Министерство науки и высшего образования Российской Федерации

Муромский институт (филиал)

федерального государственного бюджетного образовательного учреждения высшего образования

«Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»

(МИ ВлГУ)

Кафедра *УКТС*

«УТВЕРЖДАЮ»
Заместитель директора по УР
_____Д.Е. Андрианов
20.05.2025

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Аналитика потоковых данных

Направление подготовки *12.03.01 Приборостроение*

Профиль подготовки *Программирование робототехнических систем*

Семестр	Трудоемкость, час./зач. ед.	Лекции, час.	Практические занятия, час.	Лабораторные работы, час.	Консультация, час.	Контроль, час.	Всего (контактная работа), час.	СРС, час.	Форма промежуточного контроля (экз., зач., зач. с оц.)
8	**144 / 4**	**16**		**16**	**3,6**	**0,35**	**35,95**	**81,4**	**Экз.(26,65)**
Итого	**144 / 4**	**16**		**16**	**3,6**	**0,35**	**35,95**	**81,4**	**26,65**

Муром, 2025 г.

1. Цель освоения дисциплины

Цель дисциплины: формирование у студентов компетенций в области подготовки данных для дальнейшей аналитики как в режиме онлайн (OLAP-системы), так и в системах близких к реальному времени (Near Real Time), особенностей процессов и архитектур пакетного и потокового захвата изменений в данных.

Задачи:

- освоение теоретических основ обработки больших и потоковых данных;
- познание методов обработки больших и потоковых данных на практике;
- приобретение навыков определения показателей массивов больших данных, применяемых в областях своей профессиональной деятельности;
- изучение особенностей применения пакетов прикладных программ и программных сред при обработке больших данных.

2. Место дисциплины в структуре ОПОП ВО

Курс базируется на дисциплинах: хранилища данных и облачные технологии, интеллектуальная обработка мультимедиа траффика, информационная безопасность и защита информации. На данном курсе базируется выпускная квалификационная работа.

3. Планируемые результаты обучения по дисциплине

Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения ОПОП (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине, в соответствии с индикатором достижения компетенции		Наименование оценочного средства
	Индикатор достижения компетенции	Результаты обучения по дисциплине	
ПК-4 Использует технологии искусственного интеллекта при решении профессиональных задач	ПК-4.1 Применяет технологии искусственного интеллекта при построении приборов и систем	принципы объединения, обработки и анализа потоковых данных из различных источников информации (ПК-4.1) способы получения данных о взаимодействиях узлов и промежуточного оборудования локальных и глобальных сетей (ПК-4.1) использовать специализированное программное обеспечение для решения задач анализа потоковых данных (ПК-4.1) обеспечивать надёжный и эффективный сбор данных о внутрисетевых взаимодействиях (ПК-4.1) методами интеллектуального анализа потоковых данных (ПК-4.1) навыками анализа внутрисетевых взаимодействий (ПК-4.1)	отчет, тест

4. Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 4 зачетные единицы, 144 часа.

4.1. Форма обучения: очная

Уровень базового образования: среднее общее.
Срок обучения 4г.

4.1.1. Структура дисциплины

№ п\п	Раздел (тема) дисциплины	Семестр	Контактная работа обучающихся с педагогическим работником							Самостоятельная работа	Форма текущего контроля успеваемости (по неделям семестра), форма промежуточной аттестации(по семестрам)
			Лекции	Практические занятия	Лабораторные работы	Контрольные работы	КП / КР	Консультация	Контроль		
1	Особенности потоковой обработки	8	12		8					40	отчет, тестирование
2	Анализ сетевого трафика	8	4		8					41,4	отчет, тестирование
Всего за семестр		144	16		16			3,6	0,35	81,4	Экз.(26,65)
Итого		144	16		16			3,6	0,35	81,4	26,65

4.1.2. Содержание дисциплины

4.1.2.1. Перечень лекций

Семестр 8

Раздел 1. Особенности потоковой обработки

Лекция 1.

Распределенные данные и обработка, требования к скорости и надежности вычислений (2 часа).

Лекция 2.

Spark. Парадигмы, стадии обработки, типы операций, API взаимодействия (2 часа).

Лекция 3.

Spark. Работа с dataframe (2 часа).

Лекция 4.

Работа с данными типа "ключ — значение" (2 часа).

Лекция 5.

Spark. Обработка данных и ML (2 часа).

Лекция 6.

Очереди и брокеры (2 часа).

Раздел 2. Анализ сетевого трафика

Лекция 7.

Потоковая обработка с учетом состояний и основы потоковой обработки (2 часа).

Лекция 8.

Анализ сетевых взаимодействий и искусственные нейронные сети в системах информационной безопасности (2 часа).

4.1.2.2. Перечень практических занятий

Не планируется.

4.1.2.3. Перечень лабораторных работ

Семестр 8

Раздел 1. Особенности потоковой обработки

Лабораторная 1.

Установка и развертывание Apache Spark (4 часа).

Лабораторная 2.

Применение Apache Spark для считывания, обработки и записи данных (4 часа).

Раздел 2. Анализ сетевого трафика

Лабораторная 3.

Установка и развертывание Apache Kafka (4 часа).

Лабораторная 4.

Создания простейшего сервиса Apache Kafka для прослушивания источника и передачи на Apache Spark job (4 часа).

4.1.2.4. Перечень тем и учебно-методическое обеспечение самостоятельной работы

Перечень тем, вынесенных на самостоятельное изучение:

1. Особенности Spark.
2. SCC.
3. RDD.
4. YARN.
5. Spark Streaming.
6. SparkCore.
7. File System API в Spark.
8. MapReduce.
9. Apache Kafka.
10. Zookeeper в Kafka.
11. Потребители и пользователи в Kafka?.
12. Kafka в качестве системы хранения данных.
13. Разбалансировка кластера в Kafka.

Для самостоятельной работы используются методические указания по освоению дисциплины и издания из списка приведенной ниже основной и дополнительной литературы.

4.1.2.5. Перечень тем контрольных работ, рефератов, ТР, РГР, РПР

Не планируется.

4.1.2.6. Примерный перечень тем курсовых работ (проектов)

Не планируется.

4.2 Форма обучения: заочная

Уровень базового образования: среднее общее.

Срок обучения 5л.

Семестр	Трудоемкость, час./ зач. ед.	Лекции, час.	Практические занятия, час.	Лабораторные работы, час.	Консультация, час.	Контроль,час.	Всего (контактная работа), час.	СРС, час.	Форма промежуточного контроля (экз., зач., зач. с оц.)
10	**144 / 4**	**4**		**8**	**2**	**0,6**	**14,6**	**120,75**	**Экз.(8,65)**
Итого	**144 / 4**	**4**		**8**	**2**	**0,6**	**14,6**	**120,75**	**8,65**

4.2.1. Структура дисциплины

№ п\п	Раздел (тема) дисциплины	Семестр	Контактная работа обучающихся с педагогическим работником							Самостоятельная работа	Форма текущего контроля успеваемости (по неделям семестра), форма промежуточной аттестации(по семестрам)
			Лекции	Практические занятия	Лабораторные работы	Контрольные работы	КП / КР	Консультация	Контроль		
1	Особенности потоковой обработки	10	2		4					45	отчет, тестирование
2	Анализ сетевого трафика	10	2		4					75,75	отчет, тестирование
Всего за семестр		144	4		8	+		2	0,6	120,75	Экз.(8,65)
Итого		144	4		8			2	0,6	120,75	8,65

4.2.2. Содержание дисциплины

4.2.2.1. Перечень лекций

Семестр 10

Раздел 1. Особенности потоковой обработки

Лекция 1.

Распределенные данные и обработка, требования к скорости и надежности вычислений (2 часа).

Раздел 2. Анализ сетевого трафика

Лекция 2.

Потоковая обработка с учетом состояний и основы потоковой обработки (2 часа).

4.2.2.2. Перечень практических занятий

Не планируется.

4.2.2.3. Перечень лабораторных работ

Семестр 10

Раздел 1. Особенности потоковой обработки

Лабораторная 1.

Установка и развертывание Apache Spark (4 часа).

Раздел 2. Анализ сетевого трафика

Лабораторная 2.

Установка и развертывание Apache Kafka (4 часа).

4.2.2.4. Перечень тем и учебно-методическое обеспечение самостоятельной работы

Перечень тем, вынесенных на самостоятельное изучение:

1. Особенности Spark.
2. SCC.
3. RDD.
4. YARN.
5. Spark Streaming.
6. SparkCore.
7. File System API в Spark.
8. MapReduce.
9. Apache Kafka.
10. Zookeeper в Kafka.
11. Потребители и пользователи в Kafka?.
12. Kafka в качестве системы хранения данных.
13. Разбалансировка кластера в Kafka.

Для самостоятельной работы используются методические указания по освоению дисциплины и издания из списка приведенной ниже основной и дополнительной литературы.

4.2.2.5. Перечень тем контрольных работ, рефератов, ТР, РГР, РПР

1. Применение Apache Spark для считывания, обработки и записи данных.
2. Создания простейшего сервиса Apache Kafka для прослушивания источника и передачи на Apache Spark job.

4.2.2.6. Примерный перечень тем курсовых работ (проектов)

Не планируется.

5. Образовательные технологии

В процессе изучения дисциплины применяется контактная технология преподавания (за исключением самостоятельно изучаемых студентами вопросов). При проведении лабораторных работ применяется имитационный или симуляционный подход. Шаги решения задач студентам демонстрируются при помощи мультимедийной техники. В дальнейшем студенты самостоятельно решают аналогичные задания.

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.

Фонды оценочных материалов (средств) приведены в приложении.

7. Учебно-методическое и информационное обеспечение дисциплины.

7.1. Основная учебно-методическая литература по дисциплине

1. Целых, А. Н. Применение временных рядов для анализа больших данных : учебное пособие по курсу «Математические методы анализа больших данных» / А. Н. Целых, В. С. Васильев, Э. М. Котов. — Ростов-на-Дону, Таганрог : Издательство Южного федерального университета, 2021. — 84 с. - https://www.iprbookshop.ru/121929.html

2. Кадырова, Н. О. Статистический анализ больших данных: подход на основе машин опорных векторов : учебное пособие / Н. О. Кадырова, Л. В. Павлова. — Санкт-Петербург : Санкт-Петербургский политехнический университет Петра Великого, 2022. — 60 с. - https://www.iprbookshop.ru/128651.html

3. Принятие решений на основе анализа слабоструктурированных данных социальных сетей / Н. Г. Ярушкина, И. А. Андреев, А. С. Желепов, В. С. Мошкин. — Ульяновск : Ульяновский государственный технический университет, 2022. — 114 с. - https://www.iprbookshop.ru/129289.html

7.2. Дополнительная учебно-методическая литература по дисциплине

1. Воронова, Л. И. Machine Learning: регрессионные методы интеллектуального анализа данных : учебное пособие / Л. И. Воронова, В. И. Воронов. — Москва : Московский технический университет связи и информатики, 2018. — 82 с. - https://www.iprbookshop.ru/81325.html

2. Целых, А. Н. Современные методы прикладной информатики в задачах анализа данных : учебное пособие по курсу «Методы интеллектуального анализа данных» / А. Н. Целых, А. А. Целых, Э. М. Котов. — Ростов-на-Дону, Таганрог : Издательство Южного федерального университета, 2021. — 130 с. - https://www.iprbookshop.ru/117165.html

3. Лихтциндер, Б. Я. Анализ трафика мультисервисных сетей : монография / Б. Я. Лихтциндер. — Самара : Поволжский государственный университет телекоммуникаций и информатики, 2015. — 163 - https://www.iprbookshop.ru/71820.html

7.3. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем

В образовательном процессе используются информационные технологии, реализованные на основе информационно-образовательного портала института (www.mivlgu.ru/iop), и инфокоммуникационной сети института:

- предоставление учебно-методических материалов в электронном виде;
- взаимодействие участников образовательного процесса через локальную сеть института и Интернет;
- предоставление сведений о результатах учебной деятельности в электронном личном кабинете обучающегося.

Информационные справочные системы:

Набор данных для оценки обнаружения вторжений https://www.unb.ca/cic/datasets/ids-2017.html

Набор данных о сетевых атаках Kitsune https://www.kaggle.com/datasets/ymirsky/network-attack-dataset-kitsune

Программное обеспечение:

РЕД ОС (Соглашение №140/05-21У от 18.05.2021 года о сотрудничестве в области науки, развития инновационной деятельности)

Python 3.9.4 (Python Software Foundation License)

Kaspersky Endpoint Security для бизнеса – Стандартный Russian Edition (Договор №436 от 11.11.2014 года)

Open Office (Бесплатное ПО)

NetTraffic Version 2.0 (Бесплатное ПО)

Friendly Pinger 5.0.1 (Бесплатное ПО)

7.4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

iprbookshop.ru
unb.ca
kaggle.com
mivlgu.ru/iop

8. Материально-техническое обеспечение дисциплины

Лаборатория компьютерного моделирования в измерительных системах

ЭВМ Айтек Intel Core i5 2400 - 12 шт.; Лабораторный стенд изучение интерфейсов сопряжения – 12 шт. ; Видеопроектор Acer P1100 EY; Экран настенный ScreenMedia Economy-P.

9. Методические указания по освоению дисциплины

Для успешного освоения теоретического материала обучающийся: знакомится со списком рекомендуемой основной и дополнительной литературы; уточняет у преподавателя, каким дополнительным пособиям следует отдать предпочтение; ведет конспект лекций и прорабатывает лекционный материал, пользуясь как конспектом, так и учебными пособиями.

До выполнения лабораторных работ обучающийся изучает соответствующий раздел теории. Перед занятием студент знакомится с описанием заданий для выполнения работы, внимательно изучает содержание и порядок проведения лабораторной работы. Лабораторная работа проводятся в компьютерном классе. Обучающиеся выполняют индивидуальную задачу в соответствии с заданием на лабораторную работу. Полученные результаты исследований сводятся в отчет и защищаются по традиционной методике в классе на следующем лабораторном занятии. Необходимый теоретический материал, индивидуальное задание, шаги выполнения лабораторной работы и требование к отчету приведены в методических указаниях, размещенных на информационно-образовательном портале института.

Самостоятельная работа оказывает важное влияние на формирование личности будущего специалиста, она планируется обучающимся самостоятельно. Каждый обучающийся самостоятельно определяет режим своей работы и меру труда, затрачиваемого на овладение учебным содержанием дисциплины. Он выполняет внеаудиторную работу и изучение разделов, выносимых на самостоятельную работу, по личному индивидуальному плану, в зависимости от его подготовки, времени и других условий.

Форма заключительного контроля при промежуточной аттестации – экзамен. Для проведения промежуточной аттестации по дисциплине разработаны фонд оценочных средств и балльно-рейтинговая система оценки учебной деятельности студентов. Оценка по дисциплине выставляется в информационной системе и носит интегрированный характер, учитывающий результаты оценивания участия студентов в аудиторных занятиях, качества и своевременности выполнения заданий в ходе изучения дисциплины и промежуточной аттестации.

Программа составлена в соответствии с требованиями ФГОС ВО по направлению *12.03.01 Приборостроение* и профилю подготовки *Программирование робототехнических систем*
Рабочую программу составил *д.т.н., заведующий кафедрой Дорофеев Н.В.*_____

Программа рассмотрена и одобрена на заседании кафедры *УКТС*

протокол № 36 от 13.05.2025 года.
Заведующий кафедрой *УКТС* _____*Дорофеев Н.В.*
(Подпись)

Рабочая программа рассмотрена и одобрена на заседании учебно-методической комиссии факультета

протокол № 9 от 15.05.2025 года.
Председатель комиссии ФИТР _____ *Кутарова Е.И.*
(Подпись) (Ф.И.О.)

Приложение

Фонд оценочных материалов (средств) по дисциплине

Аналитика потоковых данных

1. Оценочные материалы для проведения текущего контроля успеваемости по дисциплине

Вопросы для тестирования размещены https://www.mivlgu.ru/iop/question/edit.php?courseid=4202

Общее распределение баллов текущего контроля по видам учебных работ для студентов

Рейтинг-контроль 1	1 лабораторные работы	10
Рейтинг-контроль 2	2 лабораторные работы	20
Рейтинг-контроль 3	1 лабораторные работы, тестирование	10
Посещение занятий студентом		10
Дополнительные баллы (бонусы)		5
Выполнение семестрового плана самостоятельной работы		5

2. Промежуточная аттестация по дисциплине

Перечень вопросов к экзамену / зачету / зачету с оценкой.
Перечень практических задач / заданий к экзамену / зачету / зачету с оценкой (при наличии)

Вопросы для тестирования размещены https://www.mivlgu.ru/iop/question/edit.php?courseid=4202

Вопросы для подготовки к экзамену https://www.mivlgu.ru/iop/course/view.php?id=4202

Методические материалы, характеризующие процедуры оценивания

Для оценивания сформированных у студента знаний, умений и навыков имеются типовые задания. Все типовые задания разбиты на 3 блока: блок 1 - для оценивания знаний, блок 2 - для оценивания умений, блок 3 - для оценивания навыков (владений). Каждый блок включает вопросы своего уровня сложности и оценивается определенным количеством баллов. Максимальный балл, который может набрать студент при правильном ответе на все вопросы, равняется 40.

Тест для оценки знаний, умений и навыков студента состоит из 10 вопросов и формируется на основе типовых заданий программным комплексом информационно-образовательного портала МИ ВлГУ в автоматическом режими (три вопроса из блока 1, три вопроса из блока 2 и четыре вопроса из блока 3). Программный комплекс формирует индивидуальные задания для каждого зарегистрированного в системе студента и устанавливает время прохождения тестирования. Результатом тестирования является процент правильных ответов, с учетом индивидуального семестрового рейтинга студента формируется экзаменационная оценка.

При проведении устного опроса студент отвечает на выбранные случайным образом вопросы из перечня тем и в зависимости от полноты и правильности ответа с учетом индивидуального семестрового рейтинга студента формируется экзаменационная оценка.

Максимальная сумма баллов, набираемая студентом по дисциплине равна 100.

Оценка в баллах	Оценка по шкале	Обоснование	***Уровень сформированности компетенций***
Более 80	«Отлично»	Содержание курса освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному	***Высокий уровень***
66-80	«Хорошо»	Содержание курса освоено полностью, без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками	***Продвинутый уровень***
50-65	«Удовлетворительно»	Содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий, возможно, содержат ошибки	***Пороговый уровень***
Менее 50	«Неудовлетворительно»	Содержание курса не освоено, необходимые практические навыки работы не сформированы, выполненные учебные задания содержат грубые ошибки	***Компетенции не сформированы***

3. Задания в тестовой форме по дисциплине

Примеры заданий:

1. Какие типы атак на сетевой трафик существуют?
   а) Денайал сервиса
   б) Сниффинг
   в) Спуфинг
   г) Все вышеперечисленные

2. Какие методы обнаружения аномального трафика используются в интеллектуальном анализе сетевого трафика?

а) Статистические методы
б) Машинное обучение
в) Сигнатурный анализ
г) Все вышеперечисленные

3. Какие инструменты используются для анализа сетевого трафика?

а) Wireshark
б) tcpdump
в) Snort
г) Все вышеперечисленные

4. Что такое DPI (глубокий пакетный инспектирование) и какие задачи можно решить с его помощью?

а) Определить тип приложения, генерирующего трафик
б) Обнаружить нарушения политики безопасности
в) Оптимизация производительности сети
г) Все вышеперечисленное

5. Для набора данных [“This is an error message”, “This is an info message”, “This is a success message”, “Consider this a warning”] вывести количество элементов, содержащих информацию об ошибке или предупреждение

Полный перечень тестовых заданий с указанием правильных ответов, размещен в банке вопросов на информационно-образовательном портале института по ссылке https://www.mivlgu.ru/iop/question/edit.php?courseid=4202

Оценка рассчитывается как процент правильно выполненных тестовых заданий из их общего числа.