

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
 Федеральное государственное автономное образовательное учреждение высшего образования
 «НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
 ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
ПРИЕМ 2019г.
ФОРМА ОБУЧЕНИЯ очная

Инструментарий для работы с большими данными

Направление подготовки/ специальность	09.04.04 Программная инженерия		
Образовательная программа (направленность (профиль))	Big Data Solutions / Технологии больших данных (Язык обучения: Английский)		
Специализация	Big Data Solutions / Технологии больших		
Уровень образования	высшее образование - магистратура		
Курс	1	семестр	2
Трудоемкость в кредитах (зачетных единицах)	3		
Виды учебной деятельности	Временной ресурс		
Контактная (аудиторная) работа, ч	Лекции		16
	Лабораторные занятия		16
	ВСЕГО		32
Самостоятельная работа, ч			76
ИТОГО, ч			108

Вид промежуточной аттестации	Экзамен	Обеспечивающее подразделение	ОИТ ИШИТР
Заведующий кафедрой- руководитель ОИТ на правах кафедры			Шерстнев В.С.
Руководитель ООП			Савельев А.О.
Преподаватель			Губин М.Ю.

1. Цели освоения дисциплины

Целями освоения дисциплины является формирование у обучающихся определенного ООП (п. 5.4 Общей характеристики ООП) состава компетенций для подготовки к профессиональной деятельности.

Код компетенции	Наименование компетенции	Индикаторы достижения компетенций		Составляющие результатов освоения (дескрипторы компетенций)	
		Код индикатора	Наименование индикатора достижения	Код	Наименование
ОПК(У)-3	Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	И.ОПК (У)-3.1	Анализирует профессиональную информацию оформляет и представляет ее в виде аналитических обзоров.	ОПК(У)-3.1В1	Владеет методами анализа и исследования информационных процессов и технологий
				ОПК(У)-3.1У1	Умеет выполнять анализ и проводить исследования статистически собранных данных
				ОПК(У)-3.1З1	Знает теоретические основы обработки статистически накопленной информации и методы её анализа
		И.ОПК (У)-3.2	Структурирует профессиональную информацию выделяя в ней основные элементы: цели, гипотезы, результаты, теории, классификации, аргументы и т.п.	ОПК(У)-3.2В1	Владеет навыками анализа современных достижений и методами подготовки научных докладов, публикаций и аналитических обзоров с обоснованными выводами и рекомендациями в области информационных технологий
				ОПК(У)-3.2У1	Умеет структурировать профессиональную информацию, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

2. Место дисциплины (модуля) в структуре ООП

Дисциплина относится к Вариативной части Междисциплинарного профессионального модуля учебного плана образовательной программы.

3. Планируемые результаты обучения по дисциплине

После успешного освоения дисциплины будут сформированы результаты обучения:

Планируемые результаты обучения по дисциплине		Индикатор достижения компетенции
Код	Наименование	
РД1	Применять знания подходов к обработке больших данных для определения, какая экосистема и в какой конфигурации может быть использована для решения конкретной задачи обработки больших данных.	И.ОПК (У)-3.1 И.ОПК (У)-3.2
РД2	Применять навыки администрирования хранилищ больших данных для планирования конфигурации, развертывания, настройки и администрирования хранилищ Hadoop.	И.ОПК (У)-3.1 И.ОПК (У)-3.2
РД 3	Уметь налаживать обмен данными между кластерами Hadoop и различными источниками и потребителями данных.	И.ОПК (У)-3.1 И.ОПК (У)-3.2
РД 4	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, анализ и интерпретацию полученных данных, в области больших данных.	И.ОПК (У)-3.1 И.ОПК (У)-3.2
РД 5	Уметь применять для обработки больших данных, находящихся в кластере Hadoop, инструментарий экосистемы Hadoop (Pig, Hive, Spark)	И.ОПК (У)-3.1 И.ОПК (У)-3.2

Оценочные мероприятия текущего контроля и промежуточной аттестации представлены в календарном рейтинг-плане дисциплины.

4. Структура и содержание дисциплины

Основные виды учебной деятельности

Разделы дисциплины	Формируемый результат обучения по дисциплине	Виды учебной деятельности	Объем времени, ч.
Раздел 1. Введение в обработку больших данных	РД1-РД5	Лекции	8
		Лабораторные занятия	8
		Самостоятельная работа	38
Раздел 2. Hadoop для обработки больших данных	РД1-РД5	Лекции	8
		Лабораторные занятия	8
		Самостоятельная работа	38

Содержание разделов дисциплины:

Раздел 1. Введение в обработку больших данных

Введение, обзор имеющихся технологий для обработки больших данных, излагаются особенности работы с большими данными, возникающие проблемы и их решения, рассматривается установка и настройка кластера Hadoop.

Темы лекций:

1. Концепции данных (2 ч.)
2. Центральные технологии экосистемы Hadoop (2 ч.)
3. Планирование, установка и конфигурирование кластера Hadoop (2 ч.)
4. Администрирование и управление кластером Hadoop (2 ч.)

Темы практических занятий:

1. Концепции больших данных на практике (2 ч.)
2. Основы экосистемы Hadoop (2 ч.)
3. Планирование кластера Hadoop (2 ч.)
4. Основы администрирования кластера Hadoop (2 ч.)

Названия лабораторных работ:

1. Введение в обработку больших данных (2 ч.)
2. Введение в Hadoop (2 ч.)
3. Установка и конфигурирование кластера Hadoop (2 ч.)
4. Администрирование кластера Hadoop (2 ч.)

Раздел 2. Hadoop для обработки больших данных

Описываются различные методы работы с данными в экосистеме Hadoop: с использованием технологий Pig, Spark и Hive. Рассматриваются преимущества, недостатки и границы применимости технологий.

Темы лекций:

1. Hadoop Distributed File System (HDFS). Обмен данных с кластером Hadoop (2 ч.)
2. Apache Pig (2 ч.)
3. Apache Hive (2 ч.)
4. Apache Spark (2 ч.)

Темы практических занятий:

1. Основы команд HDFS (2 ч.)
2. Оптимизация запросов Apache Pig (2 ч.)
3. Разработка архитектуры хранилища Apache Hive (2 ч.)
4. Resilient Distributed Dataset. Особенности работы с Apache Spark (2 ч.)

Названия лабораторных работ:

1. HDFS. Импорт и экспорт данных в кластер. (2 ч.)
2. Обработка данных с помощью Apache Pig (2 ч.)
3. Обработка данных с помощью Apache Hive (2 ч.)

4. Обработка данных и машинное обучение с помощью Apache Spark (2 ч.)

5. Организация самостоятельной работы студентов

Самостоятельная работа студентов при изучении дисциплины (модуля) предусмотрена в следующих видах и формах:

- Работа с лекционным материалом, поиск и обзор литературы и электронных источников информации по индивидуально заданной проблеме курса;
- Работа в электронном курсе (изучение теоретического материала, выполнение индивидуальных заданий и контролирующих мероприятий и др.);
- Изучение тем, вынесенных на самостоятельную проработку;
- Поиск, анализ, структурирование и презентация информации;
- Подготовка к лабораторным работам, к практическим и семинарским занятиям;
- Работа над междисциплинарным проектом;
- Исследовательская работа и участие в научных студенческих конференциях, семинарах и олимпиадах;
- Анализ научных публикаций по заранее определенной преподавателем теме;
- Подготовка к оценивающим мероприятиям.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Учебно-методическое обеспечение

1. White T. Hadoop: The definitive guide. – " O'Reilly Media, Inc.", 2012. [Электронный ресурс] Схема доступа:
<http://www.academia.edu/download/61663102/Oreilly.Hadoop.The.Definitive.Guide.3rd.Edition.Jan.201220200102-77747-1tk6dyq.pdf>
2. Остроух, А. В. Теория проектирования распределенных информационных систем : монография / А. В. Остроух, А. В. Помазанов. — Санкт-Петербург : Лань, 2019. — 96 с. — ISBN 978-5-8114-3417-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/116390> (дата обращения: 16.02.2019). —

6.2. Информационное и программное обеспечение

Internet-ресурсы (в т.ч. в среде LMS MOODLE и др. образовательные и библиотечные ресурсы):

1. <https://kaggle.com/> – ресурс для освоения подходов машинного обучения.

Лицензионное программное обеспечение (в соответствии с **Перечнем лицензионного программного обеспечения ТПУ**):

Для реализации курса, специализированного лицензионного программного обеспечения, не требуется.

7. Особые требования к материально-техническому обеспечению дисциплины

В учебном процессе используется следующее лабораторное оборудование для практических и лабораторных занятий:

№	Наименование специальных помещений	Наименование оборудования
1.	Аудитория для проведения учебных занятий всех типов, курсового проектирования, консультаций, текущего контроля и промежуточной аттестации (компьютерный класс) Компьютер - 15 шт. Доска аудиторная настенная - 1 шт. 634034, Томская область, г. Томск, Советская улица, 84/3 212	Шкаф для документов - 1 шт.;Комплект учебной мебели на 10 посадочных мест; Комплект (стол, кресло) - 4 шт.;
	Аудитория для проведения учебных занятий всех типов, курсового проектирования, консультаций, текущего контроля и промежуточной аттестации Проектор - 1 шт.; Компьютер - 1 шт. 634034, Томская область, г. Томск, Советская улица, 84/3 313	Доска аудиторная настенная - 1 шт. Комплект учебной мебели на 36 посадочных мест;

Рабочая программа составлена на основе Общей характеристики образовательной программы по направлению 09.04.04 Программная инженерия/ профиль Big Data Solutions / Технологии больших данных /специализации Big Data Solutions/Технологии больших данных (приема 2019 г., очная форма обучения).

Разработчик:

Должность	Подпись	ФИО
Преподаватель ОИТ ИШИТР		Губин М.Ю.

Программа одобрена на заседании Отделения информационных технологий ИШИТР (протокол №13 от 19.06.2019 г.).

Заведующий кафедрой –
руководитель ОИТ на правах кафедры


подпись /Шерстнев В.С.

Лист изменений рабочей программы дисциплины:

Учебный год	Содержание /изменение	Обсуждено на заседании ОИТ ИШИТР (протокол)
2020/2021	1. Обновлено программное обеспечение	от 01.09.2020г. № 19