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 2023

Новокузнецк
2023

СибГИУ
Подлинник электронного документа хранится в системе электронного согласования документов СибГИУ
ДОКУМЕНТ ПОДПИСАН ЭП
Сертификат: 861a8d56fc2cfb6de53bd665403c592fd02b1419
Владелец: Ректор Юрьев Алексей Борисович
Действителен: с 12 февраля 2024 г. по 7 мая 2025 г.
Дата и время: 01-03-2024 01:21:12 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 части, формируемой участниками образовательных отношений **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, умения и навыки, получаемые по одновременно изучаемым и последующим дисциплинам:

– Проектирование информационно-управляющих систем;

– Программное обеспечение информационно-управляющих систем;

– Выполнение и защита выпускной квалификационной работы.

3 Планируемые результаты обучения по учебной дисциплине

Процесс изучения учебной дисциплины направлен на формирование следующих компетенций:

– Профессиональные компетенции

Наименование категории	Код и наименование	Код и наименование индикатора	Планируемые результаты обучения

(группы) ПК	ПК	достижения ПК	
	ПК-6: Способен осуществлять подготовку данных для проведения аналитических работ по исследованию больших данных	ПК-6.2 Производит извлечение, проверку и очистку больших объемов данных из гетерогенных источников	– знать: возможности, имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данных; теоретические и прикладные основы анализа больших данных; современные методы и инструментальные средства анализа больших данных; современный опыт использования анализа больших данных. – уметь: использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, в том числе в режиме реального времени; производить очистку данных для проведения аналитических работ; проводить интеграцию и преобразование больших объемов данных. – владеть: инструментальными средствами для извлечения, преобразования, хранения и обработки данных из разнородных источников.

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;

Тема 1.2 Прикладные инструменты для работы с Big Data;

Тема 1.3 Технология MapRaduce;

Тема 1.4 Apache Hadoop;

Тема 1.5 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, стандарты;

Тема 2.2 Когнитивный анализ данных;

Тема 2.3 Визуализация больших данных.

5 Перечень тем лекций

№ раздела / темы дисциплины	Темы лекций	Трудоемкость, *академ.час*	
		всего	в форме практической подготовки
Раздел 1.	Предпосылки формирования тренда больших данных	4	
Тема 1.1.	Понятие Data Minig. Основные вызовы больших данных.	4	
Тема 1.2.	Прикладные инструменты для работы с Big Data	4	
Тема 1.3; Тема 1.4.	Базовое представление о Map Reduce и Apache Hadoop, а также Apache Spark	4	
Раздел 2.	Технологии анализа данных. Представление о работе аналитика. Процесс аналитики. Задачи и компетенции аналитиков Big Data. Big Data как рынок. Стек технологий. Игроки на рынке BD. Крупнейшие проекты BD в России.	4	
Тема 2.1.	Жизненный цикл анализа больших данных. Стандарты жизненного цикла Big Data: CRISP-DM. Принципы и инструменты аналитики.	4	
Тема 2.2.	Когнитивный анализ данных. Введение в Data Mining – понятие, структура, составляющие и сопутствующие науки. Задачи Data Mining и способы их решения. Классификация методов DM. Классы систем DM. Процесс накопления и анализа данных: Азбука когнитивного анализа.	4	
Тема 2.3.	Шкалы измерений. Генеральная совокупность и выборка. Нормальное распределение. Уровень статистической	4	

	достоверности. Свойства описательных статистик (Дескриптивный анализ). Визуальное представление данных. Меры изменчивости.		
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.	Принципы работы с большими данными без использования СУБД	1	
Раздел 1.	Ознакомление с возможностями python для анализа данных	1	
Раздел 1.	Способы подготовки и отображения данных в Python. Возможности ввода/вывода	1	
Раздел 1; Раздел 2.	Решение задач на больших графах	2	
Раздел 2.	Способы анализа данных в Python. Получение первичных элементарных характеристик о наборах данных (элементарные статистики). Способы импорта/экспорта данных.	1	
Раздел 2.	Работа с диаграммами и графиками Python с использованием Pandas/cuML	1	
Раздел 2.	Корреляционный анализ и регрессионный анализ данных	2	
Раздел 1; Раздел 2.	Решение задач Data Mining	2	
Раздел 1; Раздел 2.	Задачи классификации, кластеризации: деревья решений, RandomForest, k-means.	2	
Раздел 1; Раздел 2.	Развертывание локального кластера Apache Hadoop. Подсчёт слов в тексте, с помощью MapReduce. Параллельное	2	

	редактирование текста с использованием регулярных выражений (на базе подсчета текста). Подсчет статистик в тексте.		
Раздел 1; Раздел 2.	Создание, настройка и запуск Spark проекта в Intellij Idea (перенос кода из Hadoop)	1	
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; 3. Подготовка к практическому занятию; 4. Прохождение тестирования.	30	
Раздел 2.	1. Выполнение домашнего задания; 2. Оформление отчета о практической работе; 3. Подготовка к практическому занятию; 4. Прохождение	21	

	тестирования.		
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: 30.07.2023);

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: 30.07.2023);

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: 30.07.2023);

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: 30.07.2023);

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: 30.07.2023);

6 Лэм, Ч. Hadoop в действии : монография / Чак Лэм. – Москва : ДМК-пресс, 2012. – 424 с. – ISBN 978-5-94074-785-7. – URL: https://www.studentlibrary.ru/book/ISBN9785940747857.html (дата обращения: 30.07.2023).

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 Национальная электронная библиотека (НЭБ) : информационная система / ФГБУ «РГБ». – Москва, [2015 –]. – URL: http://rusneb.ru. – Режим доступа: по подписке;

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

6 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;

7 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;

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

9 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;

– Apache;

– LibreOffice.

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 части, формируемой участниками образовательных отношений **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, умения и навыки, получаемые по одновременно изучаемым и последующим дисциплинам:

– Проектирование информационно-управляющих систем;

– Программное обеспечение информационно-управляющих систем;
– Выполнение и защита выпускной квалификационной работы.

3 Планируемые результаты обучения по учебной дисциплине

Процесс изучения учебной дисциплины направлен на формирование следующих компетенций:

– Профессиональные компетенции

Наименование категории (группы) ПК	Код и наименование ПК	Код и наименование индикатора достижения ПК	Планируемые результаты обучения
	ПК-6: Способен осуществлять подготовку данных для проведения аналитических работ по исследованию больших данных	ПК-6.2 Производит извлечение, проверку и очистку больших объемов данных из гетерогенных источников	– знать: возможности, имеющейся у исполнителя методологической и технологической инфраструктуры анализа больших данных; теоретические и прикладные основы анализа больших данных; современные методы и инструментальные средства анализа больших данных; современный опыт использования анализа больших данных. – уметь: использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, в том числе в режиме реального времени; производить очистку данных для проведения аналитических работ; проводить интеграцию и преобразование больших объемов данных. – владеть: инструментальными средствами для

			извлечения, преобразования, хранения и обработки данных из разнородных источников.

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;

Тема 1.2 Прикладные инструменты для работы с Big Data;

Тема 1.3 Технология MapRaduce;

Тема 1.4 Apache Hadoop;

Тема 1.5 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, стандарты;

Тема 2.2 Когнитивный анализ данных;

Тема 2.3 Визуализация больших данных.

6 Составитель(и):

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