Министерство науки и высшего образования Российской Федерации

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

Кафедра автоматизации и информационных систем

УТВЕРЖДАЮ
Проректор по учебной и
воспитательной работе
_____ М.В. Темлянцев
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
Заочная форма

Срок обучения: 4 года 6 месяцев

Год начала подготовки 2022

Новокузнецк
2022

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 00:20:57 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;
– Современные СУБД и хранилища данных;
– Базы данных;
– Операционные системы;
– Основы программирования на Python.

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	Код и наименование	Код и наименование	Планируемые

категории (группы) ПК	ПК	индикатора достижения ПК	результаты обучения
	ПК-6: Способен проектировать программные комплексы, базы данных, автоматизированные информационные системы на основе современных инструментальных средств и технологий программирования	ПК-6.2 Производит извлечение больших объемов данных из гетерогенных источников и подготовку их для анализа	– 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 обучающихся с педагогическим работником включает в себя занятия лекционного типа (лекции), занятия семинарского типа (семинары, практические занятия, практикумы), промежуточную аттестацию обучающихся и иную контактную работу, предусматривающую групповую или индивидуальную работу обучающихся с педагогическим работником. Контактная работа обучающихся с педагогическим работником может быть аудиторной, внеаудиторной, а также проводиться в электронной информационно-образовательной среде.

Объем учебной дисциплины

Сессия / курс		***ИТОГО***	***1 сессия / 4 курс***	***2 сессия / 4 курс***
Форма промежуточной аттестации				*экзамен*
Трудоёмкость	*академ. час.*	***144***	*36*	*108*
	зачетных единиц	***4***	*1*	*3*
Лекции, *академ. час.*		***2***	*2*	*0*
в форме практической подготовки		***0***	*0*	*0*
Лабораторные работы, *академ. час.*		***0***	*0*	*0*
в форме практической подготовки		***0***	*0*	*0*
Практические занятия, *академ. час.*		***4***	*0*	*4*
в форме практической подготовки		***0***	*0*	*0*
Курсовая работа / проект, *академ. час.*		***0***	*0*	*0*
в форме практической подготовки		***0***	*0*	*0*
Консультации, *академ. час.*		***0***	*0*	*0*
в форме практической подготовки		***0***	*0*	*0*
Самостоятельная работа, *академ. час.*		***129***	*34*	*95*
в форме практической подготовки		***0***	*0*	*0*
Контроль, *академ. час.*		***9***	*0*	*9*
в форме практической подготовки		***0***	*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 Minig;

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 2.1.1 Индустриальный жизненный цикл;
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	0.2	
Тема 1.1.	Понятие Data Minig. Основные вызовы больших данных.	0.2	
Тема 1.2.	Прикладные инструменты для работы с Big Data	0.2	
Тема 1.3; Тема 1.4.	Базовое представление о Map Reduce и Apache Hadoop, а также Apache Spark	0.6	
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.	0.2	
Тема 2.1.	Жизненный цикл анализа больших данных. Стандарты жизненного цикла Big Data: CRISP-DM. Принципы и инструменты аналитики.	0.2	
Тема 2.2.	Когнитивный анализ данных. Введение в Data Mining – понятие, структура, составляющие	0.2	

	и сопутствующие науки. Задачи Data Mining и способы их решения. Классификация методов DM. Классы систем DM. Процесс накопления и анализа данных: Азбука когнитивного анализа.		
Тема 2.3.	Шкалы измерений. Генеральная совокупность и выборка. Нормальное распределение. Уровень статистической достоверности. Свойства описательных статистик (Дескриптивный анализ). Визуальное представление данных. Меры изменчивости.	0.2	
Итого:		**2**	**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	0.1	
Раздел 1.	Ознакомление с возможностями python для анализа данных	0.2	
Раздел 1.	Способы подготовки и отображения данных в Python. Возможности ввода/вывода	0.2	
Раздел 1; Раздел 2.	Решение задач на больших графах.	0.2	
Раздел 2.	Способы анализа данных в Python. Получение первичных элементарных характеристик о наборах данных (элементарные статистики). Способы импорта/экспорта данных.	0.2	
Раздел 2.	Работа с диаграммами и графиками Python с использованием Pandas/cuML	0.2	
Раздел 2.	Корреляционный анализ и	0.5	

	регрессионный анализ данных		
Раздел 1.	Решение задач Data Mining	0.5	
Раздел 1; Раздел 2.	Задачи классификации, кластеризации: деревья решений, RandomForest, k-means.	0.8	
Раздел 1; Раздел 2.	Развертывание локального кластера Apache Hadoop. Подсчёт слов в тексте, с помощью MapReduce. Параллельное редактирование текста с использованием регулярных выражений (на базе подсчета текста). Подсчет статистик в тексте.	1	
Раздел 1; Раздел 2.	Создание, настройка и запуск Spark проекта в Intellij Idea (перенос кода из Hadoop)	0.1	
Итого:		**4**	**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	80	

	задания; 2. Оформление отчета о практической работе; 3. Подготовка к практическому занятию.		
Раздел 2.	1. Выполнение домашнего задания; 2. Контрольная работа; 3. Оформление отчета о практической работе; 4. Подготовка к практическому занятию.	49	
Контроль	*Подготовка к экзамену*	9	
Итого:		**138**	**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: 12.03.2022);

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: 12.03.2022);

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: 12.03.2022);

4 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: 12.03.2022).

б) ресурсы информационно-телекоммуникационной сети «Интернет»:

1 Консультант студента. Электронная библиотека технического ВУЗа : электронно-библиотечная система / ООО «Политехресурс». – 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: http://www.biblioclub.ru. – Режим доступа: для авторизир. пользователей;

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

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

в) лицензионное и свободно распространяемое программное обеспечение:

– Apache Hadoop;
– Apache Spark;
– Linux CentOS;
– Linux RedHat enterprise Server;
– Microsoft Windows 7.

г) базы данных и информационно-справочные системы:

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

2 Система ГАРАНТ : электронный периодический справочник / ООО «Правовой центр «Гарант». – Кемерово, [200 –]. – Режим доступа: компьютерная сеть 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.;

4 Электронный реферативный журнал (ЭлРЖ) : база данных / ВИНИТИ РАН. – Москва, [200 –]. – Режим доступа: компьютерная сеть библиотеки 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):

ведущий специалист по информатизации, старший преподаватель Койнов Роман Сергеевич (кафедра автоматизации и информационных систем).

Рабочая программа дисциплины рассмотрена и утверждена на заседании кафедры.

Приложение А

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 – Заочная 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;
– Современные СУБД и хранилища данных;
– Базы данных;
– Операционные системы;
– Основы программирования на Python.

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 проектировать программные комплексы, базы данных, автоматизированные информационные системы на основе современных инструментальных средств и технологий программирования	ПК-6.2 Производит извлечение больших объемов данных из гетерогенных источников и подготовку их для анализа	– 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

Сессия / курс		***ИТОГО***	***1 сессия / 4 курс***	***2 сессия / 4 курс***
Форма промежуточной аттестации				*экзамен*
Трудоёмкость	*академ. час.*	***144***	*36*	*108*
	зачетных единиц	***4***	*1*	*3*
Лекции, *академ. час.*		***2***	*2*	*0*
в форме практической подготовки		***0***	*0*	*0*
Лабораторные работы, *академ. час.*		***0***	*0*	*0*
в форме практической подготовки		***0***	*0*	*0*
Практические занятия, *академ. час.*		***4***	*0*	*4*
в форме практической подготовки		***0***	*0*	*0*
Курсовая работа / проект, *академ. час.*		***0***	*0*	*0*
в форме практической подготовки		***0***	*0*	*0*
Консультации, *академ. час.*		***0***	*0*	*0*
в форме практической подготовки		***0***	*0*	*0*
Самостоятельная работа, *академ. час.*		***129***	*34*	*95*
в форме практической подготовки		***0***	*0*	*0*
Контроль, *академ. час.*		***9***	*0*	*9*
в форме практической подготовки		***0***	*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 Minig;

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 2.1.1 Индустриальный жизненный цикл;

Тема 2.2 Когнитивный анализ данных;

Тема 2.3 Визуализация больших данных.

6 Составитель(и):

ведущий специалист по информатизации, старший преподаватель Койнов Роман Сергеевич (кафедра автоматизации и информационных систем).