

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 02.06.2025 11:40:21
Уникальный программный ключ:
sa953a01204891083f939673078ef1a989dae18a

Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы»
Факультет искусственного интеллекта
(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

HADOOP, SPARK

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

**02.03.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ,
09.03.03 ПРИКЛАДНАЯ ИНФОРМАТИКА**

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: РАЗРАБОТКА И ОБУЧЕНИЕ
ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ**

(наименование (профиль/специализация) ОП ВО)

2025 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Hadoop, SPARK» входит в программу бакалавриата «Искусственный интеллект: разработка и обучение интеллектуальных систем» по направлению 02.03.02 «Фундаментальная информатика и информационные технологии» и изучается в 5 семестре 3 курса. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 6 разделов и 36 тем и направлена на изучение получения студентами фундаментальных теоретических и практических знаний в области распределённых вычислений и обработки больших данных с использованием ведущих платформ — Hadoop и Apache Spark. Курс формирует понимание архитектуры, компонент, принципов работы и практического программирования вычислительных кластеров, что критично для современных задач Data Science, AI/ML и промышленных аналитических систем.

Целью освоения дисциплины является научить студентов использовать Hadoop и Spark для эффективной работы с большими объемами данных: строить распределённые хранилища, проектировать и реализовывать задачи параллельной обработки и аналитики, масштабировать обработку, обеспечивать отказоустойчивость и интеграцию с экосистемой бигдата.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Hadoop, SPARK» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
ОПК-3	Способен к разработке алгоритмических и программных решений в области системного и прикладного программирования, математических, информационных и имитационных моделей, созданию информационных ресурсов глобальных сетей, образовательного контента, прикладных баз данных, тестов и средств тестирования систем и средств на соответствие стандартам и исходным требованиям	ОПК-3.1 Знает методы теории алгоритмов, методы системного и прикладного программирования, основные положения и концепции в области математических, информационных и имитационных моделей; ОПК-3.2 Умеет соотносить знания в области программирования, интерпретацию прочитанного, определять и создавать информационные ресурсы глобальных сетей, образовательного контента, средств тестирования систем; ОПК-3.3 Имеет практический опыт применения разработки программного обеспечения;
ПК-2	Способен эффективно работать с большими объемами данных, включая их предварительную обработку, анализ и визуализацию, с целью извлечения полезной информации для обучения моделей искусственного интеллекта	ПК-2.2 Демонстрирует навыки анализа данных с использованием статистических методов и инструментов; ПК-2.3 Владеет методами работы с различными алгоритмами машинного обучения и глубокого обучения для решения различных задач;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Hadoop, SPARK» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Hadoop, SPARK».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
ОПК-3	Способен к разработке алгоритмических и программных решений в области системного и прикладного программирования, математических, информационных и имитационных моделей, созданию информационных ресурсов глобальных сетей, образовательного контента, прикладных баз данных, тестов и средств тестирования систем и средств на соответствие стандартам и исходным требованиям	Технологическая (проектно-технологическая) практика (учебная); Эксплуатационная практика (учебная); Программирование на языке Python; Введение в базы данных; Методы разработки решений на основе искусственного интеллекта (Git, Docker); Программирование на языке C++;	Технологическая (проектно-технологическая) практика (производственная); Эксплуатационная практика (производственная); Прикладные задачи машинного обучения; Оптимизация моделей машинного обучения;
ПК-2	Способен эффективно работать с большими объемами данных, включая их предварительную обработку, анализ и визуализацию, с целью извлечения полезной информации для обучения моделей искусственного интеллекта	Эксплуатационная практика (учебная); Технологическая (проектно-технологическая) практика (учебная); Статистические методы и первичный анализ данных; Введение в базы данных; Программирование на языке Python; Лингвистические основы анализа естественного языка; Введение в компьютерное зрение; Программирование на языке C++; <i>Программирование на языке NodeJS**;</i> <i>Программирование на языке Go**;</i>	Преддипломная практика; Технологическая (проектно-технологическая) практика (производственная); Эксплуатационная практика (производственная); <i>Информационный поиск**;</i> <i>Анализ временных рядов**;</i> Нейронные сети; Оптимизация моделей машинного обучения; Практикум по обработке естественного языка (NLP); Основы глубокого обучения; Проектирование и разработка систем компьютерного зрения;

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Nadoor, SPARK» составляет «4» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			5
Контактная работа, ак.ч.	72		72
Лекции (ЛК)	36		36
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	36		36
Самостоятельная работа обучающихся, ак.ч.	45		45
Контроль (экзамен/зачет с оценкой), ак.ч.	27		27
Общая трудоемкость дисциплины	ак.ч.	144	144
	зач.ед.	4	4

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
Раздел 1	Введение в Big Data и основы Hadoop	1.1	Введение в парадигму Big Data и задачи распределённой обработки	ЛК
		1.2	Архитектура Hadoop: компоненты HDFS и MapReduce	ЛК
		1.3	Масштабируемость и отказоустойчивость в Hadoop	ЛК
		1.4	Анализ кейсов применения технологий Big Data	СЗ
		1.5	Логическая структура кластера Hadoop	СЗ
		1.6	Примеры распределённого хранения и копирования файлов в HDFS	СЗ
Раздел 2	HDFS и основы программирования MapReduce	2.1	Принципы работы HDFS: файловая система, репликация, управление файлами	ЛК
		2.2	Модель программирования MapReduce: концепция, рабочий процесс, пример	ЛК
		2.3	Проблемы, связанные с архитектурой MapReduce, и их решения	ЛК
		2.4	Практика работы с HDFS: загрузка, извлечение, удаление файлов	СЗ
		2.5	Построение простейшего приложения MapReduce	СЗ
		2.6	Отладка и мониторинг MapReduce задач	СЗ
Раздел 3	Экосистема Hadoop: инструменты, расширения, интеграция	3.1	Обзор инструментов экосистемы: Hive, Pig, HBase, Sqoop	ЛК
		3.2	ETL-процессы на базе Hadoop: принципы и сценарии использования	ЛК
		3.3	Интеграция Hadoop с внешними источниками данных	ЛК
		3.4	Запросы на Hive: SQL-подобные операции	СЗ
		3.5	Исполнение скриптов Pig для трансформации данных	СЗ
		3.6	Практика интеграции Sqoop с реляционными БД	СЗ
Раздел 4	Введение в Apache Spark: архитектура и компоненты	4.1	Архитектура Spark: RDD, DAG, resiliency	ЛК
		4.2	Компоненты Spark: Spark SQL, Streaming, MLlib, GraphX	ЛК
		4.3	Различия Spark и Hadoop MapReduce: преимущества и сценарии использования	ЛК
		4.4	Инициализация Spark-кластера и конфигурирование окружения	СЗ
		4.5	Основы работы с RDD: создание, трансформации, действия	СЗ
		4.6	Сравнительный анализ выполнения задачи на Hadoop и Spark	СЗ
Раздел 5	Работа с данными в Spark: обработка, анализ, ML	5.1	Spark DataFrame и Spark SQL: обработка структурированных данных	ЛК
		5.2	Модуль Spark MLlib: базовые алгоритмы машинного обучения	ЛК
		5.3	Spark Streaming: обработка потоковых данных	ЛК
		5.4	Задания на анализ данных с помощью DataFrame	СЗ
		5.5	Построение простых моделей MLlib (классификация, регрессия)	СЗ
		5.6	Демонстрация потоковой обработки с использованием Spark Streaming	СЗ
Раздел 6	Производственная эксплуатация Hadoop & Spark, безопасность и	6.1	Развертывание и администрирование кластера Hadoop/Spark	ЛК
		6.2	Безопасность, контроль доступа, мониторинг и	ЛК

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
	современные тренды		аудит	
		6.3	Облачные решения и будущее развития Big Data (integration AWS, Azure, Google Cloud)	ЛК
		6.4	Практика настройки отказоустойчивого кластера	СЗ
		6.5	Организация мониторинга (Ambari, Grafana)	СЗ
		6.6	Дискуссия: вызовы и перспективы Big Data-проектов, современные тренды	СЗ

* - заполняется только по **ОЧНОЙ** форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	
Семинарская	Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и техническими средствами мультимедиа презентаций.	
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Лэм Чак. Hadoop в действии [Электронный ресурс]. - М.: ДМК Пресс, 2019. 424 с. ISBN 978-5-97060-723-7 URL:

https://mega.rudn.ru/MegaPro/UserEntry?Action=Link_FindDoc&id=475248&idb=0

2. Карау Х., Конвински Э., Венделл П., Захария М. Изучаем Spark: молниеносный анализ данных. - М.: ДМК Пресс, 2015. - 304 с.: ил. ISBN 978-5-97060-323-9

Дополнительная литература:

1. Сэнди Риза, Ури Лезерсон, Шон Оуэн, Джош Уиллс. Spark для профессионалов: современные паттерны обработки больших данных. — СПб.: Питер, 2017. — 272 с.: ил. —

(Серия «Бестселлеры O'Reilly»). ISBN 978-5-496-02401-3

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров

- Электронно-библиотечная система РУДН – ЭБС РУДН

<https://mega.rudn.ru/MegaPro/Web>

- ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>

- ЭБС «Юрайт» <http://www.biblio-online.ru>

- ЭБС «Консультант студента» www.studentlibrary.ru

- ЭБС «Знаниум» <https://znaniyum.ru/>

2. Базы данных и поисковые системы

- Sage <https://journals.sagepub.com/>

- Springer Nature Link <https://link.springer.com/>

- Wiley Journal Database <https://onlinelibrary.wiley.com/>

- Научометрическая база данных Lens.org <https://www.lens.org>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля:*

1. Курс лекций по дисциплине «Hadoop, SPARK».

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

РАЗРАБОТЧИК:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность, БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность, БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность, БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.