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
_____ Г.И.Лоцев
3 сентября 2026 г.

Наука о данных и аналитика больших объемов информации

рабочая программа дисциплины (модуля)

Закреплена за кафедрой **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 **4 ЗЕТ**

Часов по учебному плану 144

в том числе:

аудиторные занятия 44

самостоятельная 97,8

Виды контроля в семестрах:

курсовая работа 1
экзамен 1

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	1 (1.1)		Итого	
Недель	18			
Вид занятий	УП	РП	УП	РП
Лекции	30	30	30	30
Практические	14	14	14	14
Контактная работа в период теоретического	2,2	2,2	2,2	2,2
В том числе инт.	8	8	8	8
В том числе в форме практ.подготовки	14	14	14	14
Итого ауд.	44	44	44	44
Контактная работа	46,2	46,2	46,2	46,2
Сам. работа	97,8	97,8	97,8	97,8
Итого	144	144	144	144

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

Наука о данных и аналитика больших объемов информации

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)

Наука о данных и аналитика больших объемов информации

_____ *наименование дисциплины*

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

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ	
1.1	Цель дисциплины — формирование знаний о получении, преобразовании, хранении, передаче, обработке и анализе больших объёмов данных и практических умений по разработке и внедрению соответствующих программных средств. Задачи: освоить характеристики больших данных и распределённую модель вычислений; изучить архитектуры обработки, MapReduce, Apache Hadoop и Spark; приобрести знания о распределённом хранении и NoSQL; освоить процессы ETL, форматы и модели данных, методы извлечения знаний, визуализации, бизнес-аналитики; выполнить практические работы по Elasticsearch и Kibana, курсовую работу и обосновать результаты аналитического проекта. Содержание основано на РПД СПбПУ 2025 года.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП	
Цикл (раздел) ООП:	Б1.В
2.1	**Требования к предварительной подготовке обучающегося:**
2.1.1	Для освоения необходимы знания бакалавриата или специалитета: основы программирования, алгоритмы и структуры данных, базы данных, компьютерные сети, веб-технологии и основы анализа данных.
2.2	**Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:**
2.2.1	Освоение дисциплины обеспечивает подготовку к изучению проектирования распределённых систем и программных сервисов обработки данных, методов машинного обучения, к выполнению курсовых проектов, практик и магистерской диссертации по программной инженерии.

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)	
ПК-1: Способен руководить работами по выявлению, формулированию, документированию и согласованию требований к программному обеспечению с заказчиком	
Знать:	
Уровень 1	Методы выявления требований к аналитическим системам и критерии реализуемости
Уровень 2	Методы описания потоков данных и процессов средствами диаграмм моделирования
Уровень 3	Подходы к согласованию архитектуры и функциональных требований аналитической системы
Уметь:	
Уровень 1	Выявлять ключевые требования к обработке и визуализации данных
Уровень 2	Разрабатывать UML-диаграммы взаимодействия компонентов и потоков данных
Уровень 3	Обосновывать и документировать решения по выбору технологий
Владеть:	
Уровень 1	Приёмами сбора и анализа требований к информационным системам
Уровень 2	Навыками построения диаграмм архитектуры и бизнес-процессов
Уровень 3	Навыками представления и защиты технических решений

ПК-2: Способен применять современные аналитические методы при разработке инструментальных средств поддержки принятия решений	
Знать:	
Уровень 1	Основные методы аналитики больших данных и их применение в решениях для бизнеса
Уровень 2	Методы машинного обучения и визуального анализа данных
Уровень 3	Критерии выбора аналитических методов для поддержки принятия решений
Уметь:	
Уровень 1	Выбирать методы анализа больших данных для прикладных задач
Уровень 2	Применять инструменты извлечения данных и визуализации результатов
Уровень 3	Сопоставлять аналитические методы по качеству и целям применения
Владеть:	
Уровень 1	Практикой применения средств анализа больших данных
Уровень 2	Навыками работы с поисковыми индексами, Elasticsearch и Kibana
Уровень 3	Навыками оценки эффективности аналитических решений

ПК-5: Способен организовывать и реализовывать мероприятия по обеспечению качества программного обеспечения на протяжении жизненного цикла проекта	
Знать:	
Уровень 1	Назначение статического и динамического анализа кода

Уровень 2	Подходы к тестированию клиентских и аналитических компонентов
Уровень 3	Критерии качества и воспроизводимости аналитического программного решения
Уметь:	
Уровень 1	Разрабатывать проверки корректности обработки и преобразования данных
Уровень 2	Проводить тестирование REST API и анализировать ошибки
Уровень 3	Фиксировать результаты испытаний, обнаруживать и устранять дефекты
Владеть:	
Уровень 1	Навыками использования средств статического анализа кода
Уровень 2	Навыками проверки интерфейсов и качества программных компонентов
Уровень 3	Навыками подготовки отчётов по тестированию и контролю качества

ПК-10: Способен анализировать потребности заинтересованных лиц и подразделений организации и подходы к исследованию больших данных	
Знать:	
Уровень 1	Характеристики, источники и жизненный цикл больших данных
Уровень 2	Современные методы сбора, обработки и визуализации больших данных
Уровень 3	Подходы к выбору инструментов анализа больших данных под потребности организации
Уметь:	
Уровень 1	Выявлять задачи и процессы организации для применения аналитики
Уровень 2	Анализировать возможности распределённой обработки, ETL и NoSQL
Уровень 3	Составлять реестр задач и обосновывать выбор инструментальных средств
Владеть:	
Уровень 1	Навыками исследования предметной области и источников данных
Уровень 2	Практикой работы с инфраструктурой хранения и визуализации данных
Уровень 3	Навыками формулирования выводов и рекомендаций по использованию аналитики

В результате освоения дисциплины обучающийся должен

3.1	**Знать:**
3.1.1	Знать: характеристики и источники больших данных, принципы горизонтального масштабирования, распределённой обработки и систем хранения; архитектуры Lambda/Kappa; подходы MapReduce, Apache Hadoop/YARN/HDFS и Spark; классификацию NoSQL; этапы ETL и интерфейсы межсервисного обмена; форматы CSV, XML, JSON, Avro, Parquet и модели OLTP/OLAP; этапы жизненного цикла аналитического проекта; методы извлечения знаний, полнотекстового поиска и визуализации; инструменты Elasticsearch/Kibana/Grafana и основные функции BI-платформ. Знать методы статического и динамического анализа ПО, основы оценки реализуемости требований и выбора методов машинного обучения.
3.2	**Уметь:**
3.2.1	Уметь: оценивать применимость технологий к конкретной задаче и требования к ПО; определять задачи для аналитики больших данных; разрабатывать схемы обработки и модели данных; настраивать учебный ELK-стек, использовать REST API Elasticsearch, создавать индексы, загружать и визуализировать данные в Kibana; выполнять аналитические исследования и оценивать результаты; готовить презентацию, выбирать технологии машинного обучения для прикладных задач; использовать статическую и динамическую проверку качества кода.
3.3	**Владеть:**
3.3.1	Владеть: навыками разработки UML-диаграмм и обоснования реализуемости требований; современными средствами хранения, извлечения, поиска и визуализации больших данных; практической работой с Elasticsearch и Kibana, настройкой визуализаций и дашбордов; инструментами статического и динамического анализа ПО; методами оценки эффективности аналитических решений и машинного обучения; навыками подготовки технического отчёта и защиты курсового проекта.

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. Введение в науку о данных и большие данные							
1.1	Цель курса и требования к слушателям /Лек/ /Лек/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			
1.2	Самостоятельная работа: Цель курса и требования к слушателям /Ср/ /Ср/	1	3	ПК-1 ПК-10	Л1.1 Л1.2			

1.3	Смежные области знаний /Лек/ /Лек/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			
1.4	Самостоятельная работа: Смежные области знаний /Ср/ /Ср/	1	3	ПК-1 ПК-10	Л1.1 Л1.2			
1.5	Предпосылки появления больших данных /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
1.6	Самостоятельная работа: Предпосылки появления больших данных /Ср/ /Ср/	1	3	ПК-2 ПК-10	Л1.1 Л1.2			
1.7	Большие данные (3V, структура и примеры) /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
1.8	Большие данные (3V, структура и примеры) /Пр/ /Пр/	1	1	ПК-2 ПК-10	Л1.1 Л1.2		1	Практическая подготовка; формы работ уточняются преподавателем
1.9	Самостоятельная работа: Большие данные (3V, структура и примеры) /Ср/ /Ср/	1	3	ПК-2 ПК-10	Л1.1 Л1.2			
1.10	Распределённая модель вычислений /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
1.11	Самостоятельная работа: Распределённая модель вычислений /Ср/ /Ср/	1	3	ПК-2 ПК-10	Л1.1 Л1.2			
	Раздел 2. Распределённая обработка и хранение больших данных							
2.1	Типы обработки информации /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
2.2	Типы обработки информации /Пр/ /Пр/	1	1	ПК-2 ПК-10	Л1.1 Л1.2	1	1	Практическая подготовка; формы работ уточняются преподавателем
2.3	Самостоятельная работа: Типы обработки информации /Ср/ /Ср/	1	3	ПК-2 ПК-10	Л1.1 Л1.2			
2.4	Архитектуры программных средств обработки информации: Lambda и Kappa /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
2.5	Самостоятельная работа: Архитектуры программных средств обработки информации: Lambda и Kappa /Ср/ /Ср/	1	3	ПК-2 ПК-10	Л1.1 Л1.2			
2.6	Алгоритм обработки MapReduce /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
2.7	Алгоритм обработки MapReduce /Пр/ /Пр/	1	1	ПК-2 ПК-10	Л1.1 Л1.2	1	1	Практическая подготовка; формы работ уточняются преподавателем
2.8	Самостоятельная работа: Алгоритм обработки MapReduce /Ср/ /Ср/	1	3	ПК-2 ПК-10	Л1.1 Л1.2			
2.9	Экосистема Apache Hadoop: HDFS и YARN /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			

2.10	Экосистема Apache Hadoop: HDFS и YARN /Пр/ /Пр/	1	1	ПК-2 ПК-10	Л1.1 Л1.2	1	1	Практическая подготовка; формы работ уточняются преподавателем
2.11	Самостоятельная работа: Экосистема Apache Hadoop: HDFS и YARN /Ср/ /Ср/	1	3	ПК-2 ПК-10	Л1.1 Л1.2			
2.12	Инструмент Apache Spark /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
2.13	Инструмент Apache Spark /Пр/ /Пр/	1	1	ПК-2 ПК-10	Л1.1 Л1.2		1	Практическая подготовка; формы работ уточняются преподавателем
2.14	Самостоятельная работа: Инструмент Apache Spark /Ср/ /Ср/	1	3	ПК-2 ПК-10	Л1.1 Л1.2			
2.15	Предпосылки появления NoSQL баз данных /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
2.16	Самостоятельная работа: Предпосылки появления NoSQL баз данных /Ср/ /Ср/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
2.17	NoSQL базы данных и их типы /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
2.18	NoSQL базы данных и их типы /Пр/ /Пр/	1	1	ПК-2 ПК-10	Л1.1 Л1.2		1	Практическая подготовка; формы работ уточняются преподавателем
2.19	Самостоятельная работа: NoSQL базы данных и их типы /Ср/ /Ср/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
2.20	Дополнительное самостоятельное изучение: документация Hadoop, Spark и распределённые вычисления /Ср/ /Ср/	1	8	ПК-2 ПК-10	Л1.1 Л1.2			
	Раздел 3. Сбор, передача, форматы данных и жизненный цикл аналитического проекта							
3.1	ETL-процесс и инструменты поддержки ETL /Лек/ /Лек/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			
3.2	Самостоятельная работа: ETL-процесс и инструменты поддержки ETL /Ср/ /Ср/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			
3.3	Извлечение данных из Web-ресурсов /Лек/ /Лек/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			
3.4	Самостоятельная работа: Извлечение данных из Web-ресурсов /Ср/ /Ср/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			
3.5	Обмен данными между сервисами: Kafka, RabbitMQ /Лек/ /Лек/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			
3.6	Обмен данными между сервисами: Kafka, RabbitMQ /Пр/ /Пр/	1	1	ПК-1 ПК-10	Л1.1 Л1.2		1	Практическая подготовка; формы работ уточняются преподавателем
3.7	Самостоятельная работа: Обмен данными между сервисами: Kafka, RabbitMQ /Ср/ /Ср/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			

3.8	Основные форматы данных и их отличительные особенности /Лек/ /Лек/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			
3.9	Самостоятельная работа: Основные форматы данных и их отличительные особенности /Ср/ /Ср/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			
3.10	Моделирование данных, OLTP и OLAP /Лек/ /Лек/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			
3.11	Моделирование данных, OLTP и OLAP /Пр/ /Пр/	1	1	ПК-1 ПК-10	Л1.1 Л1.2		1	Практическая подготовка; формы работ уточняются преподавателем
3.12	Самостоятельная работа: Моделирование данных, OLTP и OLAP /Ср/ /Ср/	1	1	ПК-1 ПК-10	Л1.1 Л1.2			
3.13	Основные этапы жизненного цикла /Лек/ /Лек/	1	1	ПК-1 ПК-2	Л1.1 Л1.2			
3.14	Самостоятельная работа: Основные этапы жизненного цикла /Ср/ /Ср/	1	1	ПК-1 ПК-2	Л1.1 Л1.2			
3.15	Отличия Business Intelligence от Data Science /Лек/ /Лек/	1	1	ПК-1 ПК-2	Л1.1 Л1.2			
3.16	Самостоятельная работа: Отличия Business Intelligence от Data Science /Ср/ /Ср/	1	1	ПК-1 ПК-2	Л1.1 Л1.2			
3.17	Роли в проекте по аналитике данных /Лек/ /Лек/	1	1	ПК-1 ПК-2	Л1.1 Л1.2			
3.18	Самостоятельная работа: Роли в проекте по аналитике данных /Ср/ /Ср/	1	1	ПК-1 ПК-2	Л1.1 Л1.2			
3.19	Дополнительное самостоятельное изучение: ETL, потоковая интеграция и REST API /Ср/ /Ср/	1	7,8	ПК-1 ПК-10	Л1.1 Л1.2			
	Раздел 4. Извлечение знаний, визуализация и бизнес-аналитика. Защита курсового проекта							
4.1	Типы и подходы извлечения знаний из данных /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
4.2	Самостоятельная работа: Типы и подходы извлечения знаний из данных /Ср/ /Ср/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
4.3	Поисковые системы: Lucene, Solr, Elasticsearch /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
4.4	Поисковые системы: Lucene, Solr, Elasticsearch /Пр/ /Пр/	1	1	ПК-2 ПК-10	Л1.1 Л1.2	1	1	Практическая подготовка; формы работ уточняются преподавателем
4.5	Самостоятельная работа: Поисковые системы: Lucene, Solr, Elasticsearch /Ср/ /Ср/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
4.6	Необходимость визуализации /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
4.7	Самостоятельная работа: Необходимость визуализации /Ср/ /Ср/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
4.8	Визуализация больших данных /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			

4.9	Визуализация больших данных /Пр/ /Пр/	1	1	ПК-2 ПК-10	Л1.1 Л1.2		1	Практическая подготовка; формы работ уточняются преподавателем
4.10	Самостоятельная работа: Визуализация больших данных /Ср/ /Ср/	1	2	ПК-2 ПК-10	Л1.1 Л1.2			
4.11	Инструментальные средства визуализации /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
4.12	Самостоятельная работа: Инструментальные средства визуализации /Ср/ /Ср/	1	2	ПК-2 ПК-10	Л1.1 Л1.2			
4.13	Основные функции BI-платформ /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
4.14	Основные функции BI-платформ /Пр/ /Пр/	1	1	ПК-2 ПК-10	Л1.1 Л1.2	1	1	Практическая подготовка; формы работ уточняются преподавателем
4.15	Самостоятельная работа: Основные функции BI-платформ /Ср/ /Ср/	1	2	ПК-2 ПК-10	Л1.1 Л1.2			
4.16	Обзор наиболее популярных платформ: Power BI, Tableau, Qlik Sense /Лек/ /Лек/	1	1	ПК-2 ПК-10	Л1.1 Л1.2			
4.17	Самостоятельная работа: Обзор наиболее популярных платформ: Power BI, Tableau, Qlik Sense /Ср/ /Ср/	1	2	ПК-2 ПК-10	Л1.1 Л1.2			
4.18	Защита курсовых проектов /Лек/ /Лек/	1	3	ПК-1 ПК-2 ПК-5 ПК-10	Л1.1 Л1.2Л2.1			
4.19	Защита курсовых проектов /Пр/ /Пр/	1	3	ПК-1 ПК-2 ПК-5 ПК-10	Л1.1 Л1.2Л2.1	3	3	
4.20	Самостоятельная работа: Защита курсовых проектов /Ср/ /Ср/	1	6	ПК-1 ПК-2 ПК-5 ПК-10	Л1.1 Л1.2Л2.1			
4.21	Дополнительное самостоятельное исследование: Elasticsearch, Kibana и подготовка практических отчётов /Ср/ /Ср/	1	10	ПК-2 ПК-10	Л1.1 Л1.2			
4.22	Самостоятельная разработка и оформление курсового проекта в соответствии с учебным планом КРСУ (дополнительное время) /Ср/ /Ср/	1	15	ПК-1 ПК-2 ПК-5 ПК-10	Л1.1 Л1.2Л2.1			
4.23	Контактная работа в период теоретического обучения (по учебному плану КРСУ). /КрТО/	1	2,2					

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

5.1. Контрольные вопросы и задания

КОНТРОЛЬНЫЕ ВОПРОСЫ ДЛЯ ОПРОСОВ, КОНТРОЛЬНЫХ РАБОТ И ЭКЗАМЕНА

Раздел 1. Введение в науку о данных и большие данные

1. Предмет, цель и задачи науки о данных.
2. Связь Data Science с информационными системами и программной инженерией.
3. Предпосылки появления больших данных и ограниченность вертикального масштабирования.
4. Основные свойства больших данных (3V) и разновидности структуры данных.
5. Горизонтальное масштабирование и уровни распределённых вычислений.

Раздел 2. Распределённая обработка и хранение больших данных
1. Пакетная, потоковая, SQL-подобная и интерактивная обработка.
2. Архитектуры Lambda и Kappa: различия и применимость.
3. Модель MapReduce: Map, Shuffle, Reduce.
4. HDFS, YARN и отказоустойчивость Apache Hadoop.
5. RDD и компоненты Apache Spark.
6. Причины перехода от реляционных баз данных к NoSQL.
7. Шардинг, репликация, согласованность и теорема CAP.
8. Key/Value, документоориентированные, графовые и ширококолоночные БД.
Раздел 3. Сбор, передача, форматы данных и жизненный цикл аналитического проекта
1. Этапы ETL, назначение Apache NiFi и Airflow.
2. Подходы к извлечению данных из Web.
3. Point-to-Point и Publish/Subscribe, брокеры Apache Kafka и RabbitMQ.
4. Сравнение CSV, XML и JSON.
5. Отличия Avro, Parquet, ORC и форматов хранения.
6. Модели данных и OLTP/OLAP.
7. Этапы жизненного цикла проекта Data Science.
8. Отличия Business Intelligence и Data Science.
9. Роли аналитика данных, Data Engineer и Data Scientist.
Раздел 4. Извлечение знаний, визуализация и бизнес-аналитика. Защита курсового проекта
1. Data Mining, Web Mining, Social Mining.
2. Назначение прямого и обратного поискового индекса.
3. Архитектура Lucene, Solr, Elasticsearch, назначение REST API.
4. Цели и формы визуализации данных.
5. Трудности визуализации больших объёмов данных и агрегации.
6. Kibana и Grafana: возможности и различия.
7. Функции BI-платформ и требования к аналитической панели.
8. Power BI, Tableau и Qlik Sense: достоинства и ограничения.
9. Структура пояснительной записки курсового проекта.
10. Обоснование методов анализа, демонстрация продукта и интерпретация результатов.
Практические задания: настройка ELK; запросы к Elasticsearch через REST API; создание индекса и запись данных; визуализация данных в Kibana; аналитический дашборд и защита курсового проекта.

5.2. Темы курсовых работ (проектов)

Лабораторный практикум не предусмотрен. По учебному плану: 30 часов лекций и 14 часов практических занятий. В учебных заданиях используются демонстрации и работа за компьютером.

5.3. Фонд оценочных средств

ПРАКТИЧЕСКИЕ РАБОТЫ И КОНТРОЛЬНЫЕ МАТЕРИАЛЫ

1. Настройка работоспособного окружения с ELK-стеком (2 ч). Установить/запустить компоненты, проверить доступность Elasticsearch и Kibana, представить конфигурацию и журнал проверок.
2. Разработка собственных визуализаций в Kibana на основе программного решения «Исследование IT-вакансий на рынке труда» (3 ч). Подготовить данные, создать минимум две визуализации, интерпретировать результат.
3. Взаимодействие с сервисами Elasticsearch через REST API (3 ч). Выполнить GET/POST-запросы, получить метаданные и результаты поиска, обработать типовые ошибки.
4. Создание индекса и запись данных в Elasticsearch (3 ч). Выбрать отображение полей, индексировать учебные документы, проверить поиск и фильтрацию.
5. Продвинутое исследование данных в Kibana: сложные рабочие столы, холсты и графы, машинное обучение (3 ч). Создать аналитический экран и пояснить выбранные показатели.
КУРСОВАЯ РАБОТА: индивидуальное аналитическое исследование с применением технологии больших данных; требования включают анализ предметной области, обоснование стека, проектирование и практическую реализацию, визуализацию и защиту результата. Исходный PDF отсылает к пособию Никифорова и соавторов «Курсовое проектирование» (2021); конкретные темы утверждает преподаватель.
КОНТРОЛЬ: письменные/устные вопросы по разделам, проверка практических заданий и отчётов, защита курсовой, экзамен.

5.4. Перечень видов оценочных средств

Контрольные вопросы и опросы по разделам; практические задания по ELK-стеку (5 работ); анализ результатов и защита практических работ; индивидуальная курсовая работа с защитой; экзамен в 1-м семестре.

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

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год
Л1.1	Барсегян А. А. и др.	Анализ данных и процессов	СПб.: БХВ-Петербург, 2009

Л1.2	Юсупова О. А., Никифоров И. В., Дробинцев П. Д.	Наука о данных и аналитика больших объемов информации. Стек технологий ELK (Elasticsearch, Logstash, Kibana)	СПб.: ПОЛИТЕХ-ПРЕСС, 2021. URL: http://elib.spbstu.ru/dl/2/i21-169.pdf

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год
Л2.1	Никифоров И. В., Воинов Н. В., Дробинцев П. Д.	Наука о данных и аналитика больших объемов информации. Курсовое проектирование	СПб.: ПОЛИТЕХ-ПРЕСС, 2021. 93 с. ISBN 978-5-7422-7352-3

6.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

Э1	Открытое образование: Наука о данных и аналитика больших объёмов информации	https://openedu.ru/course/spbstu/BIGDATA/
Э2	Юсупова О. А. и др. Стек технологий ELK. Электронная библиотека СПбПУ	http://elib.spbstu.ru/dl/2/i21-169.pdf
Э3	Курсовое проектирование по аналитике больших данных. Электронная библиотека СПбПУ	https://elib.spbstu.ru/dl/2/i21-168.pdf/en/info

6.3. Перечень информационных и образовательных технологий

6.3.1 Компетентностно-ориентированные образовательные технологии

6.3.1.1	Лекции с электронными презентациями; командные практические задания; разбор прикладных сценариев и открытых датасетов; самостоятельное изучение документации и выполнение аналитических заданий; защита курсовой работы. Используются активное обсуждение архитектур, контроль знаний, демонстрация программных решений.

6.3.2 Перечень информационных справочных систем и программного обеспечения

6.3.2.1	ОС Windows или Linux; Elasticsearch, Logstash, Kibana (ELK-стек); средства работы с REST API (например, curl/Postman); языки программирования общего назначения для написания клиентских приложений; офисные средства подготовки презентаций и отчётов; интернет-доступ. Hadoop, Spark, Apache Airflow, Kafka, RabbitMQ, Grafana, Tableau и Power BI могут демонстрироваться в учебных сценариях при наличии соответствующей инфраструктуры и лицензий.

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1	Компьютерный класс или личные компьютеры обучающихся с доступом к сети, мультимедийный проектор, ОС Windows/Linux и учебная конфигурация ELK. Для отдельных демонстраций могут использоваться виртуальные машины, контейнеры и вычислительные серверы. Доступ к вычислительному кластеру ВШПИ СПбПУ.

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

Преподаватель объявляет индивидуальное задание, исходные данные, требования к воспроизводимости и критерии оценки. До практического занятия изучаются лекция и документация. Отчёт по работе содержит цель, архитектуру/конфигурацию решения, основные действия и запросы, снимки или экспорт результатов, интерпретацию и выводы. Курсовая работа включает постановку задачи, анализ данных, обоснование выбора инструментов, реализацию прототипа или экспериментального решения, полученные показатели, сравнение с альтернативами и презентацию на защите. СР: работа с учебной литературой, самостоятельное изучение тем, подготовка к занятиям и курсовое проектирование.