



**Частное учреждение высшего образования  
«Высшая школа предпринимательства (институт)»  
(ЧУВО «ВШП»)**

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
Б1.В.05 «Технологии обработки больших данных»**

**Направление подготовки: 09.04.03 Прикладная информатика**

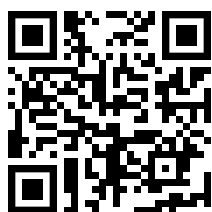
**Направленность (профиль) программы магистратуры  
«Прикладная информатика и информатизация»**

**ОДОБРЕНО**

Ученым советом ЧУВО «ВШП»  
Протокол заседания  
№01-02/24 от 30 августа 2025 г.

**УТВЕРЖДАЮ**

Ректор ЧУВО «ВШП»  
30 августа 2025 г.  
Аллабян М.Г.



Документ подписан электронной цифровой подписью  
VSHP EDS GEN 1, уникальный ключ документа:

**8F30-29EE-EB2F-GNI5**

Организация: ЧУВО «ВШП», ИНН 6903013604  
Дата подписания: 30.08.2025  
Подписал: Аллабян М.Г.

Рабочая программа учебной дисциплины **Б1.В.05 Технологии обработки больших данных**, обязательного компонента основной профессиональной образовательной программы высшего образования - программы магистратуры по направлению подготовки **09.04.03 Прикладная информатика** направленность (профиль) **«Прикладная информатика и информатизация»**, направлена на обеспечение у обучающегося способности осуществлять профессиональную деятельность в соответствующей области и сферах профессиональной деятельности, в том числе на их практическую подготовку с учётом рабочей программы воспитания и календарного плана воспитательной работы Частном учреждении высшего образования **«Высшая школа предпринимательства (институт)»** (далее — **ЧУВО «ВШП»**).

# 1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения программы магистратуры обучающийся должен овладеть следующими результатами обучения по дисциплине:

Код	Результаты освоения ООП (Содержание компетенций)	Индикаторы достижения	Перечень планируемых результатов обучения по дисциплине
УК-4	Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	УК-4.1. Выбирает на государственном и иностранном (-ых) языках коммуникативно приемлемые стили делового общения.	<p><b>Знать:</b> особенности корректного коммуникативного поведения; стили делового общения.</p> <p><b>Уметь:</b> использовать различные виды устной и письменной речи в деловом общении; применять коммуникативно приемлемые стили делового общения на государственном и иностранном (-ых) языках</p>
		УК-4.2. Использует информационно-коммуникационные технологии при поиске необходимой информации в процессе решения различных коммуникативных задач на государственном и иностранном (-ых) языках.	<p><b>Уметь:</b> коммуникативно и культурно приемлемо вести деловые разговоры в рамках академического и профессионального взаимодействия на государственном и иностранном (-ых) языках.</p>
ПК-4	Способен проектировать информационные процессы и системы с использованием инновационных инструментальных средств ИС	ПК- 4.1 Знает инструменты и методы проектирования структур баз данных, интерфейсов, программных модулей.	<p><b>Знать:</b> предметную область автоматизации и возможности ИС; инструменты и методы коммуникаций; стандарты информационного взаимодействия систем; современные методы управления организацией</p> <p><b>Уметь:</b> распределять работы и выделять ресурсы в области ИТ; проводить переговоры с различными участниками проекта; проводить реинжиниринг прикладных и информационных</p>

			процессов; принимать решения по информатизации предприятий в условиях неопределенности.
		ПК- 4.2 Демонстрирует умение проектировать информационные процессы и системы; проводить переговоры с заинтересованными сторонами проекта.	<p><b>Знать:</b> программные средства и платформы инфраструктуры информационных технологий организаций; инструменты и методы определения финансовых и производственных показателей организаций</p> <p><b>Уметь:</b> использовать инновационные подходы к проектированию ИС; управлять проектами ИС на всех стадиях жизненного цикла, оценивать эффективность и качество проекта; применять современные методы управления проектами и сервисами ИС.</p>
		ПК- 4.3 Демонстрирует навыки проектирования информационных процессов с использованием инновационных инструментальных средств ИС.	<p><b>Знать:</b> предметную область автоматизации и возможности ИС; инструменты и методы коммуникаций; стандарты информационного взаимодействия систем; современные методы управления организацией</p> <p><b>Уметь:</b> распределять работы и выделять ресурсы в области ИТ; проводить переговоры с различными участниками проекта; проводить реинжиниринг прикладных и информационных процессов; принимать решения по информатизации предприятий в условиях неопределенности.</p>

## 2. Распределение часов дисциплины по семестрам

ОФО

Семестр (курс)	3 семестр (2)
Виды деятельности	
лекционные занятия	10
лабораторные занятия	14
практические занятия/ семинарские занятия	-
руководство курсовой работой	-
клинические практические занятия (практическая подготовка)	-
контактная работа на выполнение курсового проекта	-
практическая подготовка	-
консультация перед экзаменом	2
самостоятельная работа	82
промежуточная аттестация	36
общая трудоемкость	144

## 3. Структура, тематический план и содержание учебной дисциплины

	лекционные занятия	лабораторные занятия	самостоятельная работа	формы текущего контроля
	О	О	О	
	Ф	Ф	Ф	
	О	О	О	
<b>Раздел: Введение в большие данные</b>	2	2	18	эссе
<b>Тема раздела: Введение в большие данные</b> Краткая история науки о данных: история сбора и анализа данных. Методы анализа данных: статистическое моделирование, машинное обучение, вычислительные подходы к моделированию. Статистические пределы добычи данных, принцип Бонферрони. Что такое большие данные. Источники больших данных. Масштабируемость подходов, границы масштабируемости.				
<b>Раздел: Обзор распределенных сред Hadoop и Spark</b>	2	4	22	лабораторная работа практическая работа
<b>Тема раздела: Обзор распределенных сред Hadoop и Spark</b> Введение в распределенные среды. От автономной машины к набору узлов. Экосистема Hadoop:				

архитектура, распределенная файловая система HDFS, вычислительная парадигма MapReduce, менеджер ресурсов YARN. Платформа Spark и библиотека pySpark.

**Тема раздела: Подробнее о Hadoop**

Структурные элементы Hadoop: NameNode, DataNode, Secondary NameNode, JobTracker, TaskTracker. Настройка SSH для кластера Hadoop: определение общей учетной записи, проверка правильности установки, генерация пары ключей, распространение открытого ключа, проверка входа. Запуск Hadoop в различных режимах.

<b>Раздел: MapReduce</b>	4	4	22	лабораторная работа письменная работа практическая работа
--------------------------	---	---	----	---

**Тема раздела: Введение в MapReduce**

Распределенные файловые системы: физическая организация вычислительных узлов, организация больших файловых систем. Задачи-распределители, группировка по ключу. Задачи-редукторы. Комбинаторы. Детали выполнения MapReduce. Обработка отказа узлов.

**Тема раздела: Примеры алгоритмов с использованием MapReduce**

Умножение матрицы на вектор больших размеров с применением MapReduce. Операции реляционной алгебры. Вычисление выборки и проекции с помощью MapReduce. Вычисление естественного соединения, группировки и агрегирования.

**Тема раздела: Обобщения MapReduce. Теория сложности MapReduce**

Системы потоков работ. Рекурсивные обобщения MapReduce. Система Pregel. Модель коммуникационной стоимости. Коммуникационная стоимость для сетевых задач. Физическое время. Многопутевое соединение. Размер редукции и коэффициент репликации. Пример: соединение по сходству. Графовая модель для проблем MapReduce. Схема сопоставления. Действия при отсутствии части входов. Нижняя граница коэффициента репликации. Примеры.

<b>Раздел: Практическое машинное обучение в среде Spark</b>	2	4	20	лабораторная работа
---	---	---	----	------------------------

**Тема раздела: Распространение переменных по узлам кластера в Spark**

Настройка виртуальной машины. Широковещательные переменные только для чтения. Аккумуляторные переменные только для записи. Примеры использования широкоовещательных и аккумуляторных переменных.

**Тема раздела: Машинное обучение с платформой Spark**

Платформа Spark на наборе данных KDD99. Чтение набора данных. Конструирование признаков. Выбор и обучение модели. Возможности конвейерного машинного обучения. Кросс-валидация.

<b>Итого часов</b>	<b>10</b>	<b>14</b>	<b>82</b>	
--------------------	-----------	-----------	-----------	--

#### 4. Формы текущего контроля

- эссе (шкала: значение от 0 до 8, количество: 1)

раздел дисциплины: Введение в большие данные

**Примерное задание:**

Примерные темы эссе по теме "Введение в большие данные"

1. Что такое Big Data?
2. История развития отрасли Big Data
3. Источники больших данных: интернет и мобильные устройства
4. Источники больших данных: корпоративные данные
5. Источники больших данных: результаты научных наблюдений
6. Рынки решений для управления большими данными
7. Использование больших данных в маркетинге
8. Аналитика больших данных, отличия от традиционной аналитики
9. Функции и задачи больших данных
10. Сервисы Big Data

- лабораторная работа (шкала: значение от 0 до 10, количество: 1)  
раздел дисциплины: Обзор распределенных сред Hadoop и Spark

**Примерное задание:**

Пример лабораторной работы по теме «MapReduce»

Цели:

- Формирование навыков работы с платформой MapReduce.
- Реализация алгоритмов реляционной алгебры для MapReduce.

Описание работы:

В варианте реляционной алгебры, реализованном в SQL, отношения описываются не над множествами, а над мультимножествами, то есть допускаются одинаковые кортежи. Реализуйте алгоритмы MapReduce для операций объединения, пересечения и разности множеств. Множества при этом рассматривайте как мультимножества. Пусть операции выполняются с мультимножествами R и S:

- а) объединение множеств – мультимножество, для которых количество вхождений кортежа  $t$  равно сумме двух чисел: количество его вхождений в R и S.
- б) пересечение множеств – мультимножество, для которого количество вхождений  $t$  равно минимуму количеств вхождений в R и S.
- с) пересечение множеств – мультимножество, для которого количество вхождений  $t$  равно разности количеств вхождений в R и S. Если  $t$  входит в S больше раз, чем в R, то  $t$  не включается в разность.

Для мультимножеств определена также операция выборки. Придумайте и выполните MapReduce-реализацию, которая порождает правильное количество копий каждого кортежа  $t$ , удовлетворяющего условию выборки. Предложите такой способ организации пар ключ-значение, при котором было бы легко получить из значений правильный результат выборки.

- практическая работа (шкала: значение от 0 до 10, количество: 1)  
раздел дисциплины: Обзор распределенных сред Hadoop и Spark

**Примерное задание:**

Практическая работа по теме «MapReduce»

Обработка временных рядов

Цели работы:

- Формирование навыков работы с платформой MapReduce.
- Реализация алгоритмов обработки временных рядов для MapReduce.

Описание работы:

Рассмотрите временной ряд, в котором каждая запись снабжена временной меткой, интерпретируемой как ключ, и содержит результат некоторого измерения. На выходе необходимо получить результат применения к временному ряду линейной функции вида:

$$y(t) = a_0 * x(t) + a_1 * x(t-1) + a_2 * x(t-2) + \dots + a_N * x(t-N),$$

где  $t$  – время, а  $a_0, a_1, \dots, a_N$  – известные константы.

В теории обработки сигналов это называется фильтром с конечной импульсной характеристикой (КИХ-фильтр). Особенный интерес представляет частный случай скользящего среднего, когда  $a_0 = a_1 = a_2 = \dots = a_N = 1/N$ . В этом случае каждая точка  $y$  – это среднее  $N$  предыдущих точек. Такой прием можно рассматривать как простой способ сглаживания временных рядов.

Реализуйте этот линейный фильтр в виде MapReduce программы. Используйте комбинатор. Если временные ряды хронологически упорядочены (а так обычно и бывает) и  $N$  относительно невелико, то сокращение сетевого трафика при тасовании можно добиться за счет использования комбинатора?

В качестве дополнительного упражнения реализуйте свой разбиватель, гарантирующий, что выходные данные также будут отсортированы в хронологическом порядке.

- лабораторная работа (шкала: значение от 0 до 7, количество: 1)

раздел дисциплины: MapReduce

### **Примерное задание:**

Пример лабораторной работы по теме «MapReduce»

Цели:

- Формирование навыков работы с платформой MapReduce.
- Реализация алгоритмов реляционной алгебры для MapReduce.

Описание работы:

В варианте реляционной алгебры, реализованном в SQL, отношения описываются не над множествами, а над мультимножествами, то есть допускаются одинаковые кортежи. Реализуйте алгоритмы MapReduce для операций объединения, пересечения и разности множеств. Множества при этом рассматривайте как мультимножества. Пусть операции выполняются с мультимножествами  $R$  и  $S$ :

а) объединение множеств – мультимножество, для которых количество вхождений кортежа  $t$  равно сумме двух чисел: количество его вхождений в  $R$  и  $S$ .

б) пересечение множеств – мультимножество, для которого количество вхождений  $t$  равно минимуму количеств вхождений в  $R$  и  $S$ .



с) пересечение множеств – мультимножество, для которого количество вхождений  $t$  равно разности количеств вхождений в  $R$  и  $S$ . Если  $t$  входит в  $S$  больше раз, чем в  $R$ , то  $t$  не включается в разность.

Для мультимножеств определена также операция выборки. Придумайте и выполните MapReduce-реализацию, которая порождает правильное количество копий каждого кортежа  $t$ , удовлетворяющего условию выборки. Предложите такой способ организации пар ключ-значение, при котором было бы легко получить из значений правильный результат выборки.

- письменная работа (шкала: значение от 0 до 8, количество: 1)

раздел дисциплины: MapReduce

### **Примерное задание:**

Письменная работа по теме "MapReduce"

1. Для процедуры подсчета слов в большом репозитории используются 100 распределителей и 100 редукторов.
  - а. Следует ли ожидать значительного разброса времени обработки списков значений различными редукторами, если предположить, что комбинаторы в редукторах не используются. Объясните свой ответ.
  - б. Если поручить операции редукции небольшому числу редукторов, например 10, то следует ли ожидать существенного разброса времени вычислений? Объясните свой ответ.
  - с. Предположим, что в вычислениях используются 100 распределителей и комбинатор. Следует ли ожидать существенного разброса времени вычислений? Объясните свой ответ.
2. Допустим, что задание состоит из  $n$  задач, каждая из которых занимает  $t$  секунд. Следовательно, в отсутствие отказов общее время, которое потребуется узлам для выполнения всех задач составляет  $nt$ . Предположим, что вероятность отказа задачи равна  $p$ . При отказе задачи накладные расходы добавляют ко времени выполнения задания  $10t$  секунд. Каково математическое ожидание полного времени выполнения задачи?
3. Вычислите коэффициент Жаккара для каждой пары следующих трех множеств:  $\{1,2,3,4\}$ ,  $\{2,3,5,7\}$ ,  $\{2,4,6\}$
4. Пусть универсальное множество  $U$  состоит из  $n$  элементов, из него случайным образом выбираются подмножества  $S$  и  $T$  по  $m$  элементов в каждом. Вычислите математическое ожидание коэффициента Жаккара  $S$  и  $T$

- практическая работа (шкала: значение от 0 до 10, количество: 1)

раздел дисциплины: MapReduce

### **Примерное задание:**

Практическая работа по теме «MapReduce»

Обработка временных рядов

Цели работы:

- Формирование навыков работы с платформой MapReduce.
- Реализация алгоритмов обработки временных рядов для MapReduce.

Описание работы:

Рассмотрите временной ряд, в котором каждая запись снабжена временной меткой, интерпретируемой как ключ, и содержит результат некоторого измерения. На выходе необходимо получить результат применения к временному ряду линейной функции вида:

$$y(t) = a_0 * x(t) + a_1 * x(t-1) + a_2 * x(t-2) + \dots + a_N * x(t-N),$$

где  $t$  – время, а  $a_0, a_1, \dots, a_N$  – известные константы.

В теории обработки сигналов это называется фильтром с конечной импульсной характеристикой (КИХ-фильтр). Особенный интерес представляет частный случай скользящего среднего, когда  $a_0 = a_1 = a_2 = \dots = a_N = 1/N$ . В этом случае каждая точка  $y$  – это среднее  $N$  предыдущих точек. Такой прием можно рассматривать как простой способ сглаживания временных рядов.

Реализуйте этот линейный фильтр в виде MapReduce программы. Используйте комбинатор. Если временные ряды хронологически упорядочены (а так обычно и бывает) и  $N$  относительно невелико, то сокращение сетевого трафика при тасовании можно добиться за счет использования комбинатора?

В качестве дополнительного упражнения реализуйте свой разбиватель, гарантирующий, что выходные данные также будут отсортированы в хронологическом порядке.

- лабораторная работа (шкала: значение от 0 до 7, количество: 1)

раздел дисциплины: Практическое машинное обучение в среде Spark

#### **Примерное задание:**

Роль языков программирования Python и R в аналитике больших данных. Необходимый набор библиотек. Готовые решения анализа

Проблема переобучения и регуляризация. Разбор алгоритма нейронных сетей. Разбор алгоритма SVM

### **5. Формы промежуточной аттестации**

- экзамен - 2 курс, 3 семестр (шкала: значение от 0 до 40)

#### **Примерное задание:**

Вопросы к зачету

1. Что такое большие данные.
2. Краткая история науки о данных: история сбора и анализа данных.
3. Методы анализа данных: статистическое моделирование, машинное обучение, вычислительные подходы к моделированию.
4. Статистические пределы добычи данных, принцип Бонферрони.
5. Источники больших данных.
6. Масштабируемость подходов, границы масштабируемости.
7. Введение в распределенные среды.
8. Экосистема Hadoop: архитектура, распределенная файловая система HDFS.
9. Экосистема Hadoop: вычислительная парадигма MapReduce, менеджер ресурсов YARN. Платформа Spark и библиотека pySpark.
10. Структурные элементы Hadoop: NameNode, DataNode, Secondary NameNode, JobTracker, TaskTracker.
11. Настройка SSH для кластера Hadoop: определение общей учетной записи, проверка правильности

установки, генерация пары ключей, распространение открытого ключа, проверка входа. Запуск Hadoop в различных режимах.

12. Распределенные файловые системы: физическая организация вычислительных узлов, организация больших файловых систем.

13. Задачи-распределители, группировка по ключу.

14. Задачи-редукторы.

15. Комбинаторы.

16. Обработка отказа узлов.

17. Умножение матрицы на вектор больших размеров с применением MapReduce. Операции реляционной алгебры.

18. Вычисление выборки и проекции с помощью MapReduce. Вычисление естественного соединения, группировки и агрегирования.

19. Системы потоков работ. Рекурсивные обобщения MapReduce. Система Pregel.

20. Модель коммуникационной стоимости.

21. Коммуникационная стоимость для сетевых задач.

22. Размер редукции и коэффициент репликации. Пример: соединение по сходству.

23. Графовая модель для проблем MapReduce.

24. Практическое машинное обучение в среде Spark: общая схема.

25. Spark: распространение переменных по узлам кластера

26. Spark: настройка виртуальной машины.

27. Spark: широковебательные переменные только для чтения.

28. Spark: аккумуляторные переменные только для записи.

29. Spark : примеры использования широковебательных и аккумуляторных переменных.

#### **Критерии оценивания:**

13-25 баллов: обучающийся свободно ориентируется в материале, дает обстоятельные глубокие ответы на все поставленные вопросы; демонстрирует хорошее знание понятийно-категориального аппарата изучаемой образовательной области (учебной дисциплины); умеет анализировать проблемы по дисциплине; высказывает собственную точку зрения на раскрываемые проблемы; четко грамотно формулирует свои мысли; демонстрирует учебные умения и навыки в области решения практико-ориентированных задач

0-12 баллов: обучающийся демонстрирует поверхностные знания материала, затрудняется в ответах на вопросы; не знает сущности основных понятий изучаемой образовательной области (учебной дисциплины); испытывает трудности в анализе проблем по дисциплине.

### **6. Балльная система оценивания по дисциплине**

ОФО

Семестр (Курс) - 3 (2)			
Форма текущего контроля	Раздел дисциплины	Максимальный балл	Максимальный приведенный балл
лабораторная работа	MapReduce	7	
лабораторная	Обзор распределенных сред	10	

работа	Hadoop и Spark		
лабораторная работа	Практическое машинное обучение в среде Spark	7	
письменная работа	MapReduce	8	
практическая работа	MapReduce	10	
практическая работа	Обзор распределенных сред Hadoop и Spark	10	
эссе	Введение в большие данные	8	
Максимальный текущий балл		60	60
<b>Промежуточная аттестация</b>		экзамен	
Максимальный аттестационный балл		40	40
Общий балл по дисциплине		100	100

Общий балл по дисциплине за семестр складывается из результатов, полученных по формам текущего контроля в течение семестра и аттестационного балла.

Оценка успеваемости по дисциплине в семестре пересчитывается по приведенной 100-балльной шкале независимо от шкалы, определенной преподавателем.

Перевод баллов из 100-балльной шкалы в числовой и буквенный эквивалент:

**- для экзамена, зачета с оценкой, курсовой работы (форма контроля из учебного плана):**

Сумма баллов	Отметка	Буквенный эквивалент
86-100	5	Отлично
66-85	4	Хорошо
51-65	3	Удовлетворительно
0-50	2	Неудовлетворительно

## 7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины. Электронно-библиотечные системы

### *Основная литература*

1. Анализ больших данных : учебное пособие / И. Б. Тесленко, В. Е. Крылов, А. М. Губернаторов [и др.]. — Москва : КноРус, 2024. — 295 с. — ISBN 978-5-406-14006-2. — URL: <https://book.ru/book/955989> — Текст : электронный.

### *Дополнительная литература*

1. Соловьев, В. И., Анализ данных в экономике: Теория вероятностей, прикладная статистика, обработка и анализ данных в Microsoft Excel. : учебник / В. И. Соловьев. — Москва : КноРус, 2025. — 497 с. — ISBN 978-5-406-13693-5. — URL: <https://book.ru/book/955517> — Текст : электронный.

2. Интеллектуальное право в условиях развития технологии Big Data. База данных как объект интеллектуальных и иных прав : Монография / Е.А. Войниканис, М.А. Кольздорф, В.А. Корнеев [и др.] — Москва : Проспект, 2022. — 177 с. — ISBN 978-5-392-36441-1. — URL: <https://book.ru/book/948395> — Текст : электронный.

## **8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем**

Обучающимся (магистрам) обеспечен доступ (удаленный доступ), в том числе в случае применения электронного обучения, дистанционных образовательных технологий, к современным профессиональным базам данных и информационным справочным системам (*подлежащим обновлению при необходимости*), а именно:

1. КонсультантПлюс: справочно-поисковая система [Электронный ресурс]. - <http://www.consultant.ru>
2. Мировая цифровая библиотека: <http://wdl.org/ru>
3. Научная электронная библиотека «Scopus»: <https://www.scopus.com>
4. Научная электронная библиотека ScienceDirect: <http://www.sciencedirect.com>
5. Научная электронная библиотека «eLIBRARY»: <https://elibrary.ru>
6. Портал «Гуманитарное образование» <http://www.humanities.edu.ru>
7. Федеральный портал «Российское образование» <http://www.edu.ru>
8. Федеральное хранилище «Единая коллекция цифровых образовательных ресурсов» <http://school-collection.edu.ru>
9. Поисковые системы Yandex, Rambler и др.
10. Электронная библиотека Российской Государственной Библиотеки (РГБ): <http://elibrary.rsl.ru>
11. Электронно-библиотечная система <http://www.sciteclibrary.ru>

## **9. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине**

Помещения для проведения всех видов работы, предусмотренных учебным планом, укомплектованы необходимым оборудованием и техническими средствами обучения.

<b>Наименование оборудованных учебных кабинетов, объектов для проведения практических занятий, объектов физической культуры и спорта с перечнем основного оборудования</b>	<b>Адрес (местоположение ) учебных кабинетов, объектов для проведения практических занятий, объектов физической культуры и спорта (с указанием площади и номера помещения в соответствии с документами бюро технической инвентаризации)</b>	<b>Собственность или оперативное управление, хозяйственное ведение, аренда (субаренда), безвозмездное пользование, практическая подготовка</b>	<b>Полное наименование собственника (арендодателя, ссудодателя) объекта недвижимого имущества</b>	<b>Документ – основание возникновен ия права (реквизиты и срок действия)</b>
<b>Специализирова нная многофункциона льная учебная аудитория для проведения учебных занятий лекционного и семинарского типов, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, в том числе, для организации практической подготовки обучающийся, с перечнем основного оборудования (аудитория № 3):</b> Столы для обучающихся; Стулья для обучающихся;	170001, Тверская область, г. Тверь, ул. Спартака, д. 26а (39,2 кв.м., 1 этаж, помещение № 3)	Безвозмездное пользование	Богачев Сергей Александрович	Договор безвозмездно о пользования недвижимым имуществом № 01-18/Н от 01.11.2020 с приложениям и №№ 1-3; срок действия договора: с 01.11.2020 по 30.09.2025

<p>Стол педагогического работника;</p> <p>Стул педагогического работника;</p> <p>Компьютер с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду лицензиата;</p> <p>Интерактивная доска;</p> <p>Проектор</p>				
<p><b>Специализированная многофункциональная учебная аудитория для проведения учебных занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, в том числе, для организации практической подготовки обучающийся, с перечнем основного оборудования (аудитория № 27)</b></p> <p>Компьютерные столы для обучающихся;</p> <p>Стулья для обучающихся;</p>	<p>170001, Тверская область, г. Тверь, ул. Спартака, д. 26а (31,1 кв.м., 2 этаж, помещение № 27)</p>	<p>Безвозмездное пользование</p>	<p>Богачев Сергей Александрович</p>	<p>Договор безвозмездного пользования недвижимым имуществом № 01-18/Н от 01.11.2020 с приложениям и №№ 1-3; срок действия договора: с 01.11.2020 по 30.09.2025</p>

<p>Стол педагогического работника;</p> <p>Стул педагогического работника;</p> <p>Компьютеры с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду лицензиата;</p> <p>Интерактивная доска;</p> <p>Проектор</p> <p>Сканер;</p> <p>Принтер</p>				
<p><b>Специализированная многофункциональная учебная аудитория для проведения учебных занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, в том числе, для организации практической подготовки обучающихся, с перечнем основного оборудования (аудитория № 16)</b></p> <p>Компьютерные столы для обучающихся;</p> <p>Стулья для обучающихся;</p>	<p>170001, Тверская область, г. Тверь, ул. Спартака, д. 26а (31,4 кв.м., 2 этаж, помещение № 16)</p>	<p>Безвозмездное пользование</p>	<p>Богачев Сергей Александрович</p>	<p>Договор безвозмездного пользования недвижимым имуществом № 01-18/Н от 01.11.2020 с приложениями №№ 1-3; срок действия договора: с 01.11.2020 по 30.09.2025</p>



<p>Стол педагогического работника;</p> <p>Стул педагогического работника;</p> <p>Компьютеры с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду лицензиата;</p> <p>Интерактивная доска;</p> <p>Проектор</p> <p>Сканер;</p> <p>Принтер</p>				
<p><b>Помещение для самостоятельной работы обучающихся с перечнем основного оборудования (аудитория № 22):</b></p> <p>Столы для обучающихся;</p> <p>Стулья для обучающихся;</p> <p>Компьютеры с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду лицензиата;</p> <p>Ноутбуки с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду лицензиата;</p>	<p>170001, Тверская область, г. Тверь, ул. Спартака, д. 26а (19,3 кв.м., 2 этаж, помещение № 22)</p>	<p>Безвозмездное пользование</p>	<p>Богачев Сергей Александрович</p>	<p>Договор безвозмездного пользования недвижимым имуществом № 01-18/Н от 01.11.2020 с приложениями №№ 1-3; срок действия договора: с 01.11.2020 по 30.09.2025</p>

бразовательную среду лицензиата; Принтер; Сканер				
<b>Помещение для самостоятельной работы обучающихся с перечнем основного оборудования (аудитория № 14):</b> Столы для обучающихся; Стулья для обучающихся; Компьютеры с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду лицензиата; Ноутбуки с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду лицензиата; Принтер; Сканер	170001, Тверская область, г. Тверь, ул. Спартака, д. 26а (22,5 кв.м., 1 этаж, помещение № 14)	Безвозмездное пользование	Богачев Сергей Александрович	Договор безвозмездного пользования недвижимым имуществом № 01-18/Н от 01.11.2020 с приложениями №№ 1-3; срок действия договора: с 01.11.2020 по 30.09.2025
<b>Помещение для самостоятельной работы обучающихся с перечнем основного оборудования (аудитория № 31):</b> Столы для обучающихся; Стулья для обучающихся;	170001, Тверская область, г. Тверь, ул. Спартака, д. 26а (20,3 кв.м., 2 этаж, помещение № 31)	Безвозмездное пользование	Богачев Сергей Александрович	Договор безвозмездного пользования недвижимым имуществом № 01-18/Н от 01.11.2020 с приложениями №№ 1-3; срок действия договора:

Компьютеры с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду лицензиата; Ноутбуки с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду лицензиата; Принтер; Сканер				с 01.11.2020 по 30.09.2025
--	--	--	--	-------------------------------

## 10. Образовательные технологии

Наименование образовательной технологии	Краткая характеристика
Дифференцированное обучение	Технология обучения, целью которой является создание оптимальных условий для выявления задатков, развития интересов и способностей обучающихся через разделение на группы, подразумевает наличие разных уровней учебных требований к группам в овладении ими содержанием образования.
Модульное обучение	Дисциплина структурирована по отдельным блокам, в которых учебное содержание и технология овладения объединены в систему, сопровождается контролем знаний и умений студентов, позволяет изучать дисциплину в индивидуальном темпе с учетом уровня базовой подготовки обучающихся.

## 11. Особенности реализации дисциплины для инвалидов и лиц с ограниченными возможностями здоровья

Для осуществления процедур текущего контроля успеваемости и промежуточной аттестации обучающихся созданы фонды оценочных средств, адаптированные для инвалидов и лиц с ограниченными возможностями здоровья и позволяющие оценить достижение ими запланированных в основной образовательной программе результатов обучения и уровень сформированности всех компетенций, заявленных в образовательной программе. Форма проведения текущей аттестации для

студентов-инвалидов устанавливается с учетом индивидуальных психофизических особенностей (устно, письменно на бумаге, письменно на компьютере, в форме тестирования и т.п.). При тестировании для слабовидящих студентов используются фонды оценочных средств с укрупненным шрифтом. На экзамен приглашается сопровождающий, который обеспечивает техническое сопровождение студенту. При необходимости студенту-инвалиду предоставляется дополнительное время для подготовки ответа на экзамене (или зачете). Обучающиеся с ограниченными возможностями здоровья и обучающиеся инвалиды обеспечиваются печатными и электронными образовательными ресурсами (программы, учебники, учебные пособия материалы для самостоятельной работы и т.д.) в формах, адаптированных к ограничениям их здоровья и восприятия информации:

1) для инвалидов и лиц с ограниченными возможностями здоровья **по зрению:**

- **для слепых:** задания для выполнения на семинарах и практических занятиях оформляются рельефно-точечным шрифтом Брайля или в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением для слепых, либо зачитываются ассистентом; письменные задания выполняются на бумаге рельефно-точечным шрифтом Брайля или на компьютере со специализированным программным обеспечением для слепых либо надиктовываются ассистенту; обучающимся для выполнения задания при необходимости предоставляется комплект письменных принадлежностей и бумага для письма рельефно-точечным шрифтом Брайля, компьютер со специализированным программным обеспечением для слепых;
- **для слабовидящих:** обеспечивается индивидуальное равномерное освещение не менее 300 люкс; обучающимся для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств; задания для выполнения заданий оформляются увеличенным шрифтом;

2) для инвалидов и лиц с ограниченными возможностями здоровья **по слуху:**

- **для глухих и слабослышащих:** обеспечивается наличие звукоусиливающей аппаратуры коллективного пользования, при необходимости обучающимся предоставляется звукоусиливающая аппаратура индивидуального пользования; предоставляются услуги сурдопереводчика;
- **для слепоглухих** допускается присутствие ассистента, оказывающего услуги тифлосурдопереводчика (помимо требований, выполняемых соответственно для слепых и глухих);

3) для лиц с тяжелыми нарушениями речи, глухих, слабослышащих лекции и семинары, проводимые в устной форме, проводятся в письменной форме;

4) для инвалидов и лиц с ограниченными возможностями здоровья, **имеющих нарушения опорно-двигательного аппарата:**

- для лиц с нарушениями опорно-двигательного аппарата, нарушениями двигательных функций верхних конечностей или отсутствием верхних конечностей: письменные задания выполняются на компьютере со специализированным программным обеспечением или надиктовываются ассистенту; выполнение заданий (тестов, контрольных работ), проводимые в письменной форме, проводятся в устной форме путем опроса, беседы с обучающимся.



**Частное учреждение высшего образования  
«Высшая школа предпринимательства (институт)»  
(ЧУВО «ВШП»)**

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ  
по дисциплине  
Б1.В.05 «Технологии обработки больших данных»**

**Направление подготовки: 09.04.03 Прикладная информатика**

**Направленность (профиль) программы магистратуры  
«Прикладная информатика и информатизация»**

**Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы**

В результате освоения программы магистратуры обучающийся должен овладеть следующими результатами обучения по дисциплине (модулю):

Код	Результаты освоения ООП (Содержание компетенций)	Индикаторы достижения	Перечень планируемых результатов обучения по дисциплине
УК-4	Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	УК-4.1. Выбирает на государственном и иностранном (-ых) языках коммуникативно приемлемые стили делового общения.	<p><b>Знать:</b> особенности корректного коммуникативного поведения; стили делового общения.</p> <p>П.ТВ1 П.ТВ2 П.ТВ3 П.ТВ4 П.ТВ5 П.ТВ6 П.ТВ7 П.ТВ8 П.ТВ9 П.ТВ10 П.ТВ11 П.ТВ12 П.ТВ13 П.ТВ14 П.ТВ15 П.ТВ16 П.ТВ17 П.ТВ18 П.ТВ19 П.ТВ20 П.ТВ21 П.ТВ22 П.ТВ23 П.ТВ24 П.ТВ25 П.ТВ26 П.ТВ27 П.ТВ28 П.ТВ29 П.ТВ30 П.ТВ31 П.ТВ32 П.ТВ33</p>

			П.ТВ34 П.ТВ35 П.ТВ36 Т.Э1_1 Т.Л1_2 Т.П1_2 Т.Л1_3 Т.ПР1_3 Т.П1_3 Т.Л1_4
		<b>Уметь:</b> использовать различные виды устной и письменной речи в деловом общении; применять коммуникативно приемлемые стили делового общения на государственном и иностранном (-ых) языках	П.ТВ1 П.ТВ2 П.ТВ3 П.ТВ4 П.ТВ5 П.ТВ6 П.ТВ7 П.ТВ8 П.ТВ9 П.ТВ10 П.ТВ11 П.ТВ12 П.ТВ13 П.ТВ14 П.ТВ15 П.ТВ16 П.ТВ17 П.ТВ18 П.ТВ19 П.ТВ20 П.ТВ21 П.ТВ22 П.ТВ23 П.ТВ24 П.ТВ25 П.ТВ26 П.ТВ27 П.ТВ28 П.ТВ29 П.ТВ30 П.ТВ31

			П.ТВ32 П.ТВ33 П.ТВ34 П.ТВ35 П.ТВ36 Т.Э1_1 Т.Л1_2 Т.П1_2 Т.Л1_3 Т.ПР1_3 Т.П1_3 Т.Л1_4
		УК-4.2. Использует информационно-коммуникационные технологии при поиске необходимой информации в процессе решения различных коммуникативных задач на государственном и иностранном (-ых) языках.	<b>Уметь:</b> коммуникативно и культурно приемлемо вести деловые разговоры в рамках академического и профессионального взаимодействия на государственном и иностранном (-ых) языках.         П.ТВ1 П.ТВ2 П.ТВ3 П.ТВ4 П.ТВ5 П.ТВ6 П.ТВ7 П.ТВ8 П.ТВ9 П.ТВ10 П.ТВ11 П.ТВ12 П.ТВ13 П.ТВ14 П.ТВ15 П.ТВ16 П.ТВ17 П.ТВ18 П.ТВ19 П.ТВ20 П.ТВ21 П.ТВ22 П.ТВ23 П.ТВ24 П.ТВ25 П.ТВ26 П.ТВ27



			П.ТВ28 П.ТВ29 П.ТВ30 П.ТВ31 П.ТВ32 П.ТВ33 П.ТВ34 П.ТВ35 П.ТВ36 Т.Э1_1 Т.Л1_2 Т.П1_2 Т.Л1_3 Т.ПР1_3 Т.П1_3 Т.Л1_4
ПК-4	Способен проектировать информационные процессы и системы с использованием инновационных инструментальных средств ИС	ПК- 4.1 Знает инструменты и методы проектирования структур баз данных, интерфейсов, программных модулей.	<b>Знать:</b> предметную область автоматизации и возможности ИС; инструменты и методы коммуникаций; стандарты информационного взаимодействия систем; современные методы управления организацией П.ТВ1 П.ТВ2 П.ТВ3 П.ТВ4 П.ТВ5 П.ТВ6 П.ТВ7 П.ТВ8 П.ТВ9 П.ТВ10 П.ТВ11 П.ТВ12 П.ТВ13 П.ТВ14 П.ТВ15 П.ТВ16 П.ТВ17 П.ТВ18 П.ТВ19 П.ТВ20 П.ТВ21 П.ТВ22 П.ТВ23

			П.ТВ24 П.ТВ25 П.ТВ26 П.ТВ27 П.ТВ28 П.ТВ29 П.ТВ30 П.ТВ31 П.ТВ32 П.ТВ33 П.ТВ34 П.ТВ35 П.ТВ36 Т.Э1_1 Т.Л1_2 Т.П1_2 Т.Л1_3 Т.ПР1_3 Т.П1_3 Т.Л1_4
		<b>Уметь:</b> распределять работы и выделять ресурсы в области ИТ; проводить переговоры с различными участниками проекта; проводить реинжиниринг прикладных и информационных процессов; принимать решения по информатизации предприятий в условиях неопределенности.	П.ТВ1 П.ТВ2 П.ТВ3 П.ТВ4 П.ТВ5 П.ТВ6 П.ТВ7 П.ТВ8 П.ТВ9 П.ТВ10 П.ТВ11 П.ТВ12 П.ТВ13 П.ТВ14 П.ТВ15 П.ТВ16 П.ТВ17 П.ТВ18 П.ТВ19 П.ТВ20 П.ТВ21

			П.ТB22 П.ТB23 П.ТB24 П.ТB25 П.ТB26 П.ТB27 П.ТB28 П.ТB29 П.ТB30 П.ТB31 П.ТB32 П.ТB33 П.ТB34 П.ТB35 П.ТB36 Т.Э1_1 Т.Л1_2 Т.П1_2 Т.Л1_3 Т.ПР1_3 Т.П1_3 Т.Л1_4
		ПК- 4.2 Демонстрирует умение проектировать информационные процессы и системы; проводить переговоры с заинтересованными сторонами проекта.	<b>Знать:</b> программные средства и платформы инфраструктуры информационных технологий организаций; инструменты и методы определения финансовых и производственных показателей организаций П.ТB1 П.ТB2 П.ТB3 П.ТB4 П.ТB5 П.ТB6 П.ТB7 П.ТB8 П.ТB9 П.ТB10 П.ТB11 П.ТB12 П.ТB13 П.ТB14 П.ТB15 П.ТB16 П.ТB17

			П.ТВ18 П.ТВ19 П.ТВ20 П.ТВ21 П.ТВ22 П.ТВ23 П.ТВ24 П.ТВ25 П.ТВ26 П.ТВ27 П.ТВ28 П.ТВ29 П.ТВ30 П.ТВ31 П.ТВ32 П.ТВ33 П.ТВ34 П.ТВ35 П.ТВ36 Т.Э1_1 Т.Л1_2 Т.П1_2 Т.Л1_3 Т.ПР1_3 Т.П1_3 Т.Л1_4
		<b>Уметь:</b> использовать инновационные подходы к проектированию ИС; управлять проектами ИС на всех стадиях жизненного цикла, оценивать эффективность и качество проекта; применять современные методы управления проектами и сервисами ИС.	П.ТВ1 П.ТВ2 П.ТВ3 П.ТВ4 П.ТВ5 П.ТВ6 П.ТВ7 П.ТВ8 П.ТВ9 П.ТВ10 П.ТВ11 П.ТВ12 П.ТВ13 П.ТВ14 П.ТВ15

			П.ТВ16 П.ТВ17 П.ТВ18 П.ТВ19 П.ТВ20 П.ТВ21 П.ТВ22 П.ТВ23 П.ТВ24 П.ТВ25 П.ТВ26 П.ТВ27 П.ТВ28 П.ТВ29 П.ТВ30 П.ТВ31 П.ТВ32 П.ТВ33 П.ТВ34 П.ТВ35 П.ТВ36 Т.Э1_1 Т.Л1_2 Т.П1_2 Т.Л1_3 Т.ПР1_3 Т.П1_3 Т.Л1_4
		ПК- 4.3 Демонстрирует навыки проектирования информационных процессов с использованием инновационных инструментальных средств ИС.	<b>Знать:</b> предметную область автоматизации и возможности ИС; инструменты и методы коммуникаций; стандарты информационного взаимодействия систем; современные методы управления организацией П.ТВ1 П.ТВ2 П.ТВ3 П.ТВ4 П.ТВ5 П.ТВ6 П.ТВ7 П.ТВ8 П.ТВ9 П.ТВ10 П.ТВ11

			П.ТВ12 П.ТВ13 П.ТВ14 П.ТВ15 П.ТВ16 П.ТВ17 П.ТВ18 П.ТВ19 П.ТВ20 П.ТВ21 П.ТВ22 П.ТВ23 П.ТВ24 П.ТВ25 П.ТВ26 П.ТВ27 П.ТВ28 П.ТВ29 П.ТВ30 П.ТВ31 П.ТВ32 П.ТВ33 П.ТВ34 П.ТВ35 П.ТВ36 Т.Э1_1 Т.Л1_2 Т.П1_2 Т.Л1_3 Т.ПР1_3 Т.П1_3 Т.Л1_4
		<b>Уметь:</b> распределять работы и выделять ресурсы в области ИТ; проводить переговоры с различными участниками проекта; проводить реинжиниринг прикладных и информационных процессов; принимать решения по	П.ТВ1 П.ТВ2 П.ТВ3 П.ТВ4 П.ТВ5 П.ТВ6 П.ТВ7 П.ТВ8 П.ТВ9

			информатизации предприятий в условиях неопределенности.	П.ТВ10 П.ТВ11 П.ТВ12 П.ТВ13 П.ТВ14 П.ТВ15 П.ТВ16 П.ТВ17 П.ТВ18 П.ТВ19 П.ТВ20 П.ТВ21 П.ТВ22 П.ТВ23 П.ТВ24 П.ТВ25 П.ТВ26 П.ТВ27 П.ТВ28 П.ТВ29 П.ТВ30 П.ТВ31 П.ТВ32 П.ТВ33 П.ТВ34 П.ТВ35 П.ТВ36 Т.Э1_1 Т.Л1_2 Т.П1_2 Т.Л1_3 Т.ПР1_3 Т.П1_3 Т.Л1_4
--	--	--	---	---

### Контрольные задания. Текущая аттестация

эссе - Введение в большие данные	Номер задания
1. Что такое Big Data? 2. История развития отрасли Big Data 3. Источники больших данных: интернет и мобильные устройства 4. Источники больших данных: корпоративные данные 5. Источники больших данных: результаты научных наблюдений 6. Рынки решений для управления большими данными 7. Использование больших данных в маркетинге 8. Аналитика больших данных, отличия от традиционной аналитики 9. Функции и задачи больших данных 10. Сервисы Big Data	Т.Э1_1

лабораторная работа - Обзор распределенных сред Hadoop и Spark	Номер задания
Корреляция. Регрессионный анализ. Задачи в области больших данных, решаемые методом регрессионного анализа. Постановка задачи классификации. Постановка задачи кластеризации. Задача построения ассоциативных правил.	Т.Л1_2

практическая работа - Обзор распределенных сред Hadoop и Spark	Номер задания
Обработка временных рядов Цели работы: - Формирование навыков работы с платформой MapReduce. - Реализация алгоритмов обработки временных рядов для MapReduce. Описание работы: Рассмотрим временной ряд, в котором каждая запись снабжена временной меткой, интерпретируемой как ключ, и содержит результат некоторого измерения. На выходе необходимо получить результат применения к временному ряду линейной функции вида: $y(t) = a_0 * x(t) + a_1 * x(t-1) + a_2 * x(t-2) + \dots + a_N * x(t-N),$ где $t$ – время, а $a_0, a_1, \dots, a_N$ – известные константы. В теории обработки сигналов это называется фильтром с конечной импульсной характеристикой (КИХ-фильтр). Особенный интерес представляет частный случай скользящего среднего, когда $a_0 = a_1 = a_2 = \dots = a_N = 1/N$ . В этом случае каждая точка $y$ – это среднее $N$ предыдущих точек. Такой прием можно рассматривать как простой способ сглаживания временных рядов.	Т.П1_2

лабораторная работа - MapReduce	Номер задания
Цели: - Формирование навыков работы с платформой MapReduce. - Реализация алгоритмов реляционной алгебры для MapReduce. Описание работы: В варианте реляционной алгебры, реализованном в SQL, отношения описываются не над множествами, а над мультимножествами, то есть допускаются одинаковые кортежи. Реализуйте	Т.Л1_3



<p>алгоритмы MapReduce для операций объединения, пересечения и разности множеств. Множества при этом рассматривайте как мультимножества. Пусть операции выполняются с мультимножествами R и S:</p> <p>а) объединение множеств – мультимножество, для которых количество вхождений кортежа t равно сумме двух чисел: количество его вхождений в R и S.</p> <p>б) пересечение множеств – мультимножество, для которого количество вхождений t равно минимуму количеств вхождений в R и S.</p> <p>с) пересечение множеств – мультимножество, для которого количество вхождений t равно разности количеств вхождений в R и S. Если t входит в S больше раз, чем в R, то t не включается в разность.</p> <p>Для мультимножеств определена также операция выборки. Придумайте и выполните MapReduce-реализацию, которая порождает правильное количество копий каждого кортежа t, удовлетворяющего условию выборки. Предложите такой способ организации пар ключ-значение, при котором было бы легко получить из значений правильный результат выборки.</p>	
--	--

письменная работа - MapReduce	Номер задания
<p>1. Для процедуры подсчета слов в большом репозитории используются 100 распределителей и 100 редукторов.</p> <p>а. Следует ли ожидать значительного разброса времени обработки списков значений различными редукторами, если предположить, что комбинаторы в редукторах не используются. Объясните свой ответ.</p> <p>б. Если поручить операции редукции небольшому числу редукторов, например 10, то следует ли ожидать существенного разброса времени вычислений? Объясните свой ответ.</p> <p>с. Предположим, что в вычислениях используются 100 распределителей и комбинатор. Следует ли ожидать существенного разброса времени вычислений? Объясните свой ответ.</p> <p>2. Допустим, что задание состоит из n задач, каждая из которых занимает t секунд. Следовательно, в отсутствие отказов общее время, которое потребуется узлам для выполнения всех задач составляет nt. Предположим, что вероятность отказа задачи равна p. При отказе задачи накладные расходы добавляют ко времени выполнения задания 10t секунд. Каково математическое ожидание полного времени выполнения задачи?</p> <p>3. Вычислите коэффициент Жаккара для каждой пары следующих трех множеств: {1,2,3,4}, {2,3,5,7}, {2,4,6}</p> <p>4. Пусть универсальное множество U состоит из n элементов, из него случайным образом выбираются подмножества S и T по m элементов в каждом. Вычислите математическое ожидание коэффициента Жаккара S и T</p>	Т.ПР1_3

практическая работа - MapReduce	Номер задания
<p>Реализуйте этот линейный фильтр в виде MapReduce программы. Используйте комбинатор. Если временные ряды хронологически упорядочены (а так обычно и бывает) и N относительно невелико, то сокращение сетевого трафика при тасовании можно добиться за счет использования комбинатора?</p> <p>В качестве дополнительного упражнения реализуйте свой разбиватель, гарантирующий, что выходные данные также будут отсортированы в хронологическом порядке.</p>	Т.П1_3

лабораторная работа - Практическое машинное обучение в среде Spark	Номер
--	-------

	задания
Роль языков программирования Python и R в аналитике больших данных. Необходимый набор библиотек. Готовые решения анализа Проблема переобучения и регуляризация. Разбор алгоритма нейронных сетей. Разбор алгоритма SVM	Т.Л1_4

### Контрольные задания. Промежуточная аттестация

Экзамен. Теоретический вопрос	Номер задания
Краткая история науки о данных: история сбора и анализа данных.	П.ТВ1
Методы анализа данных: статистическое моделирование, машинное обучение, вычислительные подходы к моделированию.	П.ТВ2
Статистические пределы добычи данных, принцип Бонферрони.	П.ТВ3
Что такое большие данные. Источники больших данных.	П.ТВ4
Экосистема Hadoop: архитектура, распределенная файловая система HDFS, вычислительная парадигма MapReduce, менеджер ресурсов YARN.	П.ТВ5
Платформа Spark и библиотека pySpark.	П.ТВ6
Структурные элементы Hadoop: NameNode, DataNode, Secondary NameNode, JobTracker, TaskTracker.	П.ТВ7
Настройка SSH для кластера Hadoop: определение общей учетной записи, проверка правильности установки, генерация пары ключей, распространение открытого ключа, проверка входа.	П.ТВ8
Распределенные файловые системы MapReduce: физическая организация вычислительных узлов, организация больших файловых систем.	П.ТВ9
Задачи-распределители в MapReduce, группировка по ключу. Задачи-редукторы. Комбинаторы.	П.ТВ10
Модель коммуникационной стоимости в MapReduce. Коммуникационная стоимость для сетевых задач. Физическое время. Многопутевое соединение.	П.ТВ11
Среда Spark: широковещательные переменные только для чтения, аккумуляторные переменные только для записи. Примеры использования широковещательных и аккумуляторных переменных.	П.ТВ12
Работа с объектами DataFrame в среде Spark.	П.ТВ13
Определение больших данных, ключевые характеристики. Примеры задач больших данных. Основные виды данных.	П.ТВ14
Дать краткую сравнительную характеристику инструментария ПО для анализа данных.	П.ТВ15
Охарактеризовать конструкции языка R. Перечислить типы языка R, привести примеры.	П.ТВ16
Роль аналитика по данным (Data Scientist). Ключевые компетенции аналитика. Отличия BI от Data Science.	П.ТВ17
«Жизненный цикл» проекта по аналитике больших данных. Перечислить основные роли исполнителей проекта.	П.ТВ18
Типовая архитектура проекта в области больших данных. Перечислить используемые технологии, указать степень вовлеченности каждой из технологий на каждом этапе работы	П.ТВ19

над проектом	
Что такое Data Mining? Этапы интеллектуального анализа данных	П.ТВ20
Основные задачи и методы Data Mining. Методы интеллектуального анализа данных	П.ТВ21
Что такое искусственный интеллект? Декатрон?	П.ТВ22
Роль гипотез в процессе познания. Какие факторы используются для уточнения гипотез?	П.ТВ23
Основные понятия статистики и дескриптивный анализ.	П.ТВ24
Шкалы измерений. Генеральная совокупность и выборка. Нормальное распределение. Уровень статистической достоверности.	П.ТВ25
Корреляция и регрессионный анализ. Коэффициент корреляции.	П.ТВ26
Графическое представление. Постановка задачи регрессионного анализа.	П.ТВ27
Пояснить термин "Линейная регрессия". Привести примеры использования регрессионного анализа.	П.ТВ28
Классификация и кластеризация – суть и назначение. Метрики. Постановка задачи кластеризации.	П.ТВ29
Методы кластеризации на графах. Отличие от задачи классификации. Привести примеры использования алгоритмов кластеризации.	П.ТВ30
Парадигма Map Reduce. Описать принцип работы. Нарисовать схему. Перечислить слабые и сильные стороны. Обозначить области применимости. Привести примеры использования.	П.ТВ31
Визуализация. Дать определение визуализации. Показать важность визуализации в аналитике больших данных. Привести примеры и инструменты для визуализации.	П.ТВ32
Научные проблемы больших данных. Показать значимость проблем, актуальность, связь с областями математики и инженерии.	П.ТВ33
OLAP и OLTP системы. Разница.	П.ТВ34
Репликация и шардинг.	П.ТВ35
Требования ACID. CAP-теорема, BASE архитектура NoSql. Классификация NoSql хранилищ. Их особенности. Примеры распределенных хранилищ.	П.ТВ36

#### Балльная система оценивания по дисциплине

ОФО

Семестр (Курс) - 3 (2)			
Форма текущего контроля	Раздел дисциплины	Максимальный балл	Максимальный приведенный балл
лабораторная работа	MapReduce	7	
лабораторная работа	Обзор распределенных сред Hadoop и Spark	10	
лабораторная	Практическое	7	

работа	машинное обучение в среде Spark		
письменная работа	MapReduce	8	
практическая работа	MapReduce	10	
практическая работа	Обзор распределенных сред Hadoop и Spark	10	
эссе	Введение в большие данные	8	
Максимальный текущий балл		60	60
<b>Промежуточная аттестация</b>		экзамен	
Максимальный аттестационный балл		40	40
Критерии оценивания		<p>13-25 баллов: обучающийся свободно ориентируется в материале, дает обстоятельные глубокие ответы на все поставленные вопросы; демонстрирует хорошее знание понятийно-категориального аппарата изучаемой образовательной области (учебной дисциплины); умеет анализировать проблемы по дисциплине; высказывает собственную точку зрения на раскрываемые проблемы; четко грамотно формулирует свои мысли; демонстрирует учебные умения и навыки в области решения практико-ориентированных задач</p> <p>0-12 баллов: обучающийся демонстрирует поверхностные знания материала, затрудняется в ответах на вопросы; не знает сущности основных понятий изучаемой образовательной области (учебной дисциплины); испытывает трудности в анализе проблем по дисциплине.</p>	
Общий балл по дисциплине		100	100

Общий балл по дисциплине за семестр складывается из результатов, полученных по формам текущего контроля в течение семестра и аттестационного балла.

Оценка успеваемости по дисциплине в семестре пересчитывается по приведенной 100-балльной шкале независимо от шкалы, определенной преподавателем.

Перевод баллов из 100-балльной шкалы в числовой и буквенный эквивалент:

**- для экзамена, зачета с оценкой, курсовой работы (форма контроля из учебного плана):**

Сумма баллов	Отметка	Буквенный эквивалент
86-100	5	Отлично

66-85	4	Хорошо
51-65	3	Удовлетворительно
0-50	2	Неудовлетворительно

### Список используемых сокращений

#### Текущая аттестация

Тип задания	Сокращение
внеаудиторное чтение	Т.В
доклад / конференция / реферат	Т.Д
индивидуальное задание (перевод / презентация / план урока / тезаурус / глоссарий / сценарий деловой игры / алгоритм задачи / программа / конспектирование научной литературы)	Т.И
итоговая лабораторная работа	Т.ЛР
кейс	Т.КС
коллоквиум	Т.К
контрольная работа	Т.КР
лабораторная работа	Т.Л
отчет (по научно-исследовательской работе / практике)	Т.О
письменная работа	Т.ПР
практическая работа	Т.П
расчетно-графическая работа	Т.РГ
семестровая работа	Т.СР
ситуационная задача / ситуационное задание / проект	Т.СЗ
творческая работа	Т.ТР
тест по итогам занятия	Т.Т
устный опрос / собеседование	Т.У
эссе	Т.Э

#### Промежуточная аттестация

Тип задания	Сокращение
-------------	------------

Практическое задание	П.П
Теоретический вопрос	П.ТВ
Тестовый вопрос	П.Т