

Министерство науки и высшего образования РФ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

СОГЛАСОВАНО

Заведующий кафедрой

**Кафедра прикладной
математики и компьютерной
безопасности (ПМКБ_ИКИТ)**

наименование кафедры

подпись, инициалы, фамилия

«___» _____ 20__ г.

институт, реализующий ОП ВО

УТВЕРЖДАЮ

Заведующий кафедрой

**Кафедра прикладной математики
и компьютерной безопасности
(ПМКБ_ИКИТ)**

наименование кафедры

А.А. Кытманов

подпись, инициалы, фамилия

«___» _____ 20__ г.

институт, реализующий дисциплину

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
АНАЛИЗ БОЛЬШИХ ДАННЫХ (BIG
DATA)**

Дисциплина Б1.В.05 Анализ больших данных (Big Data)

Направление подготовки /
специальность 01.04.02 Прикладная математика и
информатика,
программа 01 04 02 09 Data Science and

Направленность
(профиль)

Форма обучения

очная

Год набора

2021

Красноярск 2021

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования с учетом профессиональных стандартов по укрупненной группе

010000 «МАТЕМАТИКА И МЕХАНИКА»

Направление подготовки /специальность (профиль/специализация)

Направление 01.04.02 Прикладная математика и информатика,
программа 01.04.02.09 Data Science and Mathematical Modeling

Программу
составили

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Ознакомление студентов с существующими методами и систематикой хранения, обработки и анализа больших данных

1.2 Задачи изучения дисциплины

Изучить принципы работы с большими данными, получить представление о компонентах экосистемы Hadoop и облачных платформах для реализации решения по BigData;
овладеть стандартными инструментами работы с большими данными

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

УК-3:Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели.
УК-3.1:Вырабатывает стратегию сотрудничества и на ее основе организует отбор членов команды для достижения поставленной цели.
УК-3.2:Планирует и корректирует работу команды с учетом интересов, особенностей поведения и мнений ее членов.
УК-3.3:Разрешает конфликты и противоречия при деловом общении на основе учета интересов всех сторон.
УК-3.4:Организует дискуссии по заданной теме и обсуждение результатов работы команды с привлечением оппонентов разработанным идеям.
УК-3.5:Планирует командную работу, распределяет поручения и делегирует полномочия членам команды.
ПК-2:Способен управлять этапами жизненного цикла методологической и технологической инфраструктуры анализа данных в организации.
ПК-2.1:Знает современные методы и инструментальные средства анализа данных; методы интерпретации и визуализации данных; источники информации, в том числе информации, необходимой для обеспечения деятельности в предметной области; теоретические основы информатики и исследования операций; методы извлечения информации и знаний из гетерогенных, мульти структурированных, неструктурированных источников, в том числе при потоковой обработке.
ПК-2.2:Умеет применять современные методы и инструментальные средства анализа данных; методы интерпретации и визуализации данных; умеет пользоваться источниками информации, в том числе источниками информации, необходимыми для обеспечения деятельности в предметной области.
ПК-2.3:Способен анализировать и использовать современные методы и инструментальные средства анализа данных для решения практических и научных задач; способен применять методы интерпретации и визуализации

данных для решения научных и практических задач.
ПК-3:Способен управлять разработкой продуктов, услуг и решений на основе данных.
ПК-3.1:Знает: состояние и перспективы развития информационных технологий, технологий данных в России и в мире; современные и перспективные методы сбора, хранения и передачи данных; источники данных, интенсивность генерации данных источниками; технические средства и среды сбора, хранения и обработки данных; современные и перспективные средства визуализации и интерпретации данных; исследование операций; машинное обучение; математическое моделирование; методы сравнительного анализа.
ПК-3.2:Способен проводить аналитические и поисковые исследования по тематике информационных технологий, технологий данных.
ПК-4:Способен разрабатывать и внедрять новые методы и технологии исследования данных.
ПК-4.1:Знает: состояние и перспективы развития информационных технологий, технологий данных в России и в мире; современные и перспективные методы сбора, хранения и передачи данных; источники данных, интенсивность генерации данных источниками; технические средства и среды сбора, хранения и обработки данных; современные и перспективные средства визуализации и интерпретации данных; исследование операций; машинное обучение; математическое моделирование; методы сравнительного анализа.
ПК-4.2:Способен проводить аналитические и поисковые исследования по тематике информационных технологий, технологий данных.

1.4 Место дисциплины (модуля) в структуре образовательной программы

Продвинутые методы анализа данных

Базовые модели и методы машинного обучения

Системы хранения, обработки и управления данными

Продвинутые методы анализа данных (Advanced Methods of Data Analysis)

Базовые модели и методы машинного обучения (The Basics of Machine Learning)

Системы хранения, обработки и управления данными (Data Storage and Management Systems)

выполнение и защита выпускной квалификационной работы

Технологическая (проектно-технологическая) практика

Выполнение и защита выпускной квалификационной работы (Final certification)

Технологическая (проектно-технологическая) практика (Field Internship II)

1.5 Особенности реализации дисциплины

Язык реализации дисциплины Английский.

Дисциплина (модуль) реализуется без применения ЭО и ДОТ.

2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	Семестр
		3
Общая трудоемкость дисциплины	3 (108)	3 (108)
Контактная работа с преподавателем:	1 (36)	1 (36)
занятия лекционного типа	0,5 (18)	0,5 (18)
занятия семинарского типа		
в том числе: семинары		
практические занятия	0,5 (18)	0,5 (18)
практикумы		
лабораторные работы		
другие виды контактной работы		
в том числе: групповые консультации		
индивидуальные консультации		
иная внеаудиторная контактная работа:		
групповые занятия		
индивидуальные занятия		
Самостоятельная работа обучающихся:	2 (72)	2 (72)
изучение теоретического курса (ТО)		
расчетно-графические задания, задачи (РГЗ)		
реферат, эссе (Р)		
курсовое проектирование (КП)	Нет	Нет
курсовая работа (КР)	Нет	Нет
Промежуточная аттестация (Зачёт)		

3 Содержание дисциплины (модуля)

3.1 Разделы дисциплины и виды занятий (тематический план занятий)

№ п/п	Модули, темы (разделы) дисциплины	Занятия лекционного типа (акад. час)	Занятия семинарского типа		Самостоятельная работа, (акад. час)	Формируемые компетенции
			Семинары и/или Практические занятия (акад. час)	Лабораторные работы и/или Практикумы (акад. час)		
1	2	3	4	5	6	7
1	Batch processing	6	6	0	24	
2	Stream Data Processing	6	6	0	24	
3	Big data storage	6	6	0	24	
Всего		18	18	0	72	

3.2 Занятия лекционного типа

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
1	1	Distributed file systems	2	0	0
2	1	MapReduce Computation Model	2	0	0
3	1	Beyond MapReduce, Spark	2	0	0
4	2	Schemes for receiving and processing sweat data	2	0	0
5	2	Stream processing models and their characteristics	2	0	0
6	2	Streaming frameworks (open source Apache products)	2	0	0
7	3	Types of data warehouses (BigTable-like, Dynamo-like, SQL over BigData)	2	0	0
8	3	Architecture of data processing systems	4	0	0

Всего		18	0	0
-------	--	----	---	---

3.3 Занятия семинарского типа

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
1	1	Distributed file systems	2	0	0
2	1	MapReduce Computation Model	2	0	0
3	1	Beyond MapReduce, Spark	2	0	0
4	2	Schemes for receiving and processing sweat data	2	0	0
5	2	Stream processing models and their characteristics	2	0	0
6	2	Streaming frameworks (open source Apache products)	2	0	0
7	3	Types of data warehouses (BigTable-like, Dynamo-like, SQL over BigData)	2	0	0
8	3	Architecture of data processing systems	4	0	0
Всего			18	0	0

3.4 Лабораторные занятия

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
Всего					

5 Фонд оценочных средств для проведения промежуточной аттестации

Оценочные средства находятся в приложении к рабочим программам дисциплин.

6 Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

6.1. Основная литература

Авторы, составители	Заглавие	Издательство, год

Л1.1	Efron B., Hastie T.	Computer Age Statistical Inference: Algorithms, Evidence, and Data Science: monographs	Cambridge: Cambridge University Press, 2016
6.2. Дополнительная литература			
	Авторы, составители	Заглавие	Издательство, год
Л2.1	Rajasekaran S., Reif J.	Handbook of Parallel Computing. Models, Algorithms and Applications	London: Chapman & Hall/CRC, 2008

8 Методические указания для обучающихся по освоению дисциплины (модуля)

Для успешного усвоения учебного материала и получения зачета необходимо:

- прослушать лекции или, в случае пропуска, изучить их материал самостоятельно с помощью рекомендуемой литературы и ресурсов сети Интернет;
- написать реферат по одной из предложенных тем и выступить с докладом по теме реферата;
- выполнить проект "Облачные вычисления для анализа больших данных" и защитить на одном из практических занятий;
- успешно пройти собеседование по вопросам к зачету

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации в зависимости от нозологии:

Для лиц с нарушениями зрения:

- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа

9 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю) (при необходимости)

9.1 Перечень необходимого программного обеспечения

9.1.1	Языки программирования PHP и Python
-------	-------------------------------------

9.2 Перечень необходимых информационных справочных систем

9.2.1	• электронные информационно-справочные ресурсы научной библиотеки СФУ (http://bik.sfu-kras.ru)
-------	---

10 Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)

Учебные лаборатории и классы, оснащенные современными компьютерами, объединенными в локальные вычислительные сети с выходом в Интернет, а также периферийным и проекционным оборудованием.