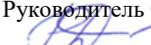




МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Дальневосточный федеральный университет»
(ДФУ)
ПЕРЕДОВАЯ ИНЖЕНЕРНАЯ ШКОЛА
«ИНСТИТУТ БИОТЕХНОЛОГИЙ, БИОИНЖЕНЕРИИ И ПИЩЕВЫХ СИСТЕМ»

СОГЛАСОВАНО
Научный руководитель ОП
 Балабанова Л.А.
(подпись) (ФИО)

Руководитель ОП
 Дремлюга Р.И.
(подпись) (ФИО)

УТВЕРЖДАЮ
И.о. декана Факультета промышленных
биотехнологий и биоинженерии
 Цыганков В.Ю.
(подпись) (И.О. Фамилия)

17 февраля 2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Анализ хранилищ больших данных
Направление подготовки 09.04.03 Прикладная информатика
Биоинформатика и анализ данных
Форма подготовки: очная

Рабочая программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта по направлению подготовки 09.04.03 Прикладная информатика, утвержденного приказом Министерства образования и науки Российской Федерации от 19 сентября 2017 г. № 916.

Директор Академии цифровой трансформации Еременко А.С.
Составители: Еременко А.С., Олейников И.С., Синягина А.Д.

Владивосток
2023

1. Рабочая программа пересмотрена на заседании Академии цифровой трансформации и утверждена на заседании Факультета промышленных биотехнологий и биоинженерии, протокол от « ____ » _____ 202__ г. № _____
2. Рабочая программа пересмотрена на заседании Академии цифровой трансформации и утверждена на заседании Факультета промышленных биотехнологий и биоинженерии, протокол от « ____ » _____ 202__ г. № _____
3. Рабочая программа пересмотрена на заседании Академии цифровой трансформации и утверждена на заседании Факультета промышленных биотехнологий и биоинженерии, протокол от « ____ » _____ 202__ г. № _____
4. Рабочая программа пересмотрена на заседании Академии цифровой трансформации и утверждена на заседании Факультета промышленных биотехнологий и биоинженерии, протокол от « ____ » _____ 202__ г. № _____
5. Рабочая программа пересмотрена на заседании Академии цифровой трансформации и утверждена на заседании Факультета промышленных биотехнологий и биоинженерии, протокол от « ____ » _____ 202__ г. № _____

Аннотация дисциплины

Анализ хранилищ больших данных

Общая трудоемкость дисциплины составляет 4 зачётных единицы / 144 академических часа. Является дисциплиной по выбору части ОП, формируемой участниками образовательных отношений, изучается на 2 курсе и завершается экзаменом. Учебным планом предусмотрено проведение лекционных занятий в объеме 18 часов, практических занятий в объеме 36 часов, а также выделены часы на самостоятельную работу студента – 90 часов (в том числе 36 часов на подготовку к экзамену).

Язык реализации: русский.

Цель: формирование у студентов знаний о сущности и значении хранилищ данных и управлении ими, изучение теоретических основ и принципов построения хранилищ данных, приобретение практических навыков аналитической обработки данных.

Задачи:

- знать основные определения, относящиеся к концепции управления хранилищами данных;
- уметь проектировать многомерные кубы данных;
- знать основные требования и средства их обеспечения к хранилищам данных;
- уметь применять глубокие специальные знания в области информационных систем и технологий для решения междисциплинарных инженерных задач по созданию интеллектуальных ГИС и технологий;
- владеть технологиями, обеспечивающими манипулирование хранилищами данных.

Для успешного изучения дисциплины у обучающихся должны быть сформированы следующие предварительные компетенции: УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий, УК-4 Способен применять современные коммуникативные технологии, в том числе на

иностранным(ых) языке(ах), для академического и профессионального взаимодействия, ОПК-1 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте, ОПК-3 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями, ОПК-4 Способен применять на практике новые научные принципы и методы исследований, ПК-1 Способен к проведению и руководству научно-исследовательскими и опытно-конструкторскими разработками при исследовании самостоятельных тем и в соответствии с тематическим планом организации, ПК-2 Способен использовать и развивать методы научных исследований и инструментария в области проектирования и управления информационными системами в прикладных областях, ПК-3 Способен проектировать и разрабатывать системные и прикладные решения по анализу больших данных, полученные в результате изучения дисциплин: «Английский язык для специальных целей», «Управление цифровой трансформацией (CDTO)», «Математические методы анализа данных», «Машинное обучение (Machine Learning)», «Биоинформатика», «Технологии сбора и анализа больших данных», «Прикладная биоинформатика: NGS-технологии и Omics-анализ». «Разработка технологических продуктов», «Правовые и этические проблемы использования технологий искусственного интеллекта», «Методы анализа и оценки IT-проектов»; обучающийся должен быть готов к прохождению производственных практик «Научно-исследовательская работа», «Преддипломная практика» и выполнению выпускной квалификационной работы, формирующих компетенции: ПК-1 Способен к проведению и руководству научно-исследовательскими и опытно-конструкторскими разработками при исследовании самостоятельных тем и в соответствии с тематическим планом организации, ПК-2 Способен использовать и развивать методы научных исследований и инструментария в области проектирования и управления

информационными системами в прикладных областях, ПК-3 Способен проектировать и разрабатывать системные и прикладные решения по анализу больших данных, ПК-4 Способен разрабатывать методики выполнения аналитических работ, ПК-5 Способен управлять проектами по информатизации прикладных задач и созданию информационных систем предприятий и организаций.

Компетенции студентов, индикаторы их достижения и результаты обучения по дисциплине:

Тип задач	Код и наименование компетенции (результат освоения)	Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения по дисциплине)
Проектный	ПК-3. Способен проектировать и разрабатывать системные и прикладные решения по анализу больших данных	ПК-3.1 Владеет инструментарием получения, хранения, передачи и обработки больших данных	Знает технологии и программное обеспечение систем хранения и обработки информации; методы и инструменты получения и приобретения, извлечения, преобразования и загрузки больших данных
			Умеет разрабатывать системы хранения и обработки больших данных, оценивать их эффективность
			Владеет навыками разработки моделей данных, адаптированных к технологиям больших данных; разработки предложений по развитию и совершенствованию системы получения, хранения, передачи и обработки больших данных
		ПК-3.2. Формулирует и решает системные и прикладные задачи анализа больших данных в области биоинформатики	Знает существующие и перспективные методы и программный инструментарий технологий больших данных в области биоинформатики, принципы и методы управления их защитой и обеспечением конфиденциальности
			Умеет проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных в области биоинформатики
			Владеет навыками создания концепции, модели, прототипа сервиса на основе аналитики

			больших данных в области биоинформатики
		ПК-3.3 Способен управлять разработкой продуктов, услуг и решений на основе больших данных	Знает существующий опыт разработки и использования продуктов, услуг и решений на основе технологий больших данных
			Умеет проводить технико-экономическое обоснование разработки новых продуктов, услуг и решений на основе больших данных
			Владеет навыками анализа и определения ценности продуктов на основе встроенной аналитики больших данных для потенциальных потребителей

Для формирования вышеуказанных компетенций в рамках дисциплины «Анализ хранилищ больших данных» применяются следующие дистанционные образовательные технологии и методы / активного / интерактивного обучения: видеоконсультация и обратная связь онлайн, лекция-беседа, работа в малых группах, разноуровневые задачи и задания.

I. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель: формирование у студентов знаний о сущности и значении хранилищ данных и управлении ими, изучение теоретических основ и принципов построения хранилищ данных, приобретение практических навыков аналитической обработки данных.

Задачи:

- знать основные определения, относящиеся к концепции управления хранилищами данных;
- уметь проектировать многомерные кубы данных;
- знать основные требования и средства их обеспечения к хранилищам данных;
- уметь применять глубокие специальные знания в области информационных систем и технологий для решения междисциплинарных инженерных задач по созданию интеллектуальных ГИС и технологий;
- владеть технологиями, обеспечивающими манипулирование хранилищами данных.

Дисциплина «Анализ хранилищ больших данных» является дисциплиной по выбору части ОП, формируемой участниками образовательных отношений. Для успешного изучения дисциплины у обучающихся должны быть сформированы следующие предварительные компетенции: УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий, УК-4 Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия, ОПК-1 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте, ОПК-3 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными

выводами и рекомендациями, ОПК-4 Способен применять на практике новые научные принципы и методы исследований, ПК-1 Способен к проведению и руководству научно-исследовательскими и опытно-конструкторскими разработками при исследовании самостоятельных тем и в соответствии с тематическим планом организации, ПК-2 Способен использовать и развивать методы научных исследований и инструментария в области проектирования и управления информационными системами в прикладных областях, ПК-3 Способен проектировать и разрабатывать системные и прикладные решения по анализу больших данных, полученные в результате изучения дисциплин: «Английский язык для специальных целей», «Управление цифровой трансформацией (CDTO)», «Математические методы анализа данных», «Машинное обучение (Machine Learning)», «Биоинформатика», «Технологии сбора и анализа больших данных», «Прикладная биоинформатика: NGS-технологии и Omics-анализ». «Разработка технологических продуктов», «Правовые и этические проблемы использования технологий искусственного интеллекта», «Методы анализа и оценки IT-проектов»; обучающийся должен быть готов к прохождению производственных практик «Научно-исследовательская работа», «Преддипломная практика» и выполнению выпускной квалификационной работы, формирующих компетенции: ПК-1 Способен к проведению и руководству научно-исследовательскими и опытно-конструкторскими разработками при исследовании самостоятельных тем и в соответствии с тематическим планом организации, ПК-2 Способен использовать и развивать методы научных исследований и инструментария в области проектирования и управления информационными системами в прикладных областях, ПК-3 Способен проектировать и разрабатывать системные и прикладные решения по анализу больших данных, ПК-4 Способен разрабатывать методики выполнения аналитических работ, ПК-5 Способен управлять проектами по информатизации прикладных задач и созданию информационных систем предприятий и организаций.

Профессиональные компетенции студентов, индикаторы их достижения и результаты обучения по дисциплине:

Тип задач	Код и наименование профессиональной компетенции (результат освоения)	Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения по дисциплине)
Проектный	ПК-3. Способен проектировать и разрабатывать системные и прикладные решения по анализу больших данных	ПК-3.1 Владеет инструментарием получения, хранения, передачи и обработки больших данных	Знает технологии и программное обеспечение систем хранения и обработки информации; методы и инструменты получения и приобретения, извлечения, преобразования и загрузки больших данных
			Умеет разрабатывать системы хранения и обработки больших данных, оценивать их эффективность
			Владеет навыками разработки моделей данных, адаптированных к технологиям больших данных; разработки предложений по развитию и совершенствованию системы получения, хранения, передачи и обработки больших данных
		ПК-3.2. Формулирует и решает системные и прикладные задачи анализа больших данных в области биоинформатики	Знает существующие и перспективные методы и программный инструментарий технологий больших данных в области биоинформатики, принципы и методы управления их защитой и обеспечением конфиденциальности
			Умеет проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных в области биоинформатики
			Владеет навыками создания концепции, модели, прототипа сервиса на основе аналитики больших данных в области биоинформатики
		ПК-3.3 Способен управлять разработкой продуктов, услуг и решений на основе больших данных	Знает существующий опыт разработки и использования продуктов, услуг и решений на основе технологий больших данных
			Умеет проводить технико-экономическое обоснование разработки новых продуктов, услуг и решений на основе больших данных
			Владеет навыками анализа и определения ценности продуктов на основе встроенной аналитики больших данных для потенциальных потребителей

II. ТРУДОЁМКОСТЬ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 4 зачётных единицы (144 академических часа).

III. СТРУКТУРА ДИСЦИПЛИНЫ

Форма обучения – очная

№	Наименование раздела дисциплины	С е м е с т р	Количество часов по видам учебных занятий и работы обучающегося					Конт- роль	Формы промежуточной аттестации
			Лек	Лаб	Пр	ОК	СР		
1.	Тема 1. Большие данные: методы и средства	4	2		6		54	36	Экзамен
2.	Тема 2. Ухищрения, применяемые для хранения больших данных	4	4		8				
3.	Тема 3. Замена требований ACID на BASE в случае с хранением больших данных	4	4		6				
4.	Тема 4. Индексы, виды индексов, наиболее часто применимые в области больших данных	4	4		8				
5.	Тема 5. Инвертированные индексы, полнотекстовые поисковые системы, первая версия инвертированного индекса от Google, принцип бареллей для хранения больших данных	4	4		8				
ИТОГО:			18		36		54	36	

IV. СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

Тема 1. Большие данные: методы и средства

Принцип трех V (Variable, Volume, Velocity). Основные средства хранения и обработки больших данных: СУБД, NoSQL СУБД, Индексы.

Тема 2. Ухищрения, применяемые для хранения больших данных

Распределенное хранение данных. Шардирование. Ранжирование. Многофазный поиск. Вероятностный поиск.

Тема 3. Замена требований ACID на BASE в случае с хранением больших данных

Замена требований ACID на BASE в случае с хранением больших данных: когда оправдана, для чего необходима, почему позволяет повысить производительность. Теорема о невозможности соблюдения ACID принципов в распределенных системах.

Тема 4. Индексы, виды индексов, наиболее часто применимые в области больших данных

Тема 5. Инвертированные индексы, полнотекстовые поисковые системы, первая версия инвертированного индекса от Google, принцип бареллей для хранения больших данных

V. СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА

ПРАКТИЧЕСКИЕ ЗАНЯТИЯ

Практическое занятие № 1. Установить хранилище clickhouse с распределенной БД (движок Distributed MergeTree) и провести следующие эксперименты:

1. Создать распределенную БД на три шарда в которых ключ шардирования совпадает / не совпадает с ключом партицирования, заполнить случайными данными, измерить скорости чтения последовательных данных, скорость чтения случайных записей. Объяснить полученную разницу и сделать выводы какой из способов пригоден для дальнейшей работы.

2. Создать в распределенной БД таблицу, ключ партицирования которой и первичный ключ – дата, время и порядковый номер (дата, время растет монотонно с ростом порядкового номера, это гарантированно условием). Требуется найти оптимальный ключ шардирования среди перечисленных ниже:

- дата, отрезанная по месяцу (шардировать каждый месяц отдельно);
- четный / не четный порядковый номер;

- случайный ключ;
- порядковый номер часа в сутках;
- порядковый номер секунды в минуте.

Практическое занятие № 2. Используя графовое хранилище данных (например, Neo4j) реализовать СУБД, хранящую информацию о текущем расписании студентов и на его основе выполнить ряд скриптов поиска в БД:

1. Скрипт, вычисляющий предметы, поставленные в «неудобное» время. Неудобным считается время, когда других предметов ни до, ни после них нет.

2. Модернизировать скрипт из задания 1, если нужно учитывать только предметы непосредственно до и после текущего (оконная функция на графовой БД).

3. Найти в БД «одинаковые» предметы. Под одинаковыми предметами понимаются предметы с одним названием. Для каждого одинакового предмета выяснить можно ли его «совместить». Совместить – означает для каждого набора дублей найти общее время, когда свободны все группы и куда можно вставить этот предмет.

Практическое занятие № 3. Для движка log в clickhouse сгенерировать или взять из сторонних сервисов стандартный лог веб-приложения, содержащий информацию о запросах, которые делали клиенты к серверу написать скрипт, который будет искать последовательности действий, совершенные одним пользователем. Пользователя можно выявлять по идентификатору сессии или по отпечатку пальца.

Практическое занятие № 4. Создать БД, содержащую данные транзакций склада или группы складов биотехнологического предприятия. На основании сгенерированных данных создать эмуляцию процессов в компании и рассчитать на каком количестве клиентов/запросов перевозок использование реляционной БД станет затрудненным и будет смысл перейти на нереляционные аналоги.

VI. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

№ п/п	Контролируемые разделы/темы дисциплины	Код и наименование индикатора достижения	Результаты обучения	Оценочные средства*		
				текущий контроль	промежуточная аттестация	
1.	Тема 1. Большие данные: методы и средства Тема 2. Ухищрения, применяемые для хранения больших данных	ПК-3.1 Владеет инструментарием получения, хранения, передачи и обработки больших данных	Знает технологии и программное обеспечение систем хранения и обработки информации; методы и инструменты получения и приобретения, извлечения, преобразования и загрузки больших данных	УО-1 ПР-2	–	
			Умеет разрабатывать системы хранения и обработки больших данных, оценивать их эффективность	ПР-7 ПР-11	–	
			Владеет навыками разработки моделей данных, адаптированных к технологиям больших данных; разработки предложений по развитию и совершенствованию системы получения, хранения, передачи и обработки больших данных	ПР-7 ПР-11	–	
	Тема 3. Замена требований ACID на BASE в случае с хранением больших данных Тема 4. Индексы, виды индексов, наиболее часто применимые в области больших данных Тема 5. Инвертированные индексы, полнотекстовые поисковые системы, первая версия инвертированного индекса от Google, принцип бареллей для хранения больших данных	ПК-3.2. Формулирует и решает системные и прикладные задачи анализа больших данных в области биоинформатики	Знает существующие и перспективные методы и программный инструментарий технологий больших данных в области биоинформатики, принципы и методы управления их защитой и обеспечением конфиденциальности	УО-1 ПР-2	–	
			Умеет проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных в области биоинформатики	ПР-7 ПР-11	–	
			Владеет навыками создания концепции, модели, прототипа сервиса на основе аналитики больших данных в области биоинформатики	ПР-7 ПР-11	–	
			ПК-3.3 Способен управлять разработкой продуктов, услуг и решений на основе больших данных	Знает существующий опыт разработки и использования продуктов, услуг и решений на основе технологий больших данных	УО-1 ПР-2	–
				Умеет проводить технико-экономическое обоснование разработки новых продуктов, услуг и решений на основе больших данных	ПР-7 ПР-11	–

			Владеет навыками анализа и определения ценности продуктов на основе встроенной аналитики больших данных для потенциальных потребителей	ПР-7 ПР-11	–
	Экзамен			–	УО-1

* Формы оценочных средств:

1) собеседование (УО-1), коллоквиум (УО-2); доклад, сообщение (УО-3); круглый стол, дискуссия, полемика, диспут, дебаты (УО-4); и т.д.

2) тесты (ПР-1); контрольные работы (ПР-2), эссе (ПР-3), рефераты (ПР-4), курсовые работы (ПР-5), научно-учебные отчеты по практикам (ПР-6); практические занятия (ПР-7); портфолио (ПР-8); проект (ПР-9); деловая и/или ролевая игра (ПР-10); разноуровневые задачи и задания (ПР-11); рабочая тетрадь (ПР-12); кроссворды (ПР-13) и т.д.

3) тренажер (ТС-1); и т.д.

VII. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Самостоятельная работа определяется как индивидуальная или коллективная учебная деятельность, осуществляемая без непосредственного руководства педагога, но по его заданиям и под его контролем. Самостоятельная работа – это познавательная учебная деятельность, когда последовательность мышления студента, его умственных и практических операций и действий зависит и определяется самим студентом.

Самостоятельная работа студентов способствует развитию самостоятельности, ответственности и организованности, творческого подхода к решению проблем учебного и профессионального уровня, что в итоге приводит к развитию навыка самостоятельного планирования и реализации деятельности.

Целью самостоятельной работы студентов является овладение необходимыми компетенциями по своему направлению подготовки, опытом творческой и исследовательской деятельности.

Формы самостоятельной работы студентов:

- работа с основной и дополнительной литературой, Интернет ресурсами;

- самостоятельное ознакомление с лекционным материалом, представленным на электронных носителях, в библиотеке образовательного учреждения;
- подготовка реферативных обзоров источников периодической печати, опорных конспектов, заранее определенных преподавателем;
- поиск информации по теме, с последующим ее представлением в аудитории в форме доклада, презентаций;
- подготовка к выполнению аудиторных контрольных работ;
- выполнение домашних контрольных работ;
- выполнение тестовых заданий, решение задач;
- составление кроссвордов, схем;
- подготовка сообщений к выступлению на семинаре, конференции;
- заполнение рабочей тетради;
- написание эссе, курсовой работы;
- подготовка к деловым и ролевым играм;
- составление резюме;
- подготовка к зачетам и экзаменам;
- другие виды деятельности, организуемые и осуществляемые образовательным учреждением и органами студенческого самоуправления.

VIII. СПИСОК ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

1. Агальцов, В.П. Базы данных: в 2 книгах. Книга 2. Распределенные и удаленные базы данных: учебник / В.П. Агальцов. - Москва: ФОРУМ : ИНФРА-М, 2021. - 271 с. - Режим доступа: <https://znanium.com/catalog/document?id=377105>
2. Агальцов, В.П. Базы данных: учебник: в 2 кн. Книга 1. Локальные базы данных / В.П. Агальцов. - Москва: ФОРУМ : ИНФРА-М, 2021. - 352 с. - Режим доступа: <https://znanium.com/catalog/document?id=398558>

3. Дадян, Э.Г. Методы, модели, средства хранения и обработки данных: учебник / Э.Г. Дадян, Ю.А. Зеленков. - Москва: Вузовский учебник : ИНФРА-М, 2022. - 168 с. - Режим доступа:

<https://znanium.com/catalog/document?id=384994>

4. Колдаев, В.Д. Структуры и алгоритмы обработки данных: учебное пособие / В.Д. Колдаев. - Москва: РИОР : ИНФРА-М, 2021. - 296 с. - Режим доступа: <https://znanium.com/catalog/document?id=398591>

5. Логунова, О.С. Обработка экспериментальных данных на ЭВМ: учебник / О.С. Логунова, П.Ю. Романов, Е.А. Ильина. - 2-е изд., испр. и доп. - Москва: ИНФРА-М, 2023. - 377 с. - Режим доступа:

<https://znanium.com/catalog/document?id=426848>

6. Мартишин, С.А. Базы данных. Практическое применение СУБД SQL и NoSQL-типа для проектирования информационных систем: учебное пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. — Москва: ФОРУМ : ИНФРА-М, 2023. - 368 с. - Режим доступа:

<https://znanium.com/catalog/document?id=429260>

7. Мартишин, С.А. Базы данных: работа с распределенными базами данных и файловыми системами на примере MongoDB и HDFS с использованием Node.js, Express.js, Apache Spark и Scala: учебное пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. - Москва: ИНФРА-М, 2023. - 235 с. - Режим доступа: <https://znanium.com/catalog/document?id=416159>

8. Полищук, Ю.В. Базы данных и их безопасность: учебное пособие / Ю.В. Полищук, А.С. Боровский. - Москва: ИНФРА-М, 2023. - 210 с. - Режим доступа: <https://znanium.com/catalog/document?id=418255>

Дополнительная литература

1. Баранчиков, А.И. Алгоритмы и модели доступа к записям баз данных / А.И. Баранчиков, П.А. Баранчиков, А.Н. Пылькин. - М.: Гор. линия-Телеком, 2011. - 182 с. - Режим доступа:

<http://znanium.com/catalog/product/326291>

2. Боровиков, В.П. Популярное введение в современный анализ данных

в системе STATISTICA: Учебное пособие для вузов / В.П. Боровиков. - М.: Гор. линия-Телеком, 2013. - 288 с. - Режим доступа: <http://znanium.com/catalog/product/425084>

3. Буренин, С.Н. Web-программирование и базы данных [Электронный ресурс]: учебный практикум / Буренин С.Н. - М.: Московский гуманитарный университет, 2014. - 120 с. - Режим доступа: <http://www.iprbookshop.ru/39683.html>

4. Жуковский О.И. Информационные технологии и анализ данных [Электронный ресурс]: учебное пособие / О.И. Жуковский. - Томск: Томский государственный университет систем управления и радиоэлектроники, Эль Контент, 2014. - 130 с. - Режим доступа: <http://www.iprbookshop.ru/72106.html>

5. Изучаем Spark: молниеносный анализ данных [Электронный ресурс] / Х. Карау [и др.]. - Москва: ДМК Пресс, 2015. - 304 с. - Режим доступа: <https://e.lanbook.com/book/90118>

6. Королев, В.Т. Технология ведения баз данных [Электронный ресурс]: учебное пособие / Королев В.Т., Контарев Е.А., Черных А.М. - М.: Российский государственный университет правосудия, 2015. - 108 с. - Режим доступа: <http://www.iprbookshop.ru/45233.html>

7. Методы и модели эконометрики. Часть 1. Анализ данных [Электронный ресурс]: учебное пособие / О.И. Бантикова [и др.]. - Оренбург: ОГУ, 2015. - 574 с. - Режим доступа: <https://e.lanbook.com/book/98113>

8. Панова, Н.Ф. FireBird. Установка, разработка баз данных, реализация запросов [Электронный ресурс]: методические указания / Панова Н.Ф. - Оренбург: Оренбургский государственный университет, ЭБС АСВ, 2014. - 45 с. - Режим доступа: <http://www.iprbookshop.ru/50015.html>

9. Селиванова, И.А. Построение и анализ алгоритмов обработки данных [Электронный ресурс]: учебно-методическое пособие / Селиванова И.А., Блинов В.А. - Екатеринбург: Уральский федеральный университет, ЭБС АСВ, 2015. - 108 с. - Режим доступа: <http://www.iprbookshop.ru/68277.html>

10. Сурина, Е.Е. Методы анализа данных [Электронный ресурс]: учебное пособие / Е.Е. Сурина. - Москва: ФЛИНТА, 2015. - 130 с. - Режим доступа: <https://e.lanbook.com/book/72701>

11. Уэс, М. Python и анализ данных [Электронный ресурс] / М. Уэс; пер. с англ. Слинкин А.А. - Москва: ДМК Пресс, 2015. - 482 с. - Режим доступа: <https://e.lanbook.com/book/73074>

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Корпоративные информационные системы. Портал «Корпоративный менеджмент». Библиотека управления, статьи и пособия. - Режим доступа: <http://www.cfin.ru/software/kis/>

2. Портал Ассоциации предприятий компьютерных и информационных технологий (АКИТ). - Режим доступа: <http://www.apkit.ru>

3. Справочная система «Кодекс». - Режим доступа: <http://www.kodeks.ru/>

4. Справочно-правовая система «Гарант». - Режим доступа: www.garant.ru

5. Справочно-правовая система «КонсультантПлюс». - Режим доступа: <http://www.consultant.ru/>

6. Электронная библиотека online статей по информационным технологиям. - Режим доступа: <http://www.citforum.ru/>

Перечень информационных технологий и программного обеспечения

1. Программное обеспечение: Microsoft Word, Microsoft Excel, Microsoft PowerPoint, ОС Windows.

IX. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Успешное освоение дисциплины предполагает активную работу студентов на всех занятиях аудиторной формы: лекциях и практиках, выполнении аттестационных мероприятий. В процессе изучения дисциплины

студенту необходимо ориентироваться на проработку теоретического материала и подготовку к практическим занятиям.

Освоение дисциплины «Анализ хранилищ больших данных» предполагает рейтинговую систему оценки знаний студентов и предусматривает со стороны преподавателя текущий контроль за посещением студентами лекций, подготовкой и выполнением всех практических заданий, выполнением всех видов самостоятельной работы.

Промежуточной аттестацией по дисциплине «Анализ хранилищ больших данных» является экзамен.

Студент считается аттестованным по дисциплине при условии выполнения всех видов текущего контроля и самостоятельной работы, предусмотренных учебной программой.

Шкала оценивания сформированности образовательных результатов по дисциплине представлена в фонде оценочных средств (ФОС).

X. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебные занятия по дисциплине «Анализ хранилищ больших данных» проводятся в помещениях, оснащенных соответствующим оборудованием и программным обеспечением.

Перечень материально-технического и программного обеспечения дисциплины приведен в таблице.

Материально-техническое и программное обеспечение дисциплины

Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
Учебная аудитория для проведения занятий лекционного и семинарского типа, оснащенная оборудованием и техническими средствами обучения (690922, Приморский край, г. Владивосток, о. Русский, п. Аякс, 10, корпус G, каб. G513)	Экран проекционный ScreenLine Trim White Ice 50 см черная кайма сверху, размер рабочей области 236x147 см Документ-камера Avervision CP355AF ЖК-панель 47", Full HD, LG M4716 ССВА Мультимедийный проектор, Mitsubishi EW330U, 3000 ANSI Lumen, 1280x800 Сетевая видеочамера Multipix MP-HD718	

<p>Учебная аудитория для проведения практических и лабораторных занятий (Компьютерный класс) (690922, Приморский край, г. Владивосток, о. Русский, п. Аякс, 10, корпус G, каб. G244)</p>	<p>Комплект учебной мебели (столы и стулья), ученическая доска, мультимедийное оборудование. Моноблок Lenovo C360 19,5 (1600x900), Pentium G3220T, 4GB DDR3-1600 (1x4GB), 500GB HDD 7200 SATA, DVD+/-RW, GigEth, Wi-Fi, BT, usb kbd/mse, Win7 Корпоративная (64- bit) (23 шт.) Экран с электроприводом 236*147 см Trim Screen Line; Проектор DLP, 3000 ANSI Lm, WXGA 1280x800, 2000:1 EW330U Mitsubishi; Подсистема специализированных креплений оборудования CORSA-2007 Tuarex; Подсистема видеокоммутации; Подсистема аудиокоммутации и звукоусиления; акустическая система для потолочного монтажа SI 3CT LP Extron; цифровой аудиопроцессор DMP 44 LC Extron</p>	
<p>Аудитории для самостоятельной работы студентов (690922, Приморский край, г. Владивосток, о. Русский, п. Аякс, 10, корпус А, каб. А1007 (А1042))</p>	<p>Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду ДВФУ. Комплекты учебной мебели (столы и стулья). Моноблок Lenovo C360G-i34164G500UDK – 115 шт. Интегрированный сенсорный дисплей Polymedia FlipBox. Копир-принтер-цветной сканер в e-mail с 4 лотками Xerox WorkCentre 5330 (WC5330C). Полноцветный копир-принтер-сканер Xerox WorkCentre 7530 (WC7530CPS). Скорость доступа в Интернет 500 Мбит/сек. Рабочие места для людей с ограниченными возможностями здоровья оснащены дисплеями и принтерами Брайля; оборудованы: портативными устройствами для чтения плоскочечатных текстов, сканирующими и читающими машинами видеоувеличителем с возможностью регуляции цветовых спектров; увеличивающими электронными лупами и ультразвуковыми маркировщиками</p>	