практических заданий	Домашнее задание к практической работе 8. Ответить на вопросы.	СИР [1-3]	5
9.	Р3.Т9.	Выполнение практических заданий	Домашнее задание к практической работе 9. Ответить на вопросы.	СИР [1-3]	8
10.	Р4.Т10.	Выполнение практических заданий	Домашнее задание к практической работе 10. Ответить на вопросы.	СИР [1-3]	6
11.	Р4.Т11.	Выполнение	Домашнее задание к практической		10

		практических заданий	работе 11. Ответить на вопросы.	СИР [1-3]	
12.	Р5.Т12.	Выполнение практических заданий	Домашнее задание к практической работе 12. Ответить на вопросы.	СИР [1-3]	5
13.	Р5.Т13.	Выполнение практических заданий	Домашнее задание к практической работе 13. Ответить на вопросы.	СИР [1-3]	9
14.	Р5.Т14.	Выполнение практических заданий	Домашнее задание к практической работе 14. Ответить на вопросы.	СИР [1-3]	9
15.	Р5.Т15.	Выполнение практических заданий	Домашнее задание к практической работе 15. Ответить на вопросы.	СИР [1-3]	5

Вопросы для самоподготовки по разделу 1.

1. Привести примеры Big-data в повседневной жизни и придумать их возможное применение на практике
2. Считать файл на языке Python в виде последовательности бит и вывести в другой файл его биты в обратной последовательности. В двух вариантах - файл имеет малый размер (помещается в оперативную память), и не помещается в оперативную память.

Вопросы для самоподготовки по разделу 2.

3. Разделить программно заданный текст на предложения. В каждом предложении посчитать количество слов и оценить количество собственных имен. Найти самое часто используемое слово.
4. Из списка сайтов найти самый влиятельный - на который чаще всего ссылаются остальные. Построить схему ссылок.
5. Существует большая база данных с несколькими текстовыми полями, в каждом из которых записан большой текст (рассказ). Предложить способ реогранизации базы данных таким образом, чтобы было возможно производить быстрый контекстный поиск. Написать программу на Питон для такой перестройки базы.

Вопросы для самоподготовки по разделу 3.

6. Построить распределение средней длины слова в тексте в двух наборах текстов - английских и русских. Построить Байесовский фильтр и определить язык неизвестного текста по известной средней длине слова в нем. Оценить вероятность ошибки.
7. Известны положения спутника в трехмерной системе координат. В предположении, что он движется по эллипсу, определить параметры эллипса.

Вопросы для самоподготовки по разделу 4.

10. На основе обучения на базовых текстах серии построить систему определения возможных участников диалога в неизвестном тексте этой серии.

11. Разработать и обучить нейронную сеть, позволяющую для заданного набора известных текстов известных годов издания оценить год, страну и возможного автора неизвестного текста.

Вопросы для самоподготовки по разделу 5.

12. По KML-файлу, полученному с базы данных землетрясений earthquake.usgs.gov определить основные центры землетрясений и их размеры.

13. Написать программу для решения задачи вычисления площади сложной 3D-фигуры, заданной формой граней (заданных координатами вершин), на многопроцессорной системе.

6.2. Методические указания по организации самостоятельной работы студентов

Самостоятельная работа студентов в рамках изучения дисциплины «Параллельные вычисления и большие данные» регламентируется общим графиком учебной работы, предусматривающим посещение практических занятий и регулярное выполнение заданий по ним, выполнение домашних заданий.

При организации самостоятельной работы по дисциплине «Параллельные вычисления и большие данные» студенту следует:

1. Внимательно изучить материалы, характеризующие курс и тематику самостоятельного изучения, что изложено в учебно-методическом комплексе по дисциплине. Это позволит четко представить, как круг изучаемых тем, так и глубину их постижения.

2. Составить подборку литературы и источников, достаточную для изучения предлагаемых тем. В программе дисциплины представлены основной и дополнительный списки литературы. Они носят рекомендательный характер, это означает, что всегда есть литература, которая может не входить в данный список, но является необходимой для освоения темы. При этом следует иметь в виду, что нужна литература различных видов: учебники, учебные и учебно-методические пособия; первоисточники, монографии, сборники научных статей, публикации в журналах, любой эмпирический материал; справочная литература – энциклопедии, словари, тематические, терминологические справочники, раскрывающие категориально- понятийный аппарат, информационные ресурсы сети Интернет.

3. Основное содержание той или иной проблемы следует уяснить, изучая учебную литературу и источники.

4. Абсолютное большинство задач носит не только теоретический, умозрительный характер, но самым непосредственным образом выходят на жизнь, они тесно связаны с практикой. Это предполагает наличие у студентов не только знания категорий и понятий, но и умения использовать их в качестве инструмента для анализа и выполнения практических задач. Иными словами, студент должен совершать собственные, интеллектуальные усилия и постоянно решать практические задачи, а не только механически заучивать понятия и положения.

7. Примерная тематика курсовых работ (проектов) (при наличии)

Курсовая работа по данной дисциплине не предусмотрена учебным планом

8. Учебно-методическое и информационное обеспечение дисциплины (модуля):

1) Основная литература

Бриллинджер Д. Временные ряды. Обработка данных и теория. М.: Мир, 1980 г. - 536 с.

2) сетевые информационные ресурсы (СИР)

1. Курс MIT https://www.edx.org/course/introduction-computational-thinking-data-mitx-6-00-2x-5
2. Курс TeamDEV https://www.youtube.com/watch?v=fIZ64zHC6sU
3. Открытый курс машинного обучения. https://habrahabr.ru/company/ods/blog/322626/
4. http://jsman.ru/mongo-book/ (https://github.com/karlseguin/the-little-mongodb-book)
5. https://www.ibm.com/developerworks/ru/library/l-hadoop-1/index.html

9. Материально-техническое обеспечение дисциплины (модуля):

Для успешного освоения предлагаемых тем необходима специализированная аудитория, оснащенная спецоборудованием для проведения лекционных занятий (средства мультимедиа). Практические занятия проводятся в компьютерном классе, объединенном в локальную сеть, желательно с выходом в Интернет. Компьютеры с установленной ОС Linux и программным обеспечением Python. Желательно наличие в компьютерах видеокарт NVIDIA либо/и нескольких процессорных ядер.

10. Образовательные технологии:

Структурно-логические, тренинговые, компьютерные образовательные технологии, проектная работа в малых группах.

11. Оценочные средства (ОС):

11.1. Оценочные средства для входного контроля (могут быть в виде тестов с закрытыми или открытыми вопросами).

№ п\п	Вид контроля	Компетенции, компоненты которых контролируются
1.	Тест	ПК-2, ПК-8, ПК-15, ПК-23, ПК-24
2.	Индивидуальный проект	ПК-2, ПК-8, ПК-15, ПК-23, ПК-24

Примерные задания оценочного теста для входного контроля

1. Считать файл на языке Python в виде последовательности бит и вывести в другой файл его биты в обратной последовательности. В двух вариантах - файл имеет малый размер (помещается в оперативную память), и не помещается в оперативную память.

2. Написать программу на Python, поворачивающую 2D фигуру, заданную своими вершинами, на заданный угол.

3. В организации была база данных дачных участков, включающая в себя их линейные размеры (периметр) и площади. Все участки квадратные и их было достаточно много (>10000). База случайно была утеряна, но осталась известна средняя площадь участков (100 кв.м.) и их средний периметр (32м). Можно-ли по этим данным оценить дисперсию периметра?

4.Условия задачи 3. Можно-ли по этим данным в приближении нормального распределения периметров оценить, какая доля всех участков имела периметр больше 56м?

5. Найти матрицу, обратную к матрице

[2, 0]
[0, 2]

6. Колесо начинает крутиться от неподвижного состояния с постоянным ускорением. За первую минуту оно делает 10 оборотов. Сколько оборотов оно сделает за третью минуту?

11.2. Оценочные средства текущего контроля.

Материалы для проведения текущего контроля знаний студентов

№ п\п	Вид контроля	Контролируемые темы (разделы)	Компетенции, компоненты которых контролируются
1.	Практические задания	Разделы 1-5	ПК-2, ПК-8, ПК-15, ПК-23, ПК-24
2.	Тест	Разделы 1-5	ПК-2, ПК-8, ПК-15, ПК-23, ПК-24

Примерный вариант тестовых заданий по основным темам.

1. Какие из перечисленных систем являются системами, описываемыми термином 'Big Data':

а) Файл-лог доступа к популярному сайту (например lenta.ru) за 7 дней - время, адрес клиента, адрес просмотренной страницы

б) Полнотекстовая необновляемая или редкообновляемая библиотека электронных текстов

в) Полнотекстовая частообновляемая библиотека электронных текстов

г) Система поиска по автору в полнотекстовой необновляемой библиотеке электронных текстов

д) Система поиска по автору в полнотекстовой обновляемой библиотеке электронных текстов

е) Система контекстного поиска по полнотекстовой необновляемой библиотеке электронных текстов

ё) Система контекстного поиска по полнотекстовой обновляемой библиотеке электронных текстов

2. Найти ошибку в программе

```
import  numpy
import hem
fileSrc='input.dat';
fileDest='output.dat';
fileout=open(fileDest, mode='rb')
with open(fileSrc, mode='wb') as file:
fileContent = file.read()
filelen=len(fileContent);
src2hemming=hem.HemmingEncodeMatrix();
for i in range(0,filelen):
data=ord(fileContent[i])
hi4=hem.str2arr(hem.int2bin(data/16),4)
hemming_hi4=hem.norm2(hem.matmult(hi4,src2hemming),1,7)
lo4=hem.str2arr(hem.int2bin(data%16),4)
hemming_lo4=hem.norm2(hem.matmult(lo4,src2hemming),1,7)
fileout.write(chr(hem.bin2int(hemming_hi4,7)%128))
fileout.write(chr(hem.bin2int(hemming_lo4,7)%128))
fileout.close;
```

3.Какой из указанных протоколов обеспечивает наибольшую защищенность передаваемых данных к подслушиванию:

анонимный ftp
ftp с авторизацией
http
http с авторизацией
https
https с авторизацией
scp
scp с авторизацией по ключу

4. Какой из указанных протоколов обеспечивает наибольшую сложность к автоматическому получению данных:

анонимный ftp
ftp с авторизацией
http
http с простой авторизацией
https
https с простой авторизацией
scp с авторизацией по паролю
scp с авторизацией по ключу

5.Данадостаточнобольшаятаблица

create table record (id integer auto_increment, str char(255),txt TEXT,unique key (id));

Расположите запросы в порядке времени, требуемого на выполнение:

SELECT count(*) FROM records
SELECT count(*) FROM records ORDER BY id DESC
SELECT count(*) FROM records ORDER BY str DESC
Обоснуйте.

6.Даныдостаточнобольшиетаблицы

record1 (id1 integer, str CHAR(255), txt TEXT, UNIQUE KEY(id));
record2 (id2 integer, id1 integer, str CHAR(255), txt TEXT, UNIQUE KEY(id));

Расположите запросы в порядке времени, требуемого на выполнение:

SELECT records1.*,records2.* FROM records1,records2 WHERE records1.id1=records2.id1
SELECT records1.*,records2.* FROM records1,records2 WHERE records1.id1=records2.id1 ORDER BY records1.id1 DESC
SELECT records1.*,records2.* FROM records1 JOIN records2 ON records1.id1=records2.id1
SELECT records1.*,records2.* FROM records1,records2 JOIN records2 ON records1.id1=records2.id1 ORDER BY records1.id1 DESC
Обоснуйте.

7. Дана строка s:

s='Мама мыла раму. Раму мыла мама.'

Производится поиск по регулярному выражению:

([\w\s]*[\.\?])

Какие символы в этом регулярном выражении не используется при анализе этой строки?

8. Дана строка s:

s='Мама мыла раму. Раму мыла мама.'

Производится поиск по регулярному выражению:

([\w\s]*[\.\?])

Как изменить это регулярное выражение, чтобы исчезли 2 последние символа каждого результата?

9. Составить граф отношения 'является предком' для стандартной семьи: 2 бабушки, 2 дедушки, отец, мать, сын, дочь. Какая вершина имеет наибольшее число ребер?

10. Вероятность какого события больше - выпадение орла на стандартной монете или выпадение четного числа на стандартной игральной кости?

11. Даны 2 числовых ряда: 1,8,7,2,5,4,3 и 0.4, 3, 2.5, 0.5, 1.5, 1.5, 1.1

Рассчитать методом наименьших квадратов регрессионную зависимость одного ряда от другого. Определить, чему будет равно следующее значение второго числового ряда, если следующее значение первого числового ряда 12. А каких пределах может находится следующее значение?

12. Выявление скрытых закономерностей.

Даны 5 рядов:

13,27,4,-6,-20,17,-16,-2,9,4

7,9,6,-8,-13,5,-7,-6,8,-3

11,-17,-9,14,9,-12,16,9,-1,7

4,-26,-15,22,22,-17,23,15,-9,10

1,-9,8,-10,-6,-7,2,-10,7,-10

Выяснить, сколько и какие из них являются линейно независимыми от других.

13. Вашей экспертной системе необходимо проверить, пойдет-ли школьник в армию. Постройте примерное дерево принятия решения для решения этой задачи.

14. Вам необходимо построить генератор на нейронной сети, а именно: после подачи на ее вход двух входных импульсов, вы должны получить периодически изменяющийся, бексонечный сигнал на выходе. Какого типа нейронная сеть подходит для этой задачи. Почему?

15. Имеется JSON объект:

```
{   "firstName": "Иван";    "lastName": "Иванов";
"address": {"streetAddress": "Московское ш., 101, кв.101"; "city": "Ленинград";        "postalCode": 101101    },
    "phoneNumbers": [  "812 123-1234";  "916 123-4567"   ]
    }
```

Найти ошибку в описании и перевести объект в XML

16.Имеются SQL ТАБЛИЦЫ:

user (user_id, name, family, age);

comments (comment_id,user_id,comment);

содержимоетаблиц:
user:
(1,'Ivan','Ivanov',20)
(2,'Petr','Pterov',30)
comments:
(1,1,'Hello!')
(2,2,'Hello!')
(3,1,'Bye!')

Перевести записи таблиц в JSON-объект
Тема 13. Технологии и реализации NoSQL
Серверные приложения. Хранение и доступ к большим данным. MongoDB.

17. Дана программа code.py, запускаемая на кластере:

```
from mpi4py import MPI
comm = MPI.COMM_WORLD
rank = comm.Get_rank()
size = comm.Get_size()
if rank == 0:
res=0
for i in range(1,size) :
data1 = comm.recv(source=i, tag=12)
res += data1['a']
print "result: %d \n" %res
elif rank > 0 :
data={'a':rank}
comm.send(data, dest=0, tag=12)

Запускается строкой:
mpirun -np5 pythoncode.py
```

Что выдаст программа и почему?

18. Верно-ли утверждение: «В технологии MapReduce два основных шага - Map и Reduce »

Примерный вариант практических заданий по основным темам.

1. Известно, что в различных видах связи существуют номера-посредники (передатчик сообщения передает сообщение посреднику, посредник передает сообщение приемнику сообщения). Допустим, есть постоянно пополняемая база данных телефонных звонков (номер передатчика, номер приемника, время звонка). Написать быстрый алгоритм для нахождения посредника, а также передатчиков и приемников, пользующихся их услугами.

2. У вас существует база автомобильных GPS-логов робокаров в неизвестном вам городе в формате время - положение - уникальный ID машины(номер) - скорость. Написать программу, определяющую наиболее нагруженную трассу, ее траекторию и координаты. Определить динамику количества машин в одной из точек этой трассы, определить возможное наличие на ней пробок.

3. По базе данных оцифрованных текстов составить словарь (слова, встречающиеся достаточно большое количество раз). Выявить и локализовать возможные опечатки в часто встречающихся словах. Опечаткой считать слова, похожие на часто употребимые, но отличающиеся от них одной измененной буквой.

4. Написать программу, которая по заданным адресам web-страниц скачивает их и

строит по их содержимому словарь наиболее часто упоминаемых адресов на другие сайты. Найти самый популярный сайт, исключая собственный.

5. Два неопытных лучника стреляют по двум мишеням на большой стене, сначала разноцветными стрелами (каждый лучник стрелой своего цвета - красный и синий), потом одинаковыми (белыми). Хотя целятся в мишень, но стреляют неточно. В файле находятся координаты попадания стрелы в стену и цвет стрелы. На основе известных распределений двух случайных величин (значения заданы в файлах), построить байесовский критерий определения порога принятия решения, оценить его ошибку. Построить байесовский алгоритм определения лучника по координатам попадания стрелы. Оценить, сколько выстрелов белыми стрелами сделал первый(красный) лучник, а сколько - второй(синий), оценить ошибку. Определить примерные координаты мишеней каждого.

6. По KML-файлу, полученному с базы данных землетрясений earthquake.usgs.gov определить основные центры землетрясений и их размеры.

7. Построить распределение средней длины слова в тексте в двух наборах текстов - английских и русских. Построить Байесовский фильтр и определить язык неизвестного текста по известной средней длине слова в нем. Оценить вероятность ошибки.

8. Построить распределение средней длины слова в тексте в двух наборах текстов - английских и русских (например, http://lib.ru/PROZA/ (для русских) и https://www.gutenberg.org/wiki/Category:Fiction_Bookshelf (для английских)). Построить Байесовский фильтр и определить язык неизвестного текста по известной средней длине слова в нем. Оценить вероятность ошибки.

9. По данным библиотечных текстов (например http://lib.ru/PROZA/) составить список популярных имен и упорядочить его в порядке популярности.

10. Написать программу для решения задачи вычисления площади сложной 3D-фигуры, заданной формой граней (заданных координатами вершин), на многопроцессорной системе.

11. В базе данных находятся данные о лекарствах, покупаемых в различных аптеках РФ, включающие в себя название лекарства, его тип (жаропонижающее, антибиотик, биодобавка, гормональное, регулятор давления, ранозаживляющее, косметика и т.д.), регион РФ и время его приобретения. Написать программу для определения среднего уровня потребления лекарств в каждом регионе и выдачи признаков возникновения эпидемий гриппа.

12. В базе данных находятся данные об эффективности рекламы в виде: тип рекламы, вид продукта, год рождения клиента, возраст, пол, регион, местное время, уникальный ID клиента (в случае наличия), эффективность (0-показано, 1-перейдено по ссылке, 2-произведена покупка). Определить по базе наиболее важные параметры, влияющие на эффективность рекламы.

13. Черный ящик инопланетян производит неизвестную операцию с четырьмя числами. На основе известного набора исходных данных и результатов (обучающий набор, приводится в файле) создать и обучить нейронную сеть, производящую похожую операцию.

11.3. Оценочные средства для промежуточной аттестации (в форме экзамена или зачета).

Материалы для проведения промежуточного контроля знаний студентов:

№ п\п	Вид контроля	Контролируемые темы (разделы)	Компетенции, компоненты которых контролируются
1	БРС	Разделы 1-5	ПК-2, ПК-6, ПК-8, ПК-15,ПК-24, ПК-23
2	Зачет	Разделы 1-5	ПК-2, ПК-6, ПК-8, ПК-15,ПК-24, ПК-23

В качестве оценочных средств для промежуточного контроля выступают результаты сдачи заданий текущей аттестации, оформленные в соответствии с бально-рейтинговой системой.

Правила начисления баллов БРС по дисциплине «Параллельные вычисления и большие данные»

1. Посещаемость занятий

Устанавливается следующее соответствие посещаемости занятий (% от общего числа академических часов по дисциплине) баллам БРС:

- менее 50% занятий – 0 баллов;
- 50%-85% занятий – 5 баллов;
- 85%-100% занятий – 10 баллов;

В случае если студент посетил менее 50% от общего числа академических часов по данной дисциплине по уважительной причине (болезни), для компенсации знаний преподавателем может быть назначено такому студенту дополнительное задание При этом баллы БРС не начисляются.

2. Активность студентов

Активность студента на занятии предполагает выполнение студентом сверх предусмотренных учебно-методическим комплексом обязательных заданий по дисциплине следующих видов работ:

Виды активности	**Баллы**
Активное обсуждение на лекциях вопросов, поднимаемых преподавателем, решение задач.	**1 балл** БРС за одно лекционное занятие (2 академических часа)
Успешное выступление на лекционном или семинарском занятии с презентацией и докладом по теме, одобренной преподавателем	Данный вид работы оценивается в **3 балла** БРС за одно выступление

3. Рубежный контроль

Рубежный контроль по данной дисциплине осуществляется на основании своевременной сдачи отчетов о выполненных практических работах.

Баллы БРС присваиваются следующим образом:

- 30 баллов – все практические работы выполнены в срок, в полном объеме, все работы достойны отличной оценки;
- 25 баллов – все практические работы выполнены в срок, в полном объеме, все работы в среднем достойны хорошей оценки;
- 20 баллов – все практические работы выполнены в срок, в полном объеме, все работы в среднем достойны удовлетворительной оценки;
- 10 баллов – все практические работы выполнены в срок, в не полном объеме (не менее 75% заданий), все работы в среднем достойны оценки не ниже хорошей;
- 0 баллов – все практические работы выполнены в срок, в полном объеме, все работы в среднем достойны неудовлетворительной оценки.

До 20 баллов студент может получить на итоговом тестировании по основным темам разделов 1 и 2 курса дисциплины.

От 20 до 40 баллов студент может получить при ответе на вопросы экзаменационного билета (раздел 1, семестр 3).

От 20 до 40 баллов студент может получить за выполнение итогового проекта (раздел 2, семестр 4).

4. Самостоятельная работа

Баллы БРС присваиваются следующим образом:

- 20 баллов – самостоятельная работа выполнена в срок, в полном объеме, все работы достойны отличной оценки;
- 15 баллов – самостоятельная работа выполнена в срок, в полном объеме, все работы в среднем достойны хорошей оценки;
- 10 баллов – самостоятельная работа выполнена в срок, в полном объеме, все работы в среднем достойны удовлетворительной оценки;
- 5 баллов – самостоятельная работа выполнена в срок, в не полном объеме (не менее 75% заданий), все работы в среднем достойны оценки не ниже хорошей;
- 0 баллов – самостоятельная работа выполнена в срок, в полном объеме, все работы в среднем достойны неудовлетворительной оценки.

5. Штрафные баллы

БРС предусматривает штрафные баллы за нарушение сроков сдачи практических и самостоятельных работ и за низкую дисциплину на занятии.

Штрафные баллы за нарушение сроков сдачи практических и самостоятельных работ составляют 5 баллов по каждой работе. Штрафные баллы суммируются, и вычитаются из баллов, начисляемых за практическую или самостоятельную работу соответственно.

Штрафные баллы (от1 до 5) за низкую дисциплину на занятии могут быть начислены студенту преподавателем за опоздание более чем на 10 минут, посторонние разговоры во время занятий, за использование мобильных устройств, неэтичное поведение при общении с другими студентами или преподавателем, использование нецензурных выражений, нарушения требований, установленных преподавателем к поведению на занятии.

Ниже приведены критерии выставления зачета и дифференциального зачет по данной дисциплине, исходя из количества заработанных студентом баллов.

Баллы полученные обучающимся в течении семестра	Академическая оценка	
60…70	удовлетворительно	зачтено
71…85	хорошо	
86…100	отлично	

Задания для подготовки к экзамену по дисциплине
«Параллельные вычисления и большие данные»

Итоговый зачетный проект (4 семестр)

ПРОЕКТ

Каждый студент в период квалификационной практики выбирает тему индивидуального проекта. В ходе выполнения задания студент должен:

1. Описать структуры и функции выбранной предметной области.

2. Описать, если требуется, дополнительное используемое программное обеспечение.
3. Выполнить разработку математической модели решаемой задачи.
4. Освоить методику решения поставленной задачи
5. Составить алгоритм решения задачи
6. Освоить программирование в среде Python.
7. Выполнить задачу по программирование в среде Python.
8. Выполнить тестирование и отладку программы
9. Оценить точности решения
10. Защита проекта (презентация с отчетом о выполнении пунктов 1-8, демонстрация приложения и кода)

Примерные направления разработки проектов по дисциплине

Задание к контрольной работе

Провести анализ задачи построения экспертной системы из заданной предметной области, изучить возможности ее решения, спроектировать модель системы, программно реализовать систему на Python.

Список типовых заданий на разработку системы:

1. На основе обучения на базовых текстах серии романов (на языке оригинала) построить систему определения участников анонимного диалога в неизвестном отрывке этой серии.

2. Построить систему, позволяющую для заданного набора известных текстов известных годов издания оценить век, страну и возможного автора неизвестного текста.

3. На основе базы данных Центрабанка (http://www.cbr.ru/currency_base/) разработать систему прогноза курса одной из валют по курсу остальных.

4. На основе базы данных комментариев Facebook (https://github.com/jbencina/facebook-news) построить бота, сочиняющего комментарии.

5. На основе базы данных IMDB (https://data.world/popculture/imdb-5000-movie-dataset) создать сервис, снимающий идеальное кино - выбирающего для заданного жанра: директора, остальных главных актеров и ключевые слова, a) максимизирующие прибыль от такого фильма и b) максимизирующего рейтинг такого фильма.

Пояснительная записка к контрольной работе должна включать:

Постановку задачи, включающую требования к функционалу системы, к программному и программному обеспечению, необходимому для ее работы, описанию исходных данных и ожидаемых результатов

Описание проекта разрабатываемой системы в виде структуры программы, идеи решения проблемы, используемых технических решений, метода отладки.

Исходный код программной реализации программной системы на объектно-ориентированном языке программирования.

Описание метода проверки решения, оценочные точности(если возможно).

К отчету по контрольной работе допускаются студенты, продемонстрировавшие работу программной системы. На отчете студент должен ответить на вопросы преподавателя по архитектуре, алгоритмам и программной реализации предложенного решения.

Разработчики:

_____	доцент	В.П. Лебедев
(подпись)	(занимаемая должность)	(инициалы, фамилия)

Программа рассмотрена на заседании кафедры естественнонаучных дисциплин
(наименование)
от « 06» марта 2018г.
Протокол № 9

Зав. кафедрой_____ А.Г. Балахчи

Настоящая программа не может быть воспроизведена ни в какой форме без предварительного письменного разрешения кафедры-разработчика программы.