

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Курганский государственный университет»
(КГУ)

Кафедра «Программное обеспечение автоматизированных систем»



УТВЕРЖДАЮ:
Первый проректор
Т.Р. Змызгова
«31» августа 2022 г.

Рабочая программа учебной дисциплины

МЕТОДЫ АНАЛИЗА БОЛЬШИХ ДАННЫХ

образовательной программы высшего образования –
программы магистратуры

09.04.04 Программная инженерия

Направленность:

**Методы и алгоритмы интеллектуальной обработки данных в информа-
ционно-вычислительных системах**

Форма обучения: очная

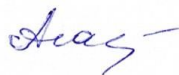
Курган 2022

Рабочая программа дисциплины «Методы анализа больших данных» составлена в соответствии с учебным планом по программе магистратуры «Программная инженерия» (Методы и алгоритмы интеллектуальной обработки данных в информационно-вычислительных системах), утвержденным для очной формы обучения «30» августа 2022 года.

Рабочая программа дисциплины одобрена на заседании кафедры «Программное обеспечение автоматизированных систем» «30» августа 2022 года, протокол № 1.

Рабочую программу составила:

Доцент кафедры
«Программное обеспечение
автоматизированных систем»



Н.В. Агапова

Согласовано:

Заведующий кафедрой
«Программное обеспечение
автоматизированных систем»



В.К. Волк

Специалист по учебно-методической
работе Учебно-методического отдела



Г.В. Казанкова

Начальник Управления образовательной
деятельности



И.В. Григоренко

1. ОБЪЕМ ДИСЦИПЛИНЫ

Всего: 5 зачетных единиц трудоемкости (180 академических часов)

Очная форма обучения

Вид учебной работы	На всю дисциплину	Семестр
		3
Аудиторные занятия (контактная работа с преподавателем), всего часов	72	72
в том числе:		
Лекции	24	24
Лабораторные работы	48	48
Аудиторные занятия в интерактивной форме, часов	-	-
Самостоятельная работа, всего часов	108	108
в том числе:		
Подготовка к экзамену	27	27
Другие виды самостоятельной работы	81	81
Вид промежуточной аттестации	экзамен	экзамен
Общая трудоемкость дисциплины и трудоемкость по семестрам, часов	180	180

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Методы анализа больших данных» относится к части блока 1, формируемой участниками образовательных отношений, модулю «Прикладные задачи интеллектуального анализа данных (элективный модуль)», дисциплина по выбору.

Программа составлена с учетом межпредметных связей с учебными дисциплинами. Основой для изучения учебной дисциплины являются следующие учебные дисциплины «Структуры и алгоритмы обработки данных», «Интеллектуальный анализ данных», «Управление данными», «Архитектуры информационно-вычислительных систем», «Методологии и технологии информационно-вычислительных систем», «Параллельное программирование и вычисления на GPU», «Язык Java», «Язык Kotlin», «Язык Haskell».

Учебная дисциплина «Методы анализа больших данных» знакомит студентов с методами хранения, сбора, обработки, выдачи больших данных.

Результаты обучения по дисциплине необходимы для изучения дисциплин: «Научно-исследовательская работа (производственная практика)» и выполнения выпускной квалификационной работы.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Цель преподавания учебной дисциплины состоит в формировании знаний умений и навыков в области разработки и администрирования программного обеспечения современных высокопроизводительных распределенных систем. В данном курсе рассматриваются программные технологии построения масштабируемых многомашинных информационно-вычислительных систем, обеспечивающих параллельную обработку сверхбольших массивов данных. За рубежом совокупность таких технологий обозначается термином Big Data (англ. - большие данные).

Рассматриваются также типовые методы и алгоритмы параллельной обработки сверхбольших массивов данных.

Компетенции, формируемые в результате освоения дисциплины:

Для 09.04.04:

- Владение методами планирования и обработки результатов экспериментальных исследований (ПК-2);

- Способность разрабатывать и использовать программное обеспечение для моделирования, анализа, распознавания и обработки информации, в том числе - в системах искусственного интеллекта (ПК-3)

В результате изучения дисциплины обучающийся должен:

Знать: математические методы обработки данных (для ПК-2, ПК-3).

Уметь: выполнять исследования процессов создания, накопления и обработки информации, включая анализ и создание моделей данных и знаний, языков их описания и манипулирования (для ПК-2, ПК-3).

Владеть: новыми методами исследования и обработки данных и их применению в самостоятельной научно-исследовательской деятельности в области профессиональной деятельности (для ПК-2, ПК-3).

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1. Учебно-тематический план

Очная форма обучения

№	Наименование раздела	Количество часов контактной работы с преподавателем	
		Лекции	Лабораторные работы
1	Что такое Data Mining? MapReduce	2	4
2	Поиск похожих объектов	2	4
3	Анализ потоков данных	2	4
4	Анализ ссылок	2	4
5	Частые предметные наборы	2	4
6	Кластеризация	2	4
7	Реклама в Интернете	2	4
8	Машинное обучение на больших данных	2	4
9	Современные BigData-решения и архитектуры	2	4
10	Технология обработки и хранения данных Hadoop	2	4
11	Системы разработки, сборки и доставки кода	2	8
12	Системы поточной обработки данных	2	-
Всего:		24	48

4.2. Содержание лекционных занятий

Тема 1. Что такое Data Mining? MapReduce

Вычислительные подходы к моделированию. Статистические пределы добычи данных. Принцип Бонферрони. Хэш-функции. Степенные зависимости. Распределенные файловые системы. Задачи-распределители. Алгоритмы, в которых используется MapReduce. Обобщения MapReduce. Рекурсивные обобщения MapReduce. Модель коммуникационной стоимости. Теория сложности MapReduce.

Тема 2. Поиск похожих объектов

Приложения поиска близкого соседям. Разбиение документов на шинглы. Сигнатуры множеств с сохранением сходства. Хэширование документов с учетом близости. Метрики. Теория функций, учитывающих близость. LSH-семейства для расстояния Жаккара. LSH-семейства для других метрик.

Тема 3. Анализ потоков данных

Потоковая модель данных. Выборка данных из потока. Фильтрация потоков. Подсчет различных элементов в потоке. Оценивание моментов. Подсчет единиц в окне. Затухающие окна. Нахождение самых популярных элементов.

Тема 4. Анализ ссылок

Определение PageRank. Эффективное вычисление PageRank. Тематический PageRank. Ссылочный спам. Хабы и авторитетные страницы.

Тема 5. Частые предметные наборы

Модель корзины покупок. Корзины покупок и алгоритм Apriori. Обработка больших наборов данных в оперативной памяти. Алгоритм с ограниченным числом проходов. Подсчет частых предметных наборов в потоке.

Тема 6. Кластеризация

Введение в методы кластеризации. Иерархическая кластеризация. Алгоритм k средних. Алгоритм CURE. Кластеризация в неевклидовых пространствах.

Тема 7. Реклама в Интернете

Проблемы онлайн-рекламы. Онлайн-алгоритмы. Задача о паросочетании. Задача о ключевых словах. Реализация алгоритма Adwords. Сопоставление предложений с поисковыми запросами

Тема 8. Машинное обучение на больших данных

Перцептроны. Обучение перцептрона с нулевым порогом. Алгоритм Winnow. Переменный порог. Многоклассовые перцептроны. Метод опорных векторов. Обучение по ближайшим соседям. Сравнение методов обучения. Модель машинного обучения. Перцептроны. Обучение перцептрона с нулевым порогом. Сходимость перцептронов. Алгоритм Winnow. Метод опорных векторов. Обучение по ближайшим соседям. Сравнение методов обучения

Тема 9. Современные BigData-решения и архитектуры

Введение. Актуальность. NoSQL системы – обзор и принцип работы. Обзор фреймворков. Ключевые метрики Big Data system. Основные компоненты BigData-систем. Типы обработки данных. Архитектура Data Lake. Архитектура Lambda. Архитектура Карра

Тема 10. Технология обработки и хранения данных Hadoop

История разработки и развития. Архитектурная концепция экосистемы Hadoop. Что такое HDFS? MapReduce на Hadoop. Компонент YARN. Различные компоненты поверх платформы

Тема 11. Системы разработки, сборки и доставки кода

Базовые компоненты классической системы очередей. Типичный жизненный цикл сообщений в системах очередей. Kafka и классические сервисы очередей. Структура данных. Consumer Groups. Consumer offsets. Apache ZooKeeper

Тема 12. Системы пакетной и поточной обработки данных

Фреймворк Apache Spark. Основные характеристики. Приложение Spark - состав. Обзор архитектуры Spark. Операции с RDD. Ленивые вычисления. Направленный ациклический граф. Циклы выполнения. Apache Flume. Apache Flume — передача данных в Hadoop. Apache Sqoop

4.3. Лабораторные работы

Номер раздела, темы	Наименование раздела	Наименование лабораторной работы	Норматив времени, час.
1	Что такое Data Mining? MapReduce	Извлечение данных	4
2	Поиск похожих объектов	Трансформация данных	4
3	Анализ потоков данных	Поиск похожих объектов	4
4	Анализ ссылок	Метод «ближайшего соседа»	4
5	Частые предметные наборы	Онлайновые алгоритмы. Сопоставление предложений с поисковыми запросами	4
6	Кластеризация	Кластеризация данных	2
Рубежный контроль 1			2
7	Реклама в Интернете	Сравнение методов обучения	4
8	Машинное обучение на больших данных	Установка Hadoop на виртуальную машину	4
9	Современные BigData-решения и архитектуры	Разработка MapReduce программы	4
10	Технология обработки и хранения данных Hadoop	Установка Spark + Sqoop программ	4
11	Системы разработки, сборки и доставки кода	Разработка Spark + Sqoop программы. Изучение организации распределённых вычислений	6
Рубежный контроль 2			2
Всего:			48

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Во время лекций по дисциплине студентам рекомендуется конспектировать теоретический материал, отмечая важные моменты, на которые заострил внимание преподаватель, участвовать в опросах и дискуссиях. Перед лекцией необходимо повторить выданный материал, зафиксировать непонятные места, чтобы обсудить их на занятии. Конспект лекций представлен в виде мультимедийных презентаций и включен в состав методического комплекса дисциплины.

Лабораторный практикум включает практические задания по основным одиннадцати разделам дисциплины. Все работы выполняются в соответствии с заданием, выданным преподавателем.

Преподавателем запланировано применение на лабораторных занятиях технологий развивающейся кооперации, коллективного взаимодействия, разбора конкретных ситуаций. Поэтому приветствуется групповой метод выполнения лабораторных работ и защиты отчетов, а также взаимооценка и обсуждение результатов выполнения лабораторных работ.

Для текущего контроля успеваемости по очной форме обучения преподавателем используется балльно-рейтинговая система контроля и оценки академической активности.

Выполнение самостоятельной работы подразумевает самостоятельное изучение разделов дисциплины, подготовку к лабораторным занятиям, к рубежным контролям (для обучающихся очной формы обучения), подготовку к экзамену.

Рекомендуемый режим самостоятельной работы

Наименование вида самостоятельной работы	Рекомендуемая трудоемкость, акад. час.
Самостоятельное изучение тем дисциплины:	53
Рекомендательные системы	8
Анализ графов социальных сетей	9
Понижение размерности	9
Машинное обучение на больших данных	9
Архитектура облачных BigData-приложений	9
Способы повышения производительности BigData-систем	9
Подготовка к контрольной работе	-
Подготовка к лабораторным работам (по 2 ч. на каждое занятие)	24
Подготовка к рубежному контролю 1	2
Подготовка к рубежному контролю 2	2
Подготовка к экзамену	27
Всего:	108

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

6.1. Перечень оценочных средств

1. Балльно-рейтинговая система контроля и оценки академической активности обучающихся в КГУ (для очной формы обучения)
2. Отчеты обучающихся по лабораторным занятиям
3. Тестовые задания
4. Банк заданий к рубежным контролям № 1, № 2 (для очной формы обучения).
5. Вопросы к экзамену

6.2. Система балльно-рейтинговой оценки работы обучающихся по дисциплине

Очная форма обучения

№	Наименование	Содержание					
		Распределение баллов, 3 семестр					
1	Распределение баллов за семестры по видам учебной работы, сроки сдачи учебной работы (<i>доводятся до сведения обучающихся на первом учебном занятии</i>)	Вид учебной работы:	Посещение лекций	Выполнение и защита результатов лабораторных работ	Рубежный контроль №1	Рубежный контроль №2	Экзамен
		Балльная оценка:	До 12	11 x 4 б = 44 б	7	7	30
		Примечания:	12 лекций по 1 баллу	11 занятий по 4 балла	На 6-й лаб.р.	На 12-й лаб.р.	
2	Критерий пересчета баллов в традиционную оценку по итогам работы в семестре и экзамена	60 и менее баллов – неудовлетворительно; 61... 73 – удовлетворительно; 74... 90 – хорошо; 91... 100 – отлично.					

3	Критерии допуска к промежуточной аттестации, возможности получения автоматического зачета по дисциплине, возможность получения бонусных баллов	<p>Для допуска к промежуточной аттестации по дисциплине за семестр обучающийся должен набрать по итогам текущего и рубежного контролей не менее 51 балла. В случае, если обучающийся набрал менее 51 балла, то к аттестационным испытаниям он не допускается.</p> <p>Для получения зачета без проведения процедуры промежуточной аттестации обучающемуся необходимо набрать в ходе текущего и рубежного контролей не менее 61 балла. В этом случае итог балльной оценки, получаемой обучающимся, определяется по количеству баллов, набранным им в ходе текущего и рубежного контролей. При этом на усмотрение преподавателя балльная оценка может быть повышена за счет получения дополнительных баллов за академическую активность.</p> <p>Обучающийся, имеющий право на получение оценки без проведения процедуры промежуточной аттестации, может повысить её путем сдачи аттестационного испытания. В случае получения обучающимся на аттестационном испытании 0 баллов итог балльной оценки по дисциплине не снижается.</p> <p>За академическую активность в ходе освоения дисциплины (участие в учебной, научно-исследовательской, спортивной, культурно-творческой и общественной деятельности) обучающемуся могут быть начислены дополнительные баллы. Максимальное количество дополнительных баллов за академическую активность составляет 30.</p> <p>Основанием для получения дополнительных баллов являются:</p> <ul style="list-style-type: none"> выполнение дополнительных заданий по дисциплине, дополнительные баллы начисляются преподавателем; участие в течение семестра в учебной, научно-исследовательской, спортивной, культурно-творческой и общественной деятельности КГУ.
4	Формы и виды учебной работы для неуспевающих (восстановившихся на курсе обучения) обучающихся для получения недостающих баллов в конце семестра	<p>В случае если к промежуточной аттестации (зачету) набрана сумма менее 51 балла, обучающемуся необходимо набрать недостающее количество баллов за счет выполнения дополнительных заданий, до конца последней (зачетной) недели семестра. При этом необходимо проработать материал всех пропущенных практических работ.</p> <p>Формы дополнительных заданий (назначаются зачету):</p> <ul style="list-style-type: none"> - выполнение и защита пропущенной практической работы (при невозможности дополнительного проведения практической работы преподаватель устанавливает форму дополнительного задания по тематике пропущенной практической работы самостоятельно) – до 8 баллов. <p>При прохождении рубежного контроля баллы ставятся в зависимости от рубежа.</p> <p>Ликвидация академических задолженностей, возникших из-за разности в учебных планах при переводе или восстановлении, проводится путем выполнения дополнительных заданий, форма и объем которых определяется преподавателем.</p>

6.3. Процедура оценивания результатов освоения дисциплины

Рубежные контроли проводятся в форме письменного тестирования и решения задач, экзамен - в виде ответа на вопросы.

Перед проведением рубежного контроля преподаватель прорабатывает с обучающимися основной материал соответствующих разделов дисциплины в форме краткой лекции-дискуссии.

Варианты заданий для рубежных контролей № 1, № 2 состоят из 10 вопросов. На каждую подготовку к ответам по рубежному контролю обучающемуся отводится 1 академический час.

Для определения баллов при проверке рубежных контролей используются интервальные оценки, представленные в таблице. Итоговая оценка получается суммированием баллов, набранных при ответе на тест и при решении задачи.

Количество правильных ответов	5-6	7-8	9-10	Решение задачи
Количество баллов	2	3	4	3

Преподаватель оценивает в баллах результаты рубежных контролей каждого обучающегося по количеству правильных ответов и заносит в ведомость учета текущей успеваемости.

На экзамене обучающемуся предлагается ответить на 2 вопроса. Вопросы к экзамену доводятся до обучающихся заранее. Каждый вопрос оценивается в 15 баллов. На подготовку ответа обучающемуся отводится 1 астрономический час.

Результаты текущего контроля успеваемости и экзамена заносятся преподавателем в экзаменационную ведомость, которая сдается в организационный отдел института в день экзамена, а также выставляются в зачетную книжку обучающегося.

6.4. Примеры оценочных средств для рубежных контролей и экзамена

6.4.1 Примеры заданий для рубежного контроля №1

1. Что такое Data Mining?
2. Принцип Бонферрони
3. Оценка TF.IDF для термина i в документе
4. Что такое хэш-функция?
5. Что такое вычислительный кластер?
6. Физическая организация вычислительных узлов
7. Поясните процесс вычислений в MapReduce
8. Редукторы и комбинаторы.
9. Как происходит Обработка отказов узлов?
10. Вычисление выборки с помощью MapReduce
11. Суть метода разбиения на «шинглы».
12. Сходство множеств по Жаккару
13. Суть коллаборативной фильтрации.
14. Как вычислить минхэш множества.
15. Вычисление минхэш-сигнатур
16. Метод разбиения на полосы
17. Редакционное расстояние
18. Расстояние Хэмминга

19. LSH-семейства для расстояния Жаккара
20. Зачем оценивают моменты при оценке потока
21. Алгоритм Датара-Гиониса-Индыка-Мотвани
22. Проблемы обработки потоков
23. Назначение фильтра Блума
24. Зачем нужен тематический PageRank
25. Итеративное вычисление PageRank с помощью MapReduce
26. TrustRank – опишите суть.
27. Решение проблемы спама термов, предложенное Google.
28. Модель корзины покупок
29. Корзины покупок и алгоритм Argioi
30. Алгоритмы обработки больших наборов данных в оперативной памяти
31. Алгоритм иерархической кластеризации
32. Основы алгоритма k средних
33. Обработка данных в алгоритме BFR
34. Кластеризация в неевклидовых пространствах
35. Модель потоковых вычислений
36. Алгоритм кластеризации потока
37. Кластеризация в параллельной среде
38. Онлайн-овые и офлайн-овые алгоритмы
39. Жадный подход к задаче о ключевых словах
40. Обобщенный алгоритм Balance
41. Архитектура машинного обучения
42. Многоклассовые перцептроны
43. Применение Argioi для поиска всех частых предметных наборов
44. Обработка больших наборов данных в оперативной памяти
45. Алгоритмы с ограниченным числом проходов
46. Подсчет частых предметных наборов в потоке

6.4.2 Примеры заданий для рубежного контроля №2

1. Что такое Data Lake?
2. Распределенные файловые системы
3. MapReduce
4. Функция Map
5. Функция Reduce
6. Операции редукции
7. Как происходит управление отказами вычислительных узлов?
8. Коэффициент репликации и размер редукции
9. Что такое схема сопоставления?
10. Умножение матриц с помощью MapReduce.
11. Apache Spark – назначение
12. Apache Kafka – назначение
13. Apache Flume – назначение
14. Apache Ignite – назначение

15. Apache Sqoop – назначение
16. Apache Hive – назначение
17. ZooKeeper– назначение
18. Компонент YARN
19. Apache Hadoop – назначение
20. Что такое оркестрация данных?
21. Что такое HDFS?
22. Два типа узлов в Hadoop
23. Ключевые метрики Big Data system
24. Архитектура лямбда
25. Что такое репликация?
26. Типичный жизненный цикл сообщений в системах очередей

6.4.3 Примеры заданий для контрольной работы

1. Apache Spark – назначение
2. Apache Kafka – назначение
3. Apache Flume – назначение
4. Apache Ignite – назначение
5. Apache Sqoop – назначение
6. Apache Hive – назначение
7. ZooKeeper– назначение
8. Компонент YARN
9. Apache Hadoop – назначение
10. Что такое оркестрация данных?
11. Что такое HDFS?
12. Два типа узлов в Hadoop
13. Ключевые метрики Big Data system
14. Архитектура лямбда
15. Решение проблемы спама термов, предложенное Google.
16. Модель корзины покупок
17. Корзины покупок и алгоритм Apriori
18. Алгоритмы обработки больших наборов данных в оперативной памяти

6.4.4 Примерный перечень вопросов для экзамена

1. Что такое Data Mining?
2. MapReduce и новый программный стек
3. Поиск похожих объектов
4. Анализ потоков данных
5. Анализ ссылок
6. Частые предметные наборы
7. Кластеризация
8. Реклама в Интернете
9. Машинное обучение на больших данных

10. Рекомендательные системы
11. Анализ графов социальных сетей
12. Понижение размерности
13. Современные BigData-решения и архитектуры.
14. Технология обработки и хранения данных Hadoop. Организация хранения данных в BigData-системах.
15. Системы разработки, сборки и доставки кода.
16. Системы пакетной обработки данных.
17. Системы поточной обработки данных.
18. Архитектура облачных BigData-приложений.
19. Способы повышения производительности BigData-систем
20. Архитектура ZooKeeper
21. Архитектура Apache Spark
22. Apache Flume – назначение

6.5. Фонд оценочных средств

Полный банк заданий для текущего, рубежных контролей и промежуточной аттестации по дисциплине, показатели, критерии, шкалы оценивания компетенций, методические материалы, определяющие процедуры оценивания образовательных результатов, приведены в учебно-методическом комплексе дисциплины.

7. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ УЧЕБНАЯ ЛИТЕРАТУРА

7.1. Основная учебная литература

1. Лесковец Ю., Раджараман А., Ульман Д. - Анализ больших наборов данных - 2016
2. Ын А., Су К. - Теоретический минимум по Big Data (Библиотека программиста) – 2019

7.2. Дополнительная учебная литература

- 1 Фрэнкс Б. - Революция в аналитике – 2020
2. Deitel P., Deitel H. - Intro to Python for Computer Science and Data Science - 2020

8. РЕСУРСЫ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫЕ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Электронная библиотека КГУ <http://dspace.kgsu.ru/xmlui/>

9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ И ИНФОРМАЦИОННЫЕ СПРАВОЧНЫЕ СИСТЕМЫ

1. ЭБС «Лань»
2. ЭБС «Консультант студента»
3. ЭБС «Znaniium.com»
4. «Гарант» - справочно-правовая система

10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Материально - техническое обеспечение по реализации дисциплины осуществляется с требованиями ФГОС ВО по данной образовательной программе.

11. ДЛЯ ОБУЧАЮЩИХСЯ С ИСПОЛЬЗОВАНИЕМ ДИСТАНЦИОННЫХ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ

При использовании электронного обучения и дистанционных образовательных технологий (далее ЭО и ДОТ) занятия полностью или частично проводятся в режиме онлайн. Объем дисциплины и распределение нагрузки по видам работ соответствует п. 4.1. Распределение баллов соответствует п. 6.2 либо может быть изменено в соответствии с решением кафедры, в случае перехода на ЭО и ДОТ в процессе обучения. Решение кафедры об используемых технологиях и системе оценивания достижений обучающихся принимается с учетом мнения ведущего преподавателя и доводится до обучающихся.

Аннотация
рабочей программы учебной дисциплины

МЕТОДЫ АНАЛИЗА БОЛЬШИХ ДАННЫХ
образовательной программы высшего образования –
программы магистратуры

09.04.04 Программная инженерия

Направленность:

Методы и алгоритмы интеллектуальной обработки данных в ИВС

Форма обучения: очная

Трудоемкость освоения дисциплины – 5 зач. ед. (144 акад. часа)

Семестры: 3-й (для очной формы обучения)

Промежуточная аттестация: экзамен

Содержание дисциплины

Цель преподавания учебной дисциплины состоит в формировании знаний умений и навыков в области разработки и администрирования программного обеспечения современных высокопроизводительных распределенных систем. В данном курсе рассматриваются программные технологии построения масштабируемых многомашинных информационно-вычислительных систем, обеспечивающих параллельную обработку сверхбольших массивов данных. За рубежом совокупность таких технологий обозначается термином Big Data (англ. - большие данные).

Рассматриваются также типовые методы и алгоритмы параллельной обработки сверхбольших массивов данных.