

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Ястребов Олег Александрович  
Должность: Ректор  
Дата подписания: 22.05.2026 11:38:35  
Уникальный программный ключ:  
ca953a0120d891083f939673078ef1a989dae18a

**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**

### **МЕТОДЫ АНАЛИЗА И ХРАНЕНИЯ БОЛЬШИХ ДАННЫХ**

(наименование дисциплины/модуля)

**Рекомендована МССН для направления подготовки/специальности:**

### **02.04.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ**

(код и наименование направления подготовки/специальности)

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

### **УПРАВЛЕНИЕ ДАННЫМИ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ**

(наименование (профиль/специализация) ОП ВО)

**2026 г.**

## 1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Методы анализа и хранения больших данных» входит в программу магистратуры «Управление данными и искусственный интеллект» по направлению 02.04.02 «Фундаментальная информатика и информационные технологии» и изучается во 2 семестре 1 курса. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 3 разделов и 26 тем и направлена на изучение формирования у студентов системных знаний и практических навыков работы с масштабируемыми инфраструктурами и методами хранения и анализа данных в условиях роста их объема, разнообразия, скорости поступления и требований к надежности и безопасности. Курс развивает компетентность в архитектуре современных систем Big Data, интеграции с распределенными и облачными решениями, автоматизации работы с данными, анализе потоковых, нереляционных и смешанных данных, а также внедрении эффективных методологий хранения, обработки и аналитики в промышленных и исследовательских задачах искусственного интеллекта. Целью освоения дисциплины является научить студентов проектировать и реализовывать решения для хранения, обработки и анализа больших и сложных потоков данных, понимать производственные и исследовательские стандарты современных распределенных систем, грамотно выбирать технологию под задачу (Hadoop, Spark, Kafka, облачные хранилища и др.), обеспечивать масштабируемость, надежность, защиту и эффективность в рамках комплексных AI/Data Science-проектов.

## 2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Методы анализа и хранения больших данных» 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)
ПК-1	Способен разрабатывать и применять алгоритмы интеллектуальной обработки данных для решения задач профессиональной деятельности	ПК-1.1 Знает существующие системы хранения и анализа данных, алгоритмы интеллектуальной обработки данных; ПК-1.2 Умеет модифицировать алгоритмы интеллектуальной обработки данных; ПК-1.3 Имеет навыки использования и применения существующих и модифицированных систем хранения и анализа данных, алгоритмов интеллектуальной обработки данных для решения задач профессиональной деятельности;

## 3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Методы анализа и хранения больших данных» 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 «Методы анализа и хранения больших данных».

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

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
ПК-1	Способен разрабатывать и применять алгоритмы интеллектуальной обработки данных для решения задач профессиональной деятельности	Обработка мультимодальных данных**; Современные устройства центров обработки больших данных**;	Преддипломная практика; Искусственный интеллект и интернет вещей**; Генеративный искусственный интеллект; Большие языковые модели (на основании трансформеров); Методы оптимизации;

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

\*\* - элективные дисциплины /практики

#### 4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Методы анализа и хранения больших данных» 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)
			2
Контактная работа, ак.ч	51		51
Лекции (ЛК)	17		17
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	34		34
Самостоятельная работа обучающихся, ак.ч.	66		66
Контроль (экзамен/зачет с оценкой), ак.ч.	27		27
Общая трудоемкость дисциплины ак.ч.	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	Наименование темы		Содержание темы	Вид учебной работы*
Раздел 1	Архитектуры и инфраструктура анализа и хранения больших данных	1.1	Архитектурные основы Big Data: концепция, требования, типы данных, сценарии применения	Рассмотрение фундаментальных принципов построения систем обработки больших данных, основные характеристики Big Data (Volume, Velocity, Variety), требования к масштабируемости и отказоустойчивости, классификация типов данных (структурированные, полуструктурированные, неструктурированные), примеры применения в различных отраслях.	ЛК
		1.2	Выбор архитектур: локальные, распределённые, облачные, гибридные решения хранения и обработки	Сравнительный анализ архитектурных подходов, on-premise решения versus облачные платформы, гибридные архитектуры, критерии выбора оптимальной архитектуры в зависимости от требований бизнеса, масштабируемость и стоимость владения.	ЛК
		1.3	Хранилища данных: Data Lake, Data Warehouse, Data Mesh, современные форматы хранения (Parquet, ORC и др.)	Различия между Data Lake и Data Warehouse, концепция Data Mesh как децентрализованного подхода, форматы хранения данных (Parquet, ORC, Avro), их преимущества и недостатки, организация метаданных и каталогизация данных.	ЛК
		1.4	Развертывание и настройка мини-кластера Hadoop/Spark или облачного хранилища	Практическая работа по установке и конфигурации компонентов экосистемы Hadoop (HDFS, YARN, MapReduce) или Apache Spark, настройка облачных сервисов (AWS S3, Azure Data Lake), базовая конфигурация и проверка работоспособности.	
		1.5	Оценка и тестирование современных форматов хранения данных: сравнение по скорости и объёму	Бенчмаркинг различных форматов файлов, тестирование скорости чтения/записи, степень компрессии, производительность запросов, рекомендации по выбору формата для конкретных задач.	
		1.6	Обзор современных платформ и сервисов управления большими данными (выбор под задачу)	Анализ коммерческих и open-source решений: Cloudera, Hortonworks, AWS EMR, Azure HDInsight, Google Cloud Dataproc, Databricks, критерии выбора платформы, сравнение функциональности и стоимости.	СЗ
		1.7	Анализ ландшафта промышленных проектов Big Data: сравнение архитектур и инфраструктур	Изучение реальных кейсов внедрения Big Data в различных компаниях, архитектурные паттерны, best practices, типичные ошибки и способы их избегания, эволюция архитектур.	СЗ

		1.8	Практический разбор кейсов проектирования инфраструктуры — от стартапа до корпорации	Разработка архитектуры под различные сценарии: MVP для стартапа, масштабирование для среднего бизнеса, корпоративная инфраструктура, миграция legacy-систем, оценка ресурсов и бюджетирование.	СЗ
Раздел 2	Технологии хранения и	2.1	Распределённые файловые	Принципы работы распределённых файловых систем, архитектура HDFS, NoSQL базы	ЛК

Номер раздела	Наименование раздела дисциплины	Наименование темы	Содержание темы	Вид учебной работы*
	обработки больших данных: распределённые и потоковые методы		системы и базы данных: HDFS, Cassandra, ClickHouse данных (Cassandra - wide-column store), колоночные СУБД (ClickHouse), CAP-теорема, модели консистентности.	
2.2		Потоковые платформы и обработка событий: Apache Kafka, Spark Streaming, Flink	Концепция потоковой обработки данных, архитектура Apache Kafka, топики и партиции, Spark Streaming и микробатчинг, Apache Flink для stream processing в реальном времени, сравнение подходов.	ЛК
2.3		Методы обеспечения согласованности, доступности, надёжности и отказоустойчивости в Big Data	CAP-теорема на практике, механизмы репликации, стратегии обеспечения consistency, availability и partition tolerance, fault tolerance механизмы, disaster recovery планирование.	ЛК
2.4		Настройка и выполнение запросов в распределённой базе данных (например, ClickHouse или Cassandra)	Практические навыки создания схем данных, написания запросов с учётом распределённой природы БД, оптимизация производительности, партиционирование и шардинг, мониторинг.	
2.5		Организация обработки потоковых данных: создание пайплайна Kafka + Spark Streaming	Построение end-to-end решения для обработки потоковых данных, настройка Kafka producers и consumers, разработка Spark Streaming приложений, обработка окон данных, интеграция с хранилищами.	
2.6		Лабораторное задание по реализации процедур резервного копирования и восстановления в продвинутых хранилищах	Практика настройки backup и restore процедур, инкрементальные копии, точки восстановления, тестирование disaster recovery сценариев, автоматизация резервного копирования.	
2.7		Разбор CAP-теоремы, типов согласованности и требований к масштабированию	Детальный анализ теоретических основ распределённых систем, eventual consistency, strong consistency, типы изоляции транзакций, практические компромиссы при проектировании систем.	СЗ
2.8		Обсуждение сценариев применения потоковых и распределённых хранилищ (индустриальные кейсы)	Реальные примеры использования: IoT-платформы, финансовые системы, телекоммуникации, e-commerce, выбор технологий под конкретные требования бизнеса.	СЗ
2.9		Практика мониторинга, поддержки и резервирования в распределённых инфраструктурах	Инструменты мониторинга (Prometheus, Grafana), метрики производительности, алертинг, логирование в распределённых системах, профилактическое обслуживание, capacity planning.	СЗ

Раздел 3	Интеллектуальный анализ данных и интеграция с AI/Data Science	3.1	Методы интеграции Big Data с аналитическими, ML и AI-системами: вызовы оркестрации, пайплайны	Архитектура ML-пайплайнов на больших данных, интеграция с аналитическими платформами, оркестрация workflows (Apache Airflow), управление экспериментами, MLOps практики.	ЛК
		3.2	Масштабируемая	Распределённое машинное обучение с использованием Spark MLlib, Dask для	ЛК

Номер раздела	Наименование раздела дисциплины	Наименование темы	Содержание темы	Вид учебной работы*
		аналитика: совместное использование Spark ML, Dask, параллельных библиотек	параллельных вычислений в Python, Ray, rapids.ai для GPU-ускорения, выбор инструментов под задачу.	
		3.3 Безопасность, приватность и этические аспекты анализа и хранения больших данных в проектах искусственного интеллекта	GDPR и регуляторные требования, дифференциальная приватность, анонимизация данных, этические вопросы использования ИИ, audit trails, управление правами доступа.	ЛК
		3.4 Интеграция распределённого хранения с ML-аналитикой (Spark ML, Dask, BigQuery ML)	Практическая интеграция ML-моделей с распределёнными хранилищами, обучение моделей на больших датасетах, feature engineering в распределённой среде, сервировка моделей.	
		3.5 Реализация сквозного аналитического пайплайна: ingestion, storage, обработка, ML-вывод	Построение полного цикла обработки данных от сбора до получения инсайтов, ETL/ELT процессы, feature store, обучение и деплой моделей, A/B тестирование, мониторинг моделей.	
		3.6 Эксперимент по оценке и обеспечению безопасности/приватности в инфраструктуре Big Data	Практические задачи по внедрению мер безопасности, шифрование данных в покое и в движении, аутентификация и авторизация, аудит доступа, тестирование на уязвимости, compliance проверки.	
		3.7 Разбор практик построения сквозных пайплайнов AI/Big Data и управления жизненным циклом данных	Изучение архитектуры end-to-end data pipelines от источников до ML-моделей, включая ingestion, processing, storage, и serving layers. Рассмотрение концепции Data Lifecycle Management (DLM) с этапами создания, хранения, обработки, использования, архивации и удаления данных. Анализ архитектурных паттернов (Medallion Architecture, Lambda/Kappa), практик версионирования данных и моделей (Delta Lake, MLflow), оркестрации workflows (Airflow, Prefect), управления метаданными и lineage tracking. Практическая реализация сквозного пайплайна с применением современных инструментов governance, включая data catalogs, quality gates и compliance автоматизацию.	СЗ
		3.8 Критический анализ реальных проектов Data Science на больших данных: bottlenecks, риски, удачные решения	Детальный разбор типичных узких мест в DS-проектах: инфраструктурные bottlenecks (data access latency, I/O constraints, memory limitations), организационные риски (expectation mismatch, siloed teams, budget overruns), технические вызовы (data drift, feature engineering at scale). Изучение реальных кейсов из e-commerce, финтех, производства с анализом проблем и их решений. Рассмотрение удачных архитектурных паттернов (Feature Platform, модульные архитектуры моделей), best practices по мониторингу и масштабированию. Практический анализ trade-offs между сложностью решения и бизнес-ценностью, критерии выбора "build vs buy", стратегии минимизации технического долга.	СЗ

		3.9	Дискуссия по вопросам приватности, этики, compliance и	Глубокое погружение в вопросы защиты приватности: техники анонимизации (k-anonymity, differential privacy), privacy-preserving ML (federated learning, homomorphic encryption), риски re-identification и inference attacks. Изучение	С3
--	--	-----	--	---	----

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
			ответственности при эксплуатации Big Data в AI	регуляторных frameworks (GDPR, CCPA, HIPAA) и их влияния на AI/ML системы, требования к explainability и automated decision-making. Анализ этических проблем: sources of bias (data, algorithmic, feedback loops), fairness metrics и их ограничения, методы mitigation на разных стадиях pipeline. Рассмотрение вопросов accountability и ответственности за ошибки AI, создание AI Ethics Board, процессы Ethical Impact Assessment. Практическое внедрение responsible AI practices: model cards, continuous fairness monitoring, diverse teams, stakeholder engagement.	

\* - 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.	
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве [Параметр] шт.), 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. Параскевов, А. В. Большие данные: учебное пособие / А. В. Параскевов, А. Э. Сергеев. - Москва; Вологда: Инфра-Инженерия, 2024. - 148 с. - ISBN 978-5-9729-2120-1. - Текст: электронный. - URL: <https://znanium.ru/catalog/product/2169699>

2. Григорьев, А. А. Методы и алгоритмы обработки данных: учебное пособие / А.А. Григорьев, Е.А. Исаев. — 2-е изд., перераб. и доп. — Москва: ИНФРА-М, 2024. — 383 с. + Доп. материалы [Электронный ресурс]. — (Высшее образование: Бакалавриат). — DOI 10.12737/1032305. - ISBN 978-5-16-015581-4. - Текст: электронный. - URL: <https://znanium.ru/catalog/product/2084190>

3. Анализ больших данных: учебное пособие / И. Б. Тесленко, В. Е. Крылов, А. М. Губернаторов [и др.]. — Москва: КноРус, 2025. — 295 с. — ISBN 978-5-406-14006-2. — URL: <https://book.ru/book/955989>

### Дополнительная литература:

1. Дадян, Э. Г. Методы, модели, средства хранения и обработки данных: учебник / Э.Г. Дадян, Ю.А. Зеленков. — Москва: Вузовский учебник: ИНФРА-М, 2024. — 168 с. - ISBN 978-5-9558-0490-3. - Текст: электронный. - URL: <https://znanium.com/catalog/product/2122966>

2. Методы и технологии обработки больших данных [Электронный ресурс]: учебно-методическое пособие / М.М. Железнов. — Москва: Издательство МИСИ – МГСУ, 2020. ISBN 978-5-7264-2193-3  
*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](http://www.studentlibrary.ru)
- ЭБС «Знаниум» <https://znanium.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 «Методы анализа и хранения больших данных».

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

## РАЗРАБОТЧИКИ

---

Должность

**РУКОВОДИТЕЛЬ БУП**

Заведующий кафедрой

---

Должность

**РУКОВОДИТЕЛЬ ОП ВО**

Заведующий кафедрой прикладного  
искусственного интеллекта

---

Должность

---

Подолько П.М.

Фамилия И.О

---

---

Подолько П.М.

Фамилия И.О

---

---

Подолько П.М.

Фамилия И.О

---