

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
"Казанский (Приволжский) федеральный университет"  
Институт вычислительной математики и информационных технологий



*подписано электронно-цифровой подписью*

## **Программа дисциплины**

Технологии и инструментальные средства для работы с большими данными

Направление подготовки: 09.03.03 - Прикладная информатика

Профиль подготовки: Большие данные и машинное обучение

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

## Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
  - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
  - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): заведующий кафедрой, к.н. Бурнашев Р.А. (кафедра анализа данных и технологий программирования, отделение фундаментальной информатики и информационных технологий), r.burnashev@inbox.ru ; старший преподаватель, б/с Фарахов Р.Р. (кафедра анализа данных и технологий программирования, отделение фундаментальной информатики и информационных технологий), RRFarakhov@kpfu.ru

## 1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-4	Способен разрабатывать и обеспечивать функционирование информационных баз данных
ПК-5	Способен осуществлять выявление, анализ, спецификацию и разработку требований в процессе разработки программного обеспечения

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

основные понятия и терминологию в области технологий обработки и анализа больших данных; способы оценки снижения качества продукции; основное программное обеспечение работы по анализу больших наборов данных

Должен уметь:

разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных; использовать и применять углубленные знания в области управления качеством; создавать алгоритмы анализа и обработки большого объема данных с применением моделей Data Mining

Должен владеть:

навыками создания и решения моделей, необходимых в сфере управления; программными инструментами для работы с большими наборами данных; методикой прогнозирования

## 2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.ДВ.03.01 Дисциплины (модули)" основной профессиональной образовательной программы 09.03.03 "Прикладная информатика (Большие данные и машинное обучение)" и относится к дисциплинам по выбору части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 3 курсе в 6 семестре.

## 3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 8 зачетных(ые) единиц(ы) на 288 часа(ов).

Контактная работа - 54 часа(ов), в том числе лекции - 18 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 36 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 180 часа(ов).

Контроль (зачёт / экзамен) - 54 часа(ов).

Форма промежуточного контроля дисциплины: экзамен в 6 семестре.

## 4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

### 4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины / модуля	Се-местр	Виды и часы контактной работы, их трудоемкость (в часах)						Само-стоя-тель-ная ра-бота
			Лекции, всего	Лекции в эл. форме	Практи-ческие занятия, всего	Практи-ческие в эл. форме	Лабора-торные работы, всего	Лабора-торные в эл. форме	
1	Тема 1. Тема 1. Современные подходы к								

обработке и хранению

N	Разделы дисциплины / модуля	Семестр	Виды и часы контактной работы, их трудоемкость (в часах)						Самостоятельная работа
			Лекции, всего	Лекции в эл. форме	Практические занятия, всего	Практические в эл. форме	Лабораторные работы, всего	Лабораторные в эл. форме	
2.	Тема 2. Тема 2. Извлечение и визуализация данных	6	3	0	0	0	6	0	30
3.	Тема 3. Тема 3. Подготовка данных к анализу. Методика извлечения знаний	6	3	0	0	0	6	0	30
4.	Тема 4. Тема 4. Алгоритмы параллельной обработки	6	3	0	0	0	6	0	30
5.	Тема 5. Тема 5. Аналитические платформы: классификация и особенности применения	6	3	0	0	0	6	0	30
6.	Тема 6. Тема 6. Понятие сценария и узла обработки Лабораторный практикум	6	3	0	0	0	6	0	30
	Итого		18	0	0	0	36	0	180

#### 4.2 Содержание дисциплины (модуля)

##### Тема 1. Тема 1. Современные подходы к обработке и хранению

В данной теме студент осваивает следующие разделы: Инструментальные средства по машинному обучению. Проверка адекватности моделей бинарного выбора. Селекция моделей на основе информационных критериев. Модели Множественного выбора с неупорядоченными альтернативами. Модель наивного байесовского классификатора. Алгоритмы дерева принятия решения для решения задач классификатора

##### Тема 2. Тема 2. Извлечение и визуализация данных

Извлечение и визуализация данных находит широкое применение в научных и статистических исследованиях. В данной теме студент осваивает следующие разделы: установление соединения с кластером. Извлечение данных с удаленного файла. Hadoop-кластер. Получение информации. Анализ представленной информации. Визуализация данных.

##### Тема 3. Тема 3. Подготовка данных к анализу. Методика извлечения знаний

В данной теме студент осваивает следующие разделы: Модель данных и информационная архитектура предприятия. Автоматизация подготовки данных. Инкрементальная и автоматическая выгрузка. Хранилище данных с несколькими источниками. Документно-ориентированная СУБД MongoDB. Использование неструктурированных данных в качестве источника.

##### Тема 4. Тема 4. Алгоритмы параллельной обработки

В данной теме студент осваивает следующие разделы: Алгоритмы случайного леса для решения задач классификации. Ансамбль алгоритмов randomforest. Классификации данных методом опорных векторов. Алгоритм линейный классификации. Метод релевантных векторов. Метод градиентного бустинга. Ресемплинг как метод бустрап-анализа.

##### Тема 5. Тема 5. Аналитические платформы: классификация и особенности применения

Инструмент разработки, планирования и мониторинга рабочих процессов. Озеро данных (Data Lake). Hadoop для масштабируемой, распределенной обработки больших наборов данных. Модель распределенных вычислений MapReduce. Поточковая обработка данных. Kafka. Облачные технологии для масштабирования систем.

##### Тема 6. Тема 6. Понятие сценария и узла обработки Лабораторный практикум

В данной теме студент осваивает следующие разделы: Импорт подготовленного текстового файла в созданный проект. Настройка параметров импорта данных из файла. Метки узла. Поддержка HDFS работы на узлах. Репликация и осведомленность об узлах.

Узлы имени в HDFS (Name Nodes). Выделение памяти в Узлах имени (NameNodes). Проблема с файлами малого размера NameNode

#### 5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-99бин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

## **6. Фонд оценочных средств по дисциплине (модулю)**

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

## **7. Перечень литературы, необходимой для освоения дисциплины (модуля)**

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;
- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

## **8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)**

Научная электронная библиотека eLIBRARY.RU - <https://elibrary.ru/>

Электронная библиотечная система ZNANIUM.COM - <http://znanium.com/>

Электронно-библиотечная система "Издательство Лань", - <https://e.lanbook.com/>

## 9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лекции	Работа обучающегося включает усвоение теоретического материала, подготовку к лекционным занятиям, работу с учебниками, иной учебной и учебно-методической литературой. Для качественного усвоения материала обучающийся должен изучить материалы лекционных занятий. В теоретическом курсе темы сопровождаются перечнем рекомендуемой литературы, которую студент может взять в библиотеке или в сети Интернет. В ходе предварительного изучения материалов обучающемуся рекомендуется просмотреть конспект сразу после занятий. Пометить материал конспекта лекций, который вызывает затруднения для понимания. Попытаться найти ответы на затруднительные вопросы, используя предлагаемую литературу. Если самостоятельно не удалось разобраться в материале, следует сформулировать вопросы и обратиться на текущей консультации или на ближайшей лекции за помощью к преподавателю. Каждую неделю рекомендуется отводить время для повторения пройденного материала, проверяя свои знания, умения и навыки по контрольным вопросам. Успешное освоение и изучение теоретического материала дисциплины возможно только при его применении в учебной и научной деятельности.
лабораторные работы	Лабораторная работа включает себя закрепление практических навыков студента под руководством преподавателя. Студент по заранее намеченному плану используя научное оборудование и специализированное программное обеспечение, выполняют определенные практические задания и в процессе их выполнения воспринимают и осмысливают новый учебный материал.
самостоятельная работа	Процесс организации самостоятельной работы студентов включает в себя следующие этапы: - подготовительный (определение целей, составление программы, подготовка методического обеспечения, подготовка оборудования); - основной (реализация программы, использование приемов поиска информации, усвоения, переработки, применения, передачи знаний, фиксирование результатов, самоорганизация процесса работы); - заключительный (оценка значимости и анализ результатов, их систематизация, оценка эффективности программы и приемов работы, выводы о направлениях оптимизации труда).
экзамен	Повторение материала: Рекомендуйте регулярно пересматривать конспекты и лекции, чтобы освежить знания перед зачётом. Создание схем и таблиц: Визуальные материалы помогут структурировать информацию и запомнить ключевые моменты. Организация групповых обсуждений может помочь в обмене знаниями и прояснении сложных тем.

## 10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

## 11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

## 12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:



- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 09.03.03 "Прикладная информатика" и профилю подготовки "Большие данные и машинное обучение".



*Приложение 2*  
*к рабочей программе дисциплины (модуля)*  
*Б1.В.ДВ.03.01 Технологии и инструментальные средства для*  
*работы с большими данными*

**Перечень литературы, необходимой для освоения дисциплины (модуля)**

Направление подготовки: 09.03.03 - Прикладная информатика

Профиль подготовки: Большие данные и машинное обучение

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

**Основная литература:**

Базы данных : методические указания / составители А. Р. Мангушева. - Казань : КНИТУ, 2017. - 16 с. - Текст : электронный // Лань : электроннобиблиотечная система. - URL: <https://e.lanbook.com/book/138352>

Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. - 3-е изд., стер. - Санкт-Петербург : Лань, 2023. - 188 с.- ISBN 978-5-507-46866-9. - Текст : электронный // Лань : электроннобиблиотечная система. - URL: <https://e.lanbook.com/book/322664>

Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. - Москва : РТУ МИРЭА, 2021. - 91 с. - Текст : электронный // Лань : лектронно-библиотечная система. - URL: <https://e.lanbook.com/book/182452>

Берикашвили, В. Ш. Статистическая обработка данных, планирование эксперимента и случайные процессы : учебное пособие для бакалавриата и магистратуры / В. Ш. Берикашвили, С. П. Оськин. - 2-е изд., испр. и доп. - Москва : Издательство Юрайт, 2019. - 164 с. - (Бакалавр и магистр. Академический курс). - ISBN 978-5-534-09216-5. - Текст : электронный // Образовательная платформа Юрайт [сайт]. - URL: <https://urait.ru/bcode/427449>

**Дополнительная литература:**

Бутаков, Н. А. Обработка больших данных с Apache Spark : учебнометодическое пособие : [16+] / Н. А. Бутаков, М. В. Петров, Д. Насонов. - Санкт-Петербург : Университет ИТМО, 2019. - 52 с. : ил. - Режим доступа: по подписке. - URL: <https://biblioclub.ru/index.php?page=book&id=566771>

Информационное обеспечение и базы данных : учебное пособие / составитель А. Ф. Похилько. - Ульяновск : УлГТУ, 2019. - 127 с. - ISBN 978-5-9795-1964-7. - Текст : электронный // Лань : электронно-библиотечная система.

*Приложение 3*  
*к рабочей программе дисциплины (модуля)*  
*Б1.В.ДВ.03.01 Технологии и инструментальные средства для*  
*работы с большими данными*

**Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем**

Направление подготовки: 09.03.03 - Прикладная информатика

Профиль подготовки: Большие данные и машинное обучение

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен обучающимся. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.