МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ,
МИНИСТЕРСТВО НАУКИ, ВЫСШЕГО ОБРАЗОВАНИЯ И ИННОВАЦИЙ
КЫРГЫЗСКОЙ РЕСПУБЛИКИ

МОО ВО Кыргызско-Российский Славянский университет
имени первого Президента Российской Федерации Б.Н. Ельцина

УТВЕРЖДАЮ
директор Политехнического института
_____ Т.В. Зуев
3 сентября 2020 г.

Системы анализа больших данных

рабочая программа дисциплины (модуля)

Закреплена за **Высшая школа Цифровых технологий**

Учебный план g090404_26_1 пи_оаирп.plx
Направление подготовки 09.04.04 - РФ, 710400 - КР Программная инженерия
Магистерская программа "Основы анализа и разработки приложений с большими объемами распределенных данных"

Квалификация **магистр**

Форма обучения **очная**

Общая трудоемкость **10 ЗЕТ**

Часов по учебному 360

в том числе:

аудиторные занятия 34

самостоятельная 289,9

Виды контроля в семестрах:

зачет с оценкой 2,3
курсовая работа 3

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	**2 (1.2)**		**3 (2.1)**		Итого	
Недель	18		19			
Вид занятий	УП	РП	УП	РП	УП	РП
Практические	30	30	4	4	34	34
Контактная работа в период теоретического	0,1	0,1			0,1	0,1
Контактная работа в период экзаменационной			0,3	0,3	0,3	0,3
Итого ауд.	30	30	4	4	34	34
Контактная работа	30,1	30,1	4,3	4,3	34,4	34,4
Сам. работа	113,9	113,9	176	176	289,9	289,9
Часы на контроль			35,7	35,7	35,7	35,7
Итого	144	144	216	216	360	360

Программу составил(и):
доцент, к.т.н., доцент СПбПУ Никифоров И.В.; специалист 1 кат. СПбПУ Вишневская Т.А. _____

Рецензент(ы):
д.т.н., профессор, Верзунов Сергей Николаевич _____

Согласовано РОП: *Лыченко Н.М.*
_____

Рабочая программа дисциплины
Системы анализа больших данных

разработана в соответствии с ФГОС 3++:
Федеральный государственный образовательный стандарт высшего образования - магистратура по направлению подготовки 09.04.04 Программная инженерия (приказ Минобрнауки России от 19.09.2017 г. № 932)

составлена на основании учебного плана:

Направление подготовки 09.04.04 - РФ, 710400 - КР Программная инженерия
Магистерская программа "Основы анализа и разработки приложений с большими объемами распределенных данных"
утвержденного учёным советом вуза от 26.06.2026 протокол № 11.

Рабочая программа одобрена на заседании ВШ Цифровых технологий

Протокол от 02.09.2026 г. № 1
Срок действия программы: 2026-2028 уч.г.
Руководитель подразделения: Хмелева Ирина Владимировна

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ
к рабочей программе дисциплины (модуля)

Системы анализа больших данных

_____ *наименование дисциплины*

№ п/п	Основание внесения изменения	Раздел/ пункт	Содержание изменения
1.			
2.			
3.			
4.			
5.			

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ	
1.1	Цель освоения дисциплины «Системы анализа больших данных» — формирование теоретических знаний и практических навыков проектирования, разработки, внедрения и исследования комплексных программных систем обработки и анализа больших объемов данных. Особое внимание уделяется ограничениям масштабирования, обеспечению безопасности и конфиденциальности, отказоустойчивости и аварийному восстановлению, использованию методов машинного обучения и экономическому обоснованию способов размещения вычислительной инфраструктуры. Задачи дисциплины: выявлять и классифицировать проблемы аналитических систем; обосновывать выбор технологий; проектировать безопасные каналы передачи и обработку данных; анализировать отказоустойчивость; реализовывать и испытывать компоненты анализа данных; оценивать производительность, точность и затраты.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП	
Цикл (раздел) ООП:	Б1.О
2.1	**Требования к предварительной подготовке обучающегося:**
2.1.1	Для изучения дисциплины требуются знания программирования, алгоритмов, баз данных, компьютерных сетей, архитектуры программных систем, основ вероятностно-статистических методов и машинного обучения на уровне подготовки бакалавра либо соответствующей подготовки при поступлении в магистратуру.
2.2	**Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:**
2.2.1	Результаты освоения используются при разработке распределённых программно-информационных систем, проектировании аналитических компонентов, испытаниях систем обработки больших данных, подготовке курсовой работы, практик и магистерской диссертации в области программной инженерии.

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)	
ПК-3: Способен применять современные информационные технологии, инструменты и среды на всех уровнях процесса создания программного обеспечения	
Знать:	
Уровень 1	Современные технологии разработки и тестирования ПО: средства обработки больших данных, библиотеки защиты и инструменты проверки компонентов.
Уровень 2	Принципы построения распределённых систем, защищённых интерфейсов, тестовых сценариев и мониторинга.
Уровень 3	Технологические ограничения и критерии выбора инструментов для комплексных систем анализа больших данных.
Уметь:	
Уровень 1	Выбирать библиотеки и среды программирования и создавать проверяемые демонстрационные компоненты.
Уровень 2	Проектировать и реализовывать защищённое взаимодействие, маршрутизацию и тестирование.
Уровень 3	Обосновывать архитектурные и инструментальные решения на основании результатов испытаний.
Владеть:	
Уровень 1	Навыками работы с современными средами разработки и тестовыми инструментами.
Уровень 2	Навыками прототипирования безопасных клиент-серверных компонентов и фиксации результатов испытаний.
Уровень 3	Навыками комплексной оценки технологического стека системы анализа больших данных.

ПК-6: Способен разработать программное обеспечение на основе встроенной аналитики больших данных	
Знать:	
Уровень 1	Подходы к построению систем обработки больших данных и их аналитических компонентов.
Уровень 2	Особенности обучения и эксплуатации моделей машинного обучения на масштабных данных.
Уровень 3	Факторы производительности, точности и стоимости аналитической инфраструктуры.
Уметь:	
Уровень 1	Определять требования к аналитическим компонентам и выбирать технологии реализации.
Уровень 2	Реализовывать простые модели машинного обучения, измерять производительность и качество.
Уровень 3	Сравнивать альтернативные архитектуры аналитики по измеримым критериям.
Владеть:	
Уровень 1	Навыками выбора средств разработки аналитических компонентов.
Уровень 2	Навыками проведения экспериментов с моделями, данными и ресурсами.
Уровень 3	Навыками документирования и обоснования решений для больших объёмов данных.

ПК-13: Способен совершенствовать и разрабатывать новые методы, модели и алгоритмы, в том числе для работы с большими данными	
Знать:	
Уровень 1	Методы сравнительного анализа технологий, проектирования планов испытаний и обработки результатов.
Уровень 2	Подходы к исследованию защищённости, отказоустойчивости и восстановления после отказов.
Уровень 3	Способы экспериментальной оценки ускорения обучения, точности моделей и экономических ограничений.
Уметь:	
Уровень 1	Формулировать план испытаний технологий и инструментов обработки больших данных.
Уровень 2	Проводить воспроизводимые сравнения и анализировать причины расхождений.
Уровень 3	Предлагать обоснованные улучшения моделей и алгоритмов по результатам эксперимента.
Владеть:	
Уровень 1	Навыками планирования экспериментов, протоколирования данных и представления выводов.
Уровень 2	Навыками тестирования отказоустойчивости, защищённости и программной производительности.
Уровень 3	Навыками разработки предложений по совершенствованию аналитической программной системы.

В результате освоения дисциплины обучающийся должен

3.1	**Знать:**
3.1.1	1. Знать ключевые проблемы обработки и анализа больших объемов данных, способы классификации и решения.
3.1.2	2. Знать средства обезличивания и обфускации, основы шифрования, HTTPS, односторонней и взаимной аутентификации TLS, сегментации сети и шлюзов доступа.
3.1.3	3. Знать принципы резервирования, резервного копирования, аварийного восстановления, способы выявления единой точки отказа.
3.1.4	4. Знать особенности технологий разработки, тестирования и мониторинга систем больших данных.
3.1.5	5. Знать ограничения обучения нейросетевых моделей на больших данных, метрики точности и подходы повышения производительности.
3.1.6	6. Знать варианты развёртывания on-premises, в облаке и у внешнего поставщика, критерии стоимости и риска.
3.2	**Уметь:**
3.2.1	1. Выбирать программные и инфраструктурные технологии с учётом объёма данных, производительности, защищённости и стоимости.
3.2.2	2. Разрабатывать демонстрационные программные компоненты обезличивания, шифрования, защищённого взаимодействия и маршрутизации.
3.2.3	3. Проектировать сценарии тестирования, мониторинга и восстановления программных систем после отказа.
3.2.4	4. Реализовывать простую модель машинного обучения, проводить измерения скорости обучения и оценивать качество прогноза.
3.2.5	5. Сопоставлять архитектурные решения и документировать результаты сравнительного анализа.
3.3	**Владеть:**
3.3.1	1. Навыками использования современных инструментов разработки и тестирования программных систем.
3.3.2	2. Навыками применения библиотек защиты данных и настройки защищённого клиент-серверного взаимодействия.
3.3.3	3. Навыками планирования и выполнения экспериментальных испытаний программных компонентов обработки больших данных.
3.3.4	4. Навыками оценки производительности, точности, отказоустойчивости и экономической эффективности решений.
3.3.5	5. Навыками подготовки технического отчёта и обоснования архитектурных и технологических решений.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетен-ции	Литература	Инте ракт.	Пр. подг.	Примечание
	Раздел 1. Обзор проблем, связанных с обработкой и анализом больших данных							
1.1	Самостоятельное изучение: классификация проблем обработки больших данных; производительность, распределённость, корректность данных. /Ср/	2	10	ПК-6 ПК-13	Л1.1 Л1.2			

1.2	Самостоятельное изучение: способы решения проблем обработки больших объёмов информации, подходы к анализу технических ограничений. /Ср/	2	10	ПК-6 ПК-13	Л1.1 Л1.2			
	Раздел 2. Способы повышения безопасности работы с данными							
2.1	Практическая работа №1. Реализация программного средства для обфускации и деобфускации учебных данных. /Пр/	2	5	ПК-3 ПК-6	Л1.1 Л1.2			Практическая подготовка; лабораторный сценарий
2.2	Самостоятельная работа: обфускация, обезличивание и защита информации; правовые ограничения использования данных. /Ср/	2	15	ПК-3 ПК-6	Л1.1 Л1.2			
2.3	Практическая работа №2. Демонстрационная программа и тесты шифрования данных стандартными библиотеками. /Пр/	2	5	ПК-3 ПК-13	Л1.1 Л1.2			Практическая подготовка
2.4	Самостоятельная работа: шифрование данных, библиотеки Bouncy Castle и другие средства, управление ключами и тестирование. /Ср/	2	15	ПК-3 ПК-6	Л1.1 Л1.2			
2.5	Практическая работа №3. Защищённое клиент-серверное приложение по HTTPS с взаимной аутентификацией TLS. /Пр/	2	5	ПК-3 ПК-13	Л1.1 Л1.2			Практическая подготовка
2.6	Самостоятельная работа: TLS, HTTPS, сертификаты, одно- и двусторонняя проверка подлинности. /Ср/	2	10	ПК-3 ПК-6	Л1.1 Л1.2			
2.7	Практическая работа №4. Программный шлюз Gateway для клиент-серверного приложения. /Пр/	2	5	ПК-3 ПК-6	Л1.1 Л1.2			Практическая подготовка
2.8	Самостоятельная работа: изоляция распределённой системы, шлюзы и балансировка запросов. /Ср/	2	15	ПК-3 ПК-6	Л1.1 Л1.2			
	Раздел 3. Аварийное восстановление (disaster recovery) программных систем после сбоев							
3.1	Самостоятельная работа: обзор аварийного восстановления, единая точка отказа, резервное копирование и резервирование. /Ср/	2	9	ПК-3 ПК-13	Л1.1 Л1.2			
3.2	Самостоятельная работа: сравнение подходов к резервированию и отказоустойчивости распределённых сервисов. /Ср/	2	10	ПК-3 ПК-13	Л1.1 Л1.2			
	Раздел 4. Проблемы применения современных методов машинного обучения в системах анализа данных							

4.1	Практическая работа №5. Реализация простой нейронной сети с обучением на данных. /Пр/	2	5	ПК-6 ПК-13	Л1.1 Л1.2Л2.1			Практическая подготовка
4.2	Самостоятельная работа: устройство нейронных сетей, обучение и глубокие модели, реализация в Java/Python. /Ср/	2	10	ПК-6 ПК-13	Л1.1 Л1.2Л2.1			
4.3	Практическая работа №6. Исследование производительности этапа обучения нейросетевой модели. /Пр/	2	5	ПК-6 ПК-13	Л1.1 Л1.2Л2.1			Практическая подготовка
4.4	Самостоятельная работа: факторы быстродействия, измерение ресурсов и варианты ускорения обучения моделей. /Ср/	2	9,9	ПК-6 ПК-13	Л1.1 Л1.2Л2.1			
4.5	Контактная работа в период теоретического обучения (семестр 2). /КрТО/	2	0,1					
4.6	Практическая работа №7. Исследование точности алгоритма машинного обучения и сравнение метрик. /Пр/	3	4	ПК-6 ПК-13	Л1.1 Л1.2Л2.1			
4.7	Самостоятельная работа: эксперименты с точностью и воспроизводимостью машинного обучения, анализ ошибок и переобучения. /Ср/	3	35	ПК-6 ПК-13	Л1.1 Л1.2Л2.1			
4.8	Самостоятельная работа: особенности масштабирования обучения и использования моделей на больших данных. /Ср/	3	30	ПК-6 ПК-13	Л1.1 Л1.2Л2.1			
4.9	Самостоятельная работа: сравнительная оценка технологий разработки, тестирования и мониторинга аналитических компонентов. /Ср/	3	30	ПК-6 ПК-13	Л1.1 Л1.2Л2.1			
	Раздел 5. Экономические трудности применения облачных и кластерных систем анализа							
5.1	Самостоятельная работа: экономический и функциональный анализ on-premises, облачного и сервисного размещения; анализ рисков данных. /Ср/	3	55	ПК-6 ПК-13	Л1.1 Л1.2Л2.1			
5.2	Курсовая работа: архитектурное проектирование и экспериментальная оценка защищённой системы анализа больших данных (26 ч включены в самостоятельную работу по плану КРСУ). /Ср/	3	26	ПК-3 ПК-6 ПК-13	Л1.1 Л1.2Л2.1			
5.3	/КрЭк/	3	0,3					

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
5.1. Контрольные вопросы и задания
ВОПРОСЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ЗАЧЁТА С ОЦЕНКОЙ

Раздел 1. Проблемы обработки больших данных
1. Характеристики больших данных и классификация проблем их обработки.
2. Узкие места масштабирования систем анализа данных.
3. Особенности многопоточной и распределённой обработки.
4. Причины некорректности данных и способы контроля качества.
5. Подходы к выбору архитектуры системы анализа.
Раздел 2. Безопасность работы с данными
6. Обезличивание, псевдонимизация и обфускация: сходства и различия.
7. Основные риски конфиденциальности больших данных.
8. Назначение симметричного и асимметричного шифрования.
9. Библиотеки и программные интерфейсы для реализации шифрования.
10. HTTPS и TLS: аутентификация сервера и защита канала.
11. Взаимная аутентификация (mutual TLS), назначение сертификатов.
12. Сегментация и изоляция вычислительной сети.
13. Назначение Gateway, балансировщиков и маршрутизаторов.
Раздел 3. Аварийное восстановление
14. Отказоустойчивость, доступность и аварийное восстановление.
15. Единая точка отказа (SPOF) и способы её устранения.
16. Резервное копирование и восстановление: стратегии и ограничения.
17. Резервирование вычислительных сервисов и данных.
18. Сравнение облачных решений для обеспечения непрерывности работы.
Раздел 4. Машинное обучение в системах анализа данных
19. Понятия нейросети, слоя, обучения и глубокой модели.
20. Ограничения вычислительных ресурсов при обучении на больших наборах данных.
21. Параллелизация и ускорение обучения моделей.
22. Методы измерения производительности обучения и инференса.
23. Переобучение, обобщающая способность и способы повышения точности.
24. Разделение выборки на обучающую, валидационную и тестовую.
25. Метрики качества классификации и регрессии.
Раздел 5. Экономика развёртывания систем анализа
26. Сравнение on-premises, облачной и сервисной моделей размещения.
27. Постоянные и переменные затраты вычислительной инфраструктуры.
28. Экономические и функциональные критерии выбора площадки размещения.
29. Риски утечки и потери данных и их влияние на выбор решения.
30. Как документировать обоснование архитектуры и технологического стека.
Темы для проверки практических результатов: демонстрация обфускации; тесты шифрования; проверка TLS; реализация Gateway; работа нейросетевой модели; графики производительности; расчёт и интерпретация метрик точности; сравнительная таблица затрат и рисков.

5.2. Темы курсовых работ (проектов)

КУРСОВАЯ РАБОТА (семестр 3) «Проектирование и сравнительный анализ защищённой и отказоустойчивой системы анализа больших данных».
Выбрать прикладной сценарий; разработать архитектуру обработки и хранения данных; обосновать способ защиты, резервирования, размещения и используемые технологии; создать прототип или демонстрационный контур; провести минимум два типа экспериментальных проверок (например, функциональные и производительные, либо отказоустойчивости); представить сравнительную оценку расходов и рисков, пояснительную записку и презентацию. Конкретное индивидуальное задание утверждает преподаватель.

5.3. Фонд оценочных средств

ПРАКТИЧЕСКИЕ РАБОТЫ

ПР №1 (семестр 2, 5 ч). Реализация программного средства обфускации и деобфускации данных. Выбрать учебный набор синтетических записей, реализовать преобразование и обратное преобразование там, где это предусмотрено; протестировать корректность, описать ограничения по конфиденциальности. Отчёт: код, тесты, выводы.

ПР №2 (семестр 2, 5 ч). Шифрование данных стандартными библиотеками (например, Bouncy Castle). Создать демонстрационное приложение шифрования и дешифрования, подготовить положительные и отрицательные тесты, описать работу с ключами и ограничения используемой схемы. Отчёт: код, сценарии и результаты.

ПР №3 (семестр 2, 5 ч). Защищённое клиент-серверное взаимодействие по HTTPS с mutual TLS. Создать учебные клиент и сервер, настроить сертификаты, проверить успешный и неуспешный обмен при изменении доверенных сертификатов. Отчёт: схема, конфигурация, протокол тестов.

ПР №4 (семестр 2, 5 ч). Маршрутизация запросов через Gateway (в исходной РПД — Netflix Zuul). Разработать демонстрационный шлюз перед двумя сервисами, проверить корректную маршрутизацию и обработку ошибок. Отчёт: архитектура, конфигурация, сценарии проверки.

ПР №5 (семестр 2, 5 ч). Реализация простой нейронной сети. Подготовить данные, создать модель, обучить её, вывести результаты на тестовой выборке. Отчёт: постановка задачи, код, параметры модели, метрики.

ПР №6 (семестр 2, 5 ч). Исследование производительности обучения нейронной сети. Провести серию воспроизводимых измерений с изменением размера обучающих данных/параметров и описать ресурсы; построить график времени обучения и сделать выводы.

ПР №7 (семестр 3, 4 ч). Анализ точности метода машинного обучения. Рассчитать метрики на отложенной выборке, сравнить хотя бы два варианта настройки, проанализировать характер ошибок. Отчёт: таблица метрик, сравнение и выводы.

КРИТЕРИИ ПРОВЕРКИ ПРАКТИЧЕСКИХ РЕЗУЛЬТАТОВ: работоспособность, корректность сценариев, воспроизводимость испытаний, безопасность выбранного подхода, аргументированность выводов и качество отчёта.

5.4. Перечень видов оценочных средств

Выполнение и защита практических работ; проверка самостоятельных заданий; тематический опрос; обсуждение технических решений; защита курсовой работы (3-й семестр); зачёт с оценкой (2-й и 3-й семестры).

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год
Л1.1	Вейцман К. и др.	Вейцман К. и др. Распределенные системы мини- и микро-ЭВМ	Москва: Финансы и статистика, 1983
Л1.2	Лазарев В. Г., Черняев В. Г.	Распределенные управляющие и вычислительные системы	Москва: Наука, 1987

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год
Л2.1	Никифоров И. В.	Курсовое проектирование по учебной дисциплине «Наука о данных и аналитика больших объемов информации»	Санкт-Петербург: Изд-во Политехн. ун-та, 2017

6.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

Э1	Параллельное программирование с использованием OpenMP и MPI (ресурс из исходной РПД СПбПУ)	https://openedu.ru/course/tgu/PROGR/
Э2	Сайт курса исходной РПД СПбПУ; доступ может быть ограничен	https://dl.spbstu.ru/user/index.php?id=1863

6.3. Перечень информационных и образовательных технологий

6.3.1 Компетентностно-ориентированные образовательные технологии

6.3.1.1	Практические занятия по выбранной предметной области, в том числе в малых командах; самостоятельный анализ технической документации; обсуждение сценариев и архитектур (case study); проектная работа; экспериментальное тестирование программных компонентов; документирование и представление результатов.

6.3.2 Перечень информационных справочных систем и программного обеспечения

6.3.2.1	Операционные системы Windows или Linux; офисные программы для текстов, таблиц и презентаций; среды разработки для Java и C#/.NET; Python с библиотеками машинного обучения; средства настройки и тестирования TLS/HTTPS; библиотеки шифрования (в том числе Bouncy Castle); средства контроля версий и профилирования; доступ в Интернет. Конкретный состав устанавливается преподавателем с учётом доступной инфраструктуры.

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1	Для занятий необходим компьютерный класс или собственные компьютеры обучающихся с доступом в Интернет, средства разработки и тестирования, мультимедийный проектор, демонстрационные серверные компоненты. Для практических исследований допустимы виртуальные машины, контейнеры или учебные вычислительные ресурсы. Необходим доступ к вычислительному кластеру ВШПИ СПбПУ.

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Практические задания выполняются индивидуально или в малой команде. Отчёт содержит постановку задачи, описание выбранных технологий, схему/алгоритм, результаты работы программы, протоколы испытаний, оценку ограничений и выводы. При работе с персональными данными используют учебные синтетические наборы. Рекомендуется заранее изучить документацию библиотек и зафиксировать версии программ. Самостоятельная работа включает изучение разделов,

подготовку к практическим работам, исследование и анализ измерений, курсовую работу в 3-м семестре. Порядок выставления баллов и условия зачётов с оценкой определяются утверждённым ФОС и технологической картой.