Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Сибирский государственный индустриальный университет»

Кафедра автоматизации и информационных систем

УТВЕРЖДАЮ
Директор института
информационных технологий и
автоматизированных систем
_____ Л.Д. Павлова
подпись
«____» _____ 20___ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Инструменты для работы с большими данными

09.03.02 «Информационные системы и технологии»
(направленность (профиль): «Информационные системы и технологии»)

Квалификация выпускника
Бакалавр

Форма обучения
Очная форма

Срок обучения: 4 года

Год начала подготовки 2024

Новокузнецк
2024

СибГИУ
Подлинник электронного документа хранится в системе электронного согласования документов СибГИУ
ДОКУМЕНТ ПОДПИСАН ЭП
Сертификат: f71ddb1db49f86d988698c2b5d8babab266584d5
Владелец: Директор ИИТиАС Павлова Лариса Дмитриевна
Действителен: с 20 марта 2024 г. по 13 июня 2025 г.
Дата и время: 14-05-2024 01:24:01 ПРОВЕРИТЬ ПОДПИСЬ

1 Цели и задачи освоения учебной дисциплины

Целями учебной дисциплины являются:

– формирование у обучающихся профессиональной компетенции в области разработки и использования систем обработки и анализа больших массивов данных.

Задачами учебной дисциплины являются:

– освоение обучающимся методов анализа и хранения больших объемов данных, этапов жизненного цикла обработки больших данных, языков, наиболее приспособленные для обработки и аналитики больших данных, способов организации хранения и доступа к большим данным;

– освоение обучающимися технологий разработки специализированных программных систем, отвечающих за обработку больших данных.

2 Место учебной дисциплины в структуре ООП по направлению подготовки (специальности)

Учебная дисциплина относится к учебным дисциплинам обязательной части **Блока 1 «Дисциплины (модули)»** ООП по направлению подготовки (специальности) 09.03.02 «Информационные системы и технологии».

Учебная дисциплина базируется на предварительном усвоении обучающимися учебных дисциплин:

– Техническое обеспечение информационно-управляющих систем;

– Основы облачных платформ и распределенных систем;

– Технологии сбора и обмена информацией в системах больших данных;

– Современные СУБД и хранилища данных.

Учебная дисциплина дополняет знания и умения, получаемые по одновременно изучаемым и последующим учебным дисциплинам:

– Проектирование информационно-управляющих систем;

– Программное обеспечение информационно-управляющих систем.

3 Планируемые результаты обучения по учебной дисциплине

Процесс изучения учебной дисциплины направлен на формирование следующих компетенций:

– Общепрофессиональные компетенции

Наименование категории (группы) ОПК	Код и наименование ОПК	Код и наименование индикатора	Планируемые результаты обучения

		достижения ОПК	
	ОПК-1: Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ОПК-1.3 Применяет методы математического анализа и моделирования в теоретических и экспериментальных исследованиях профессиональной деятельности	– знать: методы математического анализа больших данных. – уметь: использовать регрессионные модели при работе с большими данными.
	ОПК-2: Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	ОПК-2.3 Применяет современные информационные технологии и программные средства при решении прикладных задач	– знать: прикладной инструментарий работы с большими данными. – уметь: использовать информационные технологии решения прикладных задач больших данных.
	ОПК-3: Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	ОПК-3.3 Решает поставленные задачи на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий	– знать: методы и способы поиска информации. – уметь: применять информационно-коммуникационные технологии для поиска информации.

4 Объем и содержание учебной дисциплины

Учебные занятия по учебной дисциплине проводятся в форме контактной работы и в форме самостоятельной работы обучающихся.

Контактная работа включает в себя занятия лекционного типа (лекции), занятия семинарского типа (семинары, практические занятия, практикумы), промежуточную аттестацию обучающихся и иные формы взаимодействия обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации ООП на иных условиях, в том числе при проведении промежуточной аттестации обучающихся. Контактная работа может проводиться с применением электронного обучения, дистанционных образовательных технологий.

Объем учебной дисциплины

Семестр / курс		***ИТОГО***	***7 семестр***
Форма промежуточной аттестации			*экзамен*
Трудоёмкость	*академ. час.*	***144***	*144*
	зачетных единиц	***4***	*4*
Лекции, *академ. час.*		***32***	*32*
в форме практической подготовки		***0***	*0*
Лабораторные работы, *академ. час.*		***0***	*0*
в форме практической подготовки		***0***	*0*
Практические занятия, *академ. час.*		***16***	*16*
в форме практической подготовки		***0***	*0*
Курсовая работа / проект, *академ. час.*		***0***	*0*
в форме практической подготовки		***0***	*0*
Консультации, *академ. час.*		***0***	*0*
в форме практической подготовки		***0***	*0*
Самостоятельная работа, *академ. час.*		***51***	*51*
в форме практической подготовки		***0***	*0*
Контроль, *академ. час.*		***45***	*45*
в форме практической подготовки		***0***	*0*

Содержание учебной дисциплины

Раздел 1 Введение в большие данные (Понятие Data Minig. Прикладные инструменты для работы с Big Data. Технология MapRaduce. Hadoop.);

Тема 1.1 Понятие Data Mining (Основные понятия Data Mining. Понятия аналитической обработки данных, витрины данных. Data Warehouse. Data Lakes. Современные методы построения хранилищ данных.);

Тема 1.2 Прикладные инструменты для работы с Big Data (Инструменты для работы с большими данными. Технологии параллельной (parallel) обработки информации. Технология mpp. Общие сведения о пакетах Apache Hadoop, Apache Spark, Kafka, RabbitMQ.);

Тема 1.3 Технология MapReduce (Технология работы с большими массивами данных. Основные этапы технологии MapReduce. Использование MapReduce в современных пакетах работы с большими данными.);

Тема 1.4 Apache Hadoop (Концепции Apache Hadoop. Типы узлов Hadoop. Концепция арбитра в распределенных многопоточных системах. Файловая система HDFS. Принцип работы Hadoop.);

Тема 1.5 Apache Spark (Концепции Apache Spark. Типы узлов Spark. Механизмы передачи и обработки данных Apache Spark. Файловая система RDD. Разъединенные универсальные коллекции Apache Spark.);

Раздел 2 Технологии анализа данных (Жизненный цикл анализа больших данных, стандарты. Когнитивный анализ данных. Визуализация больших данных.);

Тема 2.1 Жизненный цикл анализа больших данных (Основные этапы подготовки и хранения больших данных для анализа. Сбор больших данных. Подготовка к анализу больших данных. Анализ данных.);

Тема 2.2 Статистический анализ больших данных (Методы анализа данных. Цель анализа данных. Синтаксический анализ данных. Когнитивный анализ данных. Способы и технологии когнитивного анализа данных);

Тема 2.3 Визуализация больших данных (Задача визуализации данных. Технологии визуализации данных. Инструментарий визуализации больших данных).

5 Перечень тем лекций

№ раздела / темы дисциплины	**Темы лекций**	**Трудоемкость, *академ.час***	
		всего	**в форме практической подготовки**
Раздел 1.	Введение в большие данные		
Тема 1.1.	Понятие Data Mining	2	
Тема 1.2.	Прикладные инструменты для работы с Big Data	4	
Тема 1.3.	Технология MapReduce	4	
Тема 1.4.	Apache Hadoop	4	
Тема 1.5.	Apache Spark	6	
Раздел 2.	Технологии анализа данных		
Тема 2.1.	Жизненный цикл анализа больших данных	4	
Тема 2.2.	Статистический анализ больших данных	6	
Тема 2.3.	Визуализация больших данных	2	
Итого:		**32**	**0**

6 Перечень тем практических занятий (семинаров)

№ раздела / темы дисциплины	Темы практических занятий (семинаров)	Трудоемкость, *академ.час*	
		всего	**в форме практической подготовки**
Раздел 1.	Введение в большие данные		
Тема 1.1.	Понятие Data Mining	2	
Тема 1.2.	Прикладные инструменты для работы с Big Data	2	
Тема 1.3.	Технология MapReduce	2	
Тема 1.4.	Apache Hadoop	2	
Тема 1.5.	Apache Spark	2	
Раздел 2.	Технологии анализа данных		
Тема 2.1.	Жизненный цикл анализа больших данных, стандарты	2	
Тема 2.2.	Когнитивный анализ данных	2	
Тема 2.3.	Визуализация больших данных	2	
Итого:		**16**	**0**

7 Перечень тем лабораторных работ

№ раздела / темы дисциплины	Темы лабораторных работ	Трудоемкость, *академ.час*	
		всего	**в форме практической подготовки**
	Отсутствуют		
Итого:		**0**	**0**

8 Перечень тем курсовых работ (проектов)

№ раздела / темы дисциплины	Темы курсовых работ (проектов)	Трудоемкость, *академ.час*	
		всего	**в форме практической подготовки**
	Отсутствуют		
Итого:		**0**	**0**

9 Виды самостоятельной работы

№ раздела / темы дисциплины	Виды самостоятельной работы	Трудоемкость, *академ.час*	
		всего	**в форме практической подготовки**
Раздел 1.	1. Выполнение домашнего задания; 2. Оформление отчета о практической работе;	30	

	3. Подготовка к практическому занятию; 4. Прохождение тестирования.		
Раздел 2.	1. Выполнение домашнего задания; 2. Оформление отчета о практической работе; 3. Подготовка к практическому занятию; 4. Прохождение тестирования.	21	
Контроль	*Подготовка к экзамену*	45	
Итого:		**96**	**0**

10 Учебно-методическое и информационное обеспечение учебной дисциплины

а) литература:

1 Базы данных в высокопроизводительных информационных системах : учебное пособие / авт.-сост. Е. И. Николаев ; Северо-Кавказский федеральный университет. – Ставрополь : СКФУ, 2016. – 163 с. – URL: https://biblioclub.ru/index.php?page=book&id=466799 (дата обращения: 13.04.2024);

2 Жуковский, О. И. Информационные технологии и анализ данных : учебное пособие / О. И. Жуковский ; Томский Государственный университет систем управления и радиоэлектроники. – Томск : Эль Контент, 2014. – 130 с. – ISBN 978-5-4332-0158-3. – URL: https://biblioclub.ru/index.php?page=book&id=480500 (дата обращения: 13.04.2024);

3 Бутаков, Н. А. Обработка больших данных с Apache Spark : учебно-методическое пособие / Н. А. Бутаков, М. В. Петров, Д. Насонов. – Санкт-Петербург : Университет ИТМО, 2019. – 52 с. – URL: https://biblioclub.ru/index.php?page=book&id=566771 (дата обращения: 13.04.2024);

4 Советов, Б. Я. Базы данных : учебник для вузов / Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской. – 3-е изд., перераб. и доп. – Москва : Издательство Юрайт, 2023. – 420 с. – ISBN 978-5-534-07217-4. – URL: https://urait.ru/bcode/510752 (дата обращения: 13.04.2024);

5 Нестеров, С. А. Базы данных : учебник и практикум для вузов / С. А. Нестеров. — Москва : Издательство Юрайт, 2023. – 230 с. – ISBN 978-5-534-00874-6. – URL: https://urait.ru/bcode/511650 (дата обращения: 13.04.2024);

6 Лэм, Ч. Hadoop в действии / Чак Лэм. – Москва : ДМК-пресс, 2012. – 424 с. – ISBN 978-5-94074-785-7. – URL:

https://www.studentlibrary.ru/book/ISBN9785940747857.html (дата обращения: 13.04.2024).

б) ресурсы информационно-телекоммуникационной сети «Интернет»:

1 Консультант студента : электронно-библиотечная система / ООО «КОНСУЛЬТАНТ СТУДЕНТА». – Москва, [200 –]. – URL: http://www.studentlibrary.ru. – Режим доступа: для авторизир. пользователей;

2 ЛАНЬ : электронно-библиотечная система : [коллекция «Инженерно-технические науки»] / ООО «Издательство ЛАНЬ». – Санкт-Петербург, [200 –]. – URL: http://e.lanbook.com. – Режим доступа: для авторизир. пользователей;

3 НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА eLIBRARY.RU : база данных / ООО «НЭБ». – Москва, [200 –]. – URL: http://elibrary.ru. – Режим доступа: по подписке;

4 Образовательная платформа ЮРАЙТ / ООО «Электронное издательство ЮРАЙТ». – Москва, [200 –]. – URL: https://urait.ru. – Режим доступа: для авторизир. пользователей;

5 Университетская библиотека онлайн : электронно-библиотечная система / ООО «Директ-Медиа». – Москва, [200 –]. – URL: https://biblioclub.ru. – Режим доступа: для авторизир. пользователей. – URL: http://www.biblioclub.ru;

6 Электронная библиотека // Научно-техническая библиотека СибГИУ : сайт. – Новокузнецк, [200 –]. – URL: http://library.sibsiu.ru/LibrELibraryFullText.asp. – Режим доступа: для авторизир. пользователей. – URL: https://library.sibsiu.ru/LibrELibraryFullText.asp;

7 Электронные периодические издания ИВИС : универсальная база данных / ООО «ИВИС». – Москва, [200 –]. – URL: http://eivis.ru. – Режим доступа: по подписке;

8 Электронный каталог : сайт / Научно-техническая библиотека СибГИУ. – Новокузнецк, [199 –]. – URL: http://libr.sibsiu.ru. – URL: https://libr.sibsiu.ru.

в) лицензионное и свободно распространяемое программное обеспечение:

– 7-Zip;
– Adobe Acrobat Reader;
– Astra Linux Special Edition;
– Kaspersky Endpoint Security;
– Microsoft Office;
– Microsoft Windows;
– OnlyOffice;
– Р7-Офис.

г) базы данных и информационно-справочные системы:

1 ГАРАНТ : справочно-правовая система / ООО «Правовой центр «Гарант». – Кемерово, [200 –]. – Режим доступа: компьютерная сеть Сиб. гос. индустр. ун-та.;

2 КонсультантПлюс : справочно-правовая система / ООО «Информационный центр АНВИК». – Новокузнецк, [199 –]. – Режим доступа: компьютерная сеть библиотеки Сиб. гос. индустр. ун-та.;

3 Техэксперт : информационно-справочная система / ООО «Группа компаний «Кодекс». – Кемерово, [200 –]. – Режим доступа: компьютерная сеть Сиб. гос. индустр. ун-та.

11 Материально-техническое обеспечение учебной дисциплины

Материально-техническое обеспечение учебной дисциплины включает учебные аудитории, оснащенные оборудованием, компьютерной техникой, и техническими средствами обучения, в том числе:

- учебную аудиторию для проведения занятий лекционного типа, оборудованную учебной доской, экраном и мультимедийным проектором;
- учебную аудиторию для проведения занятий семинарского типа (практических занятий), оснащенную компьютерным и мультимедийным оборудованием;
- учебную аудиторию (помещения) для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации;
- помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду, научно-техническую библиотеку СибГИУ.

Рабочая программа дисциплины составлена в соответствии с требованиями ФГОС ВО по направлению подготовки (специальности) 09.03.02 «Информационные системы и технологии».

Составитель(и):

доцент Добрынин Алексей Сергеевич (кафедра автоматизации и информационных систем).

Рабочая программа дисциплины рассмотрена и утверждена на заседании кафедры.

Приложение

Аннотация
рабочей программы дисциплины «Инструменты для работы с большими данными»

по направлению подготовки (специальности)
09.03.02 «Информационные системы и технологии»
(направленность (профиль): «Информационные системы и технологии»)
форма обучения – Очная форма

1 Цели и задачи освоения учебной дисциплины

Целями учебной дисциплины являются:

– формирование у обучающихся профессиональной компетенции в области разработки и использования систем обработки и анализа больших массивов данных.

Задачами учебной дисциплины являются:

– освоение обучающимся методов анализа и хранения больших объемов данных, этапов жизненного цикла обработки больших данных, языков, наиболее приспособленные для обработки и аналитики больших данных, способов организации хранения и доступа к большим данным;

– освоение обучающимися технологий разработки специализированных программных систем, отвечающих за обработку больших данных.

2 Место учебной дисциплины в структуре ООП по направлению подготовки (специальности)

Учебная дисциплина относится к учебным дисциплинам обязательной части **Блока 1 «Дисциплины (модули)»** ООП по направлению подготовки (специальности) 09.03.02 «Информационные системы и технологии».

Учебная дисциплина базируется на предварительном усвоении обучающимися учебных дисциплин:

– Техническое обеспечение информационно-управляющих систем;
– Основы облачных платформ и распределенных систем;
– Технологии сбора и обмена информацией в системах больших данных;
– Современные СУБД и хранилища данных.

Учебная дисциплина дополняет знания и умения, получаемые по одновременно изучаемым и последующим учебным дисциплинам:

– Проектирование информационно-управляющих систем;
– Программное обеспечение информационно-управляющих систем.

3 Планируемые результаты обучения по учебной дисциплине

Процесс изучения учебной дисциплины направлен на формирование следующих компетенций:

– Общепрофессиональные компетенции

Наименование категории (группы) ОПК	Код и наименование ОПК	Код и наименование индикатора достижения ОПК	Планируемые результаты обучения
	ОПК-1: Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ОПК-1.3 Применяет методы математического анализа и моделирования в теоретических и экспериментальных исследованиях профессиональной деятельности	– знать: методы математического анализа больших данных. – уметь: использовать регрессионные модели при работе с большими данными.
	ОПК-2: Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	ОПК-2.3 Применяет современные информационные технологии и программные средства при решении прикладных задач	– знать: прикладной инструментарий работы с большими данными. – уметь: использовать информационные технологии решения прикладных задач больших данных.
	ОПК-3: Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с	ОПК-3.3 Решает поставленные задачи на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий	– знать: методы и способы поиска информации. – уметь: применять информационно-коммуникационные технологии для поиска информации.

	учетом основных требований информационной безопасности		

4 Объем учебной дисциплины

Семестр / курс		*ИТОГО*	*7 семестр*
Форма промежуточной аттестации			*экзамен*
Трудоёмкость	*академ. час.*	***144***	*144*
	зачетных единиц	***4***	*4*
Лекции, *академ. час.*		***32***	*32*
в форме практической подготовки		***0***	*0*
Лабораторные работы, *академ. час.*		***0***	*0*
в форме практической подготовки		***0***	*0*
Практические занятия, *академ. час.*		***16***	*16*
в форме практической подготовки		***0***	*0*
Курсовая работа / проект, *академ. час.*		***0***	*0*
в форме практической подготовки		***0***	*0*
Консультации, *академ. час.*		***0***	*0*
в форме практической подготовки		***0***	*0*
Самостоятельная работа, *академ. час.*		***51***	*51*
в форме практической подготовки		***0***	*0*
Контроль, *академ. час.*		***45***	*45*
в форме практической подготовки		***0***	*0*

5 Краткое содержание учебной дисциплины

В структуре учебной дисциплины выделяются следующие основные разделы (темы):

Раздел 1 Введение в большие данные (Понятие Data Minig. Прикладные инструменты для работы с Big Data. Технология MapRaduce. Hadoop.);

Тема 1.1 Понятие Data Mining (Основные понятия Data Mining. Понятия аналитической обработки данных, витрины данных. Data Warehouse. Data Lakes. Современные методы построения хранилищ данных.);

Тема 1.2 Прикладные инструменты для работы с Big Data (Инструменты для работы с большими данными. Технологии параллельной (parallel) обработки информации. Технология mpp. Общие сведения о пакетах Apache Hadoop, Apache Spark, Kafka, RabbitMQ.);

Тема 1.3 Технология MapReduce (Технология работы с большими массивами данных. Основные этапы технологии MapReduce. Использование MapReduce в современных пакетах работы с большими данными.);

Тема 1.4 Apache Hadoop (Концепции Apache Hadoop. Типы узлов Hadoop. Концепция арбитра в распределенных многопоточных системах. Файловая система HDFS. Принцип работы Hadoop.);

Тема 1.5 Apache Spark (Концепции Apache Spark. Типы узлов Spark. Механизмы передачи и обработки данных Apache Spark.

Файловая система RDD. Разъединенные универсальные коллекции Apache Spark.);

Раздел 2 Технологии анализа данных (Жизненный цикл анализа больших данных, стандарты. Когнитивный анализ данных. Визуализация больших данных.);

Тема 2.1 Жизненный цикл анализа больших данных (Основные этапы подготовки и хранения больших данных для анализа. Сбор больших данных. Подготовка к анализу больших данных. Анализ данных.);

Тема 2.2 Статистический анализ больших данных (Методы анализа данных. Цель анализа данных. Синтаксический анализ данных. Когнитивный анализ данных. Способы и технологии когнитивного анализа данных);

Тема 2.3 Визуализация больших данных (Задача визуализации данных. Технологии визуализации данных. Инструментарий визуализации больших данных).

6 Составитель(и):

доцент Добрынин Алексей Сергеевич (кафедра автоматизации и информационных систем).