



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Дальневосточный федеральный университет»
(ДФУ)
ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ (ШКОЛА)

СОГЛАСОВАНО
Руководитель ОП


А.С. Клеин

«УТВЕРЖДАЮ»

И.о. директора департамента


Еременко А.С.

« 27 » сентября 2021 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Администрирование хранилищ больших данных

Направление подготовки - 09.04.01 Информатика и вычислительная техника
(Искусственный интеллект и большие данные (совместно с ПАО Сбербанк))

Форма подготовки очная

курс 2 семестр 3
лекции 18 час.
практические занятия 36 час.
лабораторные работы 00 час.
в том числе с использованием МАО практические занятия 36 час
всего часов аудиторной нагрузки 54 час.
самостоятельная работа 54 час.
в том числе на подготовку к экзамену 00 час.
контрольные работы (количество) не предусмотрены
курсовая работа / курсовой проект не предусмотрены
зачет с оценкой 3 семестр
экзамен не предусмотрен

Рабочая программа дисциплины разработана в соответствии с федеральным государственным образовательным стандартом (ФГОС) высшего образования (ВО) – магистратура по направлению подготовки 09.04.01 Информатика и вычислительная техника, утвержденного приказом Министерства образования и науки Российской Федерации от 19.09.2017 г. № 918 (с изменениями и дополнениями).

Рабочая программа обсуждена на заседании Академии цифровой трансформации протокол № 9.2 от 27 сентября 2021 г.

И.о. директора Академии цифровой трансформации

Еременко А.С.

Составители: к.т.н. Еременко А.С., Олейников И.С.

Владивосток
2021

Оборотная сторона титульного листа РПУД

I. Рабочая программа пересмотрена на заседании департамента:

Протокол от « _____ » _____ 20__ г. № _____

Директор департамента _____
(подпись) (И.О. Фамилия)

II. Рабочая программа пересмотрена на заседании департамента:

Протокол от « _____ » _____ 20__ г. № _____

Директор департамента _____
(подпись) (И.О. Фамилия)

III. Рабочая программа пересмотрена на заседании департамента:

Протокол от « _____ » _____ 20__ г. № _____

Директор департамента _____
(подпись) (И.О. Фамилия)

IV. Рабочая программа пересмотрена на заседании департамента:

Протокол от « _____ » _____ 20__ г. № _____

Директор департамента _____
(подпись) (И.О. Фамилия)

I. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины - 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.

В результате изучения данной дисциплины у обучающихся формируются следующие профессиональные компетенции:

Код и наименование профессиональной компетенции (результат освоения)	Код и наименование индикатора достижения компетенции
ПК-2 Способен разрабатывать методики выполнения аналитических работ	ПК-2.1 Умеет выявлять проблемы и сложности в существующих практиках выполнения аналитических работ в организации; описывать методики выполнения аналитических работ ПК-2.2 Владеет навыками выполнения аналитических работ, их апробации и доработки на выбранных проектах
ПК-4 Способен ставить цели и принимать управленческие решения, основанные на анализе больших данных	ПК-4.1 Владеет навыками стратегического управления развитием методологической и технологической инфраструктуры анализа больших данных в организации ПК-4.2 Определяет необходимые ресурсы и инструменты для решения задач с использованием анализа данных; руководит работой команды, вырабатывая командную стратегию на основе анализа данных

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения)
ПК-2.1 Умеет выявлять проблемы и сложности в существующих практиках выполнения аналитических работ в организации; описывать методики выполнения аналитических работ	Знает основные методики и практики выполнения аналитических работ
	Умеет выявлять проблемы и сложности в существующих практиках выполнения аналитических работ в организации
	Владеет навыками выполнения аналитических работ, их апробации и доработки на выбранных проектах
ПК-2.2 Владеет навыками выполнения аналитических работ, их апробации и доработки на выбранных проектах	Знает методы, применяемые для функционального и оперативного управления предприятиями
	Планировать проектные работы
	Владеет навыками выполнения аналитических работ, их апробации и доработки на выбранных проектах
ПК-4.1 Владеет навыками стратегического управления развитием методологической и технологической инфраструктуры анализа больших данных в организации	Знает существующие и перспективные методы и программный инструментарий технологий больших данных
	Умеет управлять развитием технологической инфраструктуры анализа больших данных
	Владеет навыками стратегического управления
ПК-4.2 Определяет необходимые ресурсы и инструменты для решения задач с использованием анализа данных; руководит работой команды, вырабатывая командную стратегию на основе анализа данных	Знает методы создания программного обеспечения для анализа и обработки данных
	Умеет использовать методы проектирования систем анализа и обработки данных
	Владеет навыками работы в распределенных командах

II. ТРУДОЁМКОСТЬ ДИСЦИПЛИНЫ И ВИДОВ УЧЕБНЫХ ЗАНЯТИЙ ПО ДИСЦИПЛИНЕ

Общая трудоемкость дисциплины составляет 3 зачётных единицы (108 академических часов).

Видами учебных занятий и работы обучающегося по дисциплине могут являться:

Обозначение	Виды учебных занятий и работы обучающегося
Лек	Лекции
Пр	Практические занятия
СР	Самостоятельная работа обучающегося в период теоретического обучения

Структура дисциплины:

Форма обучения – очная.

№	Наименование раздела дисциплины	Семестр	Количество часов по видам учебных занятий и работы обучающегося					Контроль	Формы текущего контроля успеваемости и промежуточной аттестации
			Лек	Лаб	Пр	ОК	СР		
1	Большие данные, что это такое, принцип трех V	3	2		6		54		УО-1, ПР-7; УО-1, ПР-9; УО-1, ПР-9;
2	Ухищрения, применяемые для хранения больших данных	3	4		6				
3	Замена требований ACID на BASE в случае с хранением больших данных	3	4		8				
4	Индексы, виды индексов, наиболее часто применимые в области больших данных	3	4		8				
5	Инвертированные индексы, полнотекстовые поисковые системы, первая версия инвертированного индекса от Google, принцип бареллей для хранения больших данных	3	4		8				
Итого:			18		36		54		Зачет с оценкой

III. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

(18 часов)

Тема 1. Большие данные, что это такое, 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

IV. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА

Практическая часть курса включает в себя практические работы (36 часов), в процессе которых решаются конкретные прикладные задачи распознавания образов.

Практические работы (36 час.)

Практическая работа 1. Установить хранилище clickhouse с распределенной БД (движок Distributed MergeTree) В котором 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 — датавремя и порядковый номер (датавремя растет монотонно с ростом порядкового номера, это гарантировано условием) Требуется найти оптимальный ключ шардирования среди перечисленных ниже:

1. дата, отрезанная по месяцу (шардировать каждый месяц отдельно)
2. четный/не четный порядковый номер
3. случайный ключ
4. порядковый номер часа в сутках
5. порядковый номер секунды в минуте

Практическая задача 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

- ✓ Скрипт, вычисляющий предметы, поставленные в «неудобное» время. Неудобным считается время, когда других предметов ни до ни после них нет
- ✓ Модернизировать скрипт из задания 1, если нужно учитывать только предметы НЕПОСРЕДСТВЕННО до и после текущего (оконная функция на графовой БД)
- ✓ Найти в БД «одинаковые» предметы. Под одинаковыми предметами понимаем предметы с одним названием, читаемые одним и тем же лектором для разных групп одной специальности в разное время.
- ✓ Бонус задание к пункту 3 — 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

Практическая работа 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 1) — идентификатору сессии или 2) отпечатку пальца

Практическая работа 4. Создать БД, содержащую данные транзакций склада или группы складов предприятия по логистике. 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.

Применяются методы интерактивного обучения, тематическая дискуссия.

V. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

План-график выполнения самостоятельной работы по дисциплине, в том числе примерные нормы времени на выполнение по каждому заданию.

Примерная дата проведения	Наименование контрольного мероприятия	Форма контроля	Нормы времени на выполнение
1-6 недели	Работа с конспектом	Просмотр и проверка выполнения самостоятельной работы преподавателем, обсуждение результатов выполненной работы на занятии	20 час.
7-12 недели	Работа с конспектом, работа с литературой, подготовка к практическим работам	Устный опрос, собеседование с группой.	20 час.
13-18 недели	Защита практических работ	Презентация проектов	14 час.
Итого			54 часа

Учебно-методическое обеспечение самостоятельной работы студента по дисциплине «Администрирование хранилищ больших данных» предусматривает:

- поиск дополнительной литературы, к которой студенты могут прибегать при возникновении особой заинтересованности в конкретной теме;
- определение перечня контрольных вопросов, позволяющих студентам самостоятельно проверить качество полученных знаний;
- организацию консультаций преподавателя со студентами для разъяснения вопросов, вызывающих у студентов затруднения при самостоятельном освоении учебного материала.

Дополнительными формами самостоятельной работы являются групповые и индивидуальные задания, выступающие продолжением аудиторных занятий и направленные на овладение практическими навыками по основным разделам дисциплины.

Материалы для организации самостоятельной работы студентов

Самостоятельная работа студентов состоит из подготовки к практическим занятиям, работы над рекомендованной литературой, написания докладов по теме занятия, подготовки презентаций, решения творческих задач, подготовка проектов.

При организации самостоятельной работы преподаватель должен учитывать уровень подготовки каждого студента и предвидеть трудности, которые могут возникнуть при выполнении самостоятельной работы. Преподаватель дает каждому студенту индивидуальные и дифференцированные задания. Некоторые из них могут осуществляться в группе (например, подготовка доклада и презентации по одной теме могут делать несколько студентов с разделением своих обязанностей – один готовит научно-теоретическую часть, а второй проводит анализ практики).

Рекомендации к самостоятельной работе на лекции

Студенту необходимо быть готовым к лекции до прихода лектора в аудиторию, так как именно в первую минуту объявляется тема, формулируется основная цель, дается перечень важнейших вопросов. Без этого дальнейшее понимание лекции затрудняется.

Эффективность познавательной деятельности студента при слушании всецело зависит от направленности его внимания. Внимание обусловлено единством субъективных и объективных причин. В зависимости от действия этих причин оно может быть произвольным, т.е. возникает помимо сознательного намерения человека, и произвольным, сознательно регулируемым, направляемым. Работа студента на лекции – сложный процесс, включающий в себя слушание, осмысливание и собственно конспектирование (запись).

Умение студента слышать на лекции преподавателя является лишь первым шагом в процессе осмысленного слушания, который включает в себя несколько этапов, начиная от восприятия речи и кончая оценкой сказанного.

Лекцию необходимо записывать, вести краткие конспекты, где формулировались бы наиболее важные моменты, основные положения, излагаемые лектором. Обычно запись производится в специальной тетради. При оформлении конспекта лекции необходимо оставлять поля, где студент может записать свои собственные мысли, возникающие параллельно с мыслями, высказанными лектором, а также вопросы, которые могут возникнуть в процессе слушания, чтобы получить на них ответы при самостоятельной проработке материала лекции, при изучении рекомендованной литературы или непосредственно у преподавателя в конце лекции.

Основное отличие конспекта от текста – отсутствие или значительное снижение избыточности, то есть удаление отдельных слов или частей текста, не выражающих значимой информации, а также замена развернутых оборотов текста более лаконичными словосочетаниями (свертывание). При конспектировании основную информацию следует записывать подробно, а

дополнительные и вспомогательные сведения, примеры – очень кратко. Умение отделять основную информацию от второстепенной – одно из основных требований к конспектирующему. Хорошие результаты в выработке умения выделять основную информацию дает известный приём, названный условно приемом фильтрации и сжатия текста, который включает в себя две операции:

1. Разбивку текста на части по смыслу.
2. Нахождение в каждой части текста одного слова краткой фразы или обобщающей короткой формулировки, выражающих основу содержания этой части.

Рекомендуется применять систему условных сокращений. В первую очередь сокращаются длинные слова и те, что повторяются в речи лектора чаще всего. При этом само сокращение должно быть по возможности кратким. Основные термины, повторяющиеся наиболее часто, могут быть выделены как ключевые слова и обозначены начальными заглавными буквами этих слов (сокращение, называемое аббревиатурой). Ключевые слова записываются первый раз полностью, после чего в скобках дается их аббревиатура. Процесс записи значительно облегчается при использовании сокращений общепринятых вспомогательных слов. В самостоятельной работе над лекцией целесообразным является использование студентами логических схем. Они в наглядной форме раскрывают содержание и взаимосвязь категорий, законов, понятий, наиболее важных фактов.

Прослушанный материал лекции студент должен проработать. Насколько эффективно он это сделает, зависит и прочность усвоения знаний. Опыт показывает, что только многообразная, планомерная и целенаправленная обработка лекционного материала обеспечивает его надежное закрепление в долговременной памяти человека.

Повторение нужно разнообразить. При первом повторении изучаются все параграфы и абзацы, при втором, возможно, будет достаточно рассмотреть только отдельные параграфы, а в дальнейшем лишь тему лекции.

Необходимым является подготовка студента к предстоящей лекции. Основным требованием, предъявляемым к такой работе, является, прежде всего, систематичность ее проведения. Она включает ряд важных познавательно-практических этапов: чтение записей, сделанных в процессе слушания и конспектирования предыдущей лекции, вынесение на поля всего, что требуется при дальнейшей работе с конспектом и учебником; техническое оформление записей (подчеркивание, выделение главного, выводов, доказательств); выполнение практических заданий преподавателя; знакомство с материалом предстоящей лекции по учебнику и дополнительной литературе.

Методические рекомендации для написания конспектов

Конспекты, написанные от руки, предоставляются преподавателю для оценки (зачёт/незачёт). Учитывая, что в большинстве случаев тексты первоисточников весьма объёмные, для конспектирования можно выбрать только страницы, разделы или главы (30-50 стр. печатного текста). Объём законспектированного текста в тетради определяется самим студентом.

VI. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

Для текущей аттестации при изучении дисциплины «Администрирование хранилищ баз данных» используются следующие оценочные средства:

1) Устный опрос (УО):

Собеседование (консультация с преподавателем) (УО-1)

2) Письменные работы (ПР):

Конспект (ПР-7)

Проект (ПР-9)

№ п/п	Контролируемые разделы / темы дисциплины	Код и наименование индикатора достижения	Результаты обучения	Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Большие данные, что это такое, принцип трех V	ПК-2.1, ПК-2.2 ПК-4.1, ПК-4.2	знает	УО-1, ПР-7, ПР-9	Вопросы к зачету
			умеет		
			владеет		
2	Ухищрения, применяемые для хранения больших данных	ПК-2.1, ПК-2.2 ПК-4.1, ПК-4.2	знает	УО-1, ПР-7, ПР-9	Вопросы к зачету
			умеет		
			владеет		
3	Замена требований ACID на BASE в случае с хранением больших данных	ПК-2.1, ПК-2.2 ПК-4.1, ПК-4.2	знает	УО-1, ПР-7, ПР-9	Вопросы к зачету
			умеет		
			владеет		
4	Индексы, виды индексов, наиболее часто применимые в области больших данных	ПК-2.1, ПК-2.2 ПК-4.1, ПК-4.2	знает	УО-1, ПР-7, ПР-9	Вопросы к зачету
			умеет		
			владеет		
5	Инвертированные индексы, полнотекстовые поисковые системы, первая версия инвертированного индекса от Google, принцип бареллей для хранения больших данных	ПК-2.1, ПК-2.2 ПК-4.1, ПК-4.2	знает	УО-1, ПР-7, ПР-9	Вопросы к зачету
			умеет		
			владеет		

VII. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

(печатные и электронные издания)

1. Oracle. Что такое Big Data? [Электронный ресурс]. URL: <https://www.oracle.com/cis/big-data/what-is-big-data/>
2. Turtle H., Flood J. Query evaluation: Strategies and optimizations [Электронный ресурс]. URL: <https://www.sciencedirect.com/science/article/abs/pii/030645739500020H>
3. Manyika J. et al. Big data: The next frontier for innovation, competition, and productivity [Электронный ресурс]. URL: http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx
4. Yandex Cloud. Шардирование таблиц в ClickHouse [Электронный ресурс]. URL: <https://cloud.yandex.ru/docs/managed-clickhouse/tutorials/sharding>
5. ClickHouse. Движок Distributed в clickhouse [Электронный ресурс]. URL: <https://clickhouse.com/docs/ru/engines/table-engines/special/distributed>
6. Маркитантов А. FAISS: Быстрый поиск лиц и клонов на многомиллионных данных [Электронный ресурс]. URL: <https://habr.com/ru/companies/okkamgroup/articles/509204/> (
7. Gadgets Help. Отказ от ACID в пользу BASE в области разработки баз данных [Электронный ресурс]. URL: <https://gadgetshelp.com/prilozheniia/otkaz-ot-acid-v-polzu-base-v-oblasti-razrabotki-baz-dannykh/>
8. Yandex Cloud. NoSQL: виды, особенности и применение [Электронный ресурс]. URL: <https://cloud.yandex.ru/blog/posts/2022/10/nosql>
9. Трифионов А.А. Алгоритмы построения инвертированного индекса для коллекции текстовых данных [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/algoritmy-postroeniya-invertirovannogo-indeksa-dlya-kollektsii-tekstovyyh-dannyh>
10. Ливеровский П. Как бороться с репостами или пара слов о перцептивных хешах [Электронный ресурс]. URL: <https://habr.com/ru/articles/237307/>

11. Алиар А. «Выглядит похоже». Как работает перцептивный хэш [Электронный ресурс]. URL: <https://habr.com/ru/articles/120562/>
12. Brin S., Page L. The anatomy of a large-scale hypertextual Web search engine [Электронный ресурс]. URL: <https://www.sciencedirect.com/science/article/abs/pii/S016975529800110X>
13. Подольский П. Устройство поисковых систем: базовый поиск и инвертированный индекс [Электронный ресурс]. URL: <https://habr.com/ru/articles/545634/> (дата обращения: 22.10.2022).

Дополнительная литература

(печатные и электронные издания)

1. Отказ от ACID в пользу BASE в области разработки баз данных, <https://gadgetshelp.com/prilozheniia/otkaz-ot-acid-v-polzu-base-v-oblasti-razrabotki-baz-dannykh/>
2. NoSQL: виды, особенности и применение, <https://cloud.yandex.ru/blog/posts/2022/10/nosql>

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. «ИТ-образование в Рунете». Образовательные ресурсы Рунета: <http://ifets.ieee.org/russian/depository/resource.htm>
2. «Российский общеобразовательный портал»: <http://www.school.edu.ru/>
3. «Издание литературы в электронном виде»: <http://www.magister.msk.ru/library/library.htm>
4. Annual Review: <http://www.annualreviews.org/ebvc>
5. Scopus - мультидисциплинарная реферативная база данных: <http://www.scopus.com/>
6. Единая коллекция образовательных ресурсов: <http://school-collection.edu.ru/catalog/>
7. Информационные ресурсы Российской Библиотечной Ассоциации (РБА): <http://www.rba.ru/>
8. Каталог электронных ресурсов научной библиотеки ДВФУ: <http://www.dvfu.ru/web/library/elib>
9. Коллекция журналов издательства Elsevier на портале ScienceDirect: <http://www.sciencedirect.com/>.
10. Научная электронная библиотека (НЭБ): <http://www.elibrary.ru/>
11. Портал «Гуманитарное образование» <http://www.humanities.edu.ru/index.html>

12. Российская государственная библиотека (электронный каталог):
<http://www.rsl.ru/>
13. Университетская информационная система Россия (УИС Россия):
<http://uisrussia.msu.ru>
14. Электронная библиотечная система «Айбукс»: <http://ibooks.ru/>
15. Электронная библиотечная система «Университетская библиотека»:
www.biblioclub.ru.
16. Электронная библиотечная система издательства «Лань»:
<http://e.lanbook.com/>

Перечень информационных технологий и программного обеспечения

При осуществлении образовательного процесса студентами и профессорско-преподавательским составом используется следующее программное обеспечение: Microsoft Office (Access, Excel, PowerPoint, Word и т. д), Open Office, Skype, программное обеспечение электронного ресурса сайта ДВФУ, включая ЭБС ДВФУ.

VIII. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Работа с теоретическими материалами. Изучение дисциплины следует начинать с проработки тематического плана лекций, уделяя особое внимание структуре и содержанию темы и основных понятий. Изучение «сложных» тем следует начинать с составления логической схемы основных понятий, категорий, связей между ними. Целесообразно прибегнуть к классификации материала, в частности при изучении тем, в которых присутствует большое количество незнакомых понятий, категорий, теорий, концепций, либо насыщенных информацией типологического характера. Студенты должны составлять конспекты лекций, систематически готовиться к практическим занятиям, вести глоссарий и быть готовы ответить на контрольные вопросы в ходе лекций и аудиторных занятий. Успешное освоение программы курса предполагает прочтение ряда оригинальных работ и выполнение практических заданий.

Подготовка и выполнение практических заданий. По каждой теме дисциплины предлагаются вопросы и практические задания. Перед выполнением заданий изучите теорию вопроса, предполагаемого к исследованию. Самостоятельная работа студентов заключается:

- в подготовке к практическим занятиям в форме консультаций и дискуссий;
- в выполнении индивидуальных и групповых заданий,

- в подготовке к защите курсовой работы,
- в подготовке к итоговому собеседованию.

Цель практических (семинарских) занятий – научить студентов самостоятельно анализировать учебную и научную литературу и вырабатывать у них опыт самостоятельного мышления по проблемам курса, а также выработать навыки практического применения теоретических знаний. Как правило, семинары проводятся в виде практик-консультаций с элементами дискуссии. При этой форме работы отдельным студентам могут поручаться сообщения по тому или иному вопросу, а также ставя дополнительные вопросы, как всей аудитории, так и определенным участникам обсуждения.

Самостоятельная работа студентов, предусмотренная учебным планом, соответствует более глубокому усвоению изучаемого курса, формирует навыки исследовательской работы и ориентирует на умение применять теоретические знания на практике.

Материалом для подготовки могут стать конспекты лекций, профессиональная литература, учебно-методическое обеспечение дисциплины.

Методические рекомендации для написания конспектов

Конспекты, написанные от руки, предоставляются преподавателю для оценки (зачёт/незачёт). Учитывая, что в большинстве случаев тексты первоисточников весьма объёмные, для конспектирования можно выбрать только страницы, разделы или главы (30-50 стр. печатного текста). Объём законспектированного текста в тетради определяется самим студентом.

Методические указания к выполнению проектного задания

Выполнение проектного задания в рамках дисциплины является обязательным и предполагает индивидуальную или групповую работу.

Проект – совокупность мероприятий, направленных на достижение определённой и четко структурированной цели в конкретные сроки с привлечением оптимальных средств и ресурсов.

Проект представляет собой конечный продукт, получаемый в результате планирования и выполнения комплекса учебных и исследовательских заданий. Позволяет оценить умения обучающихся самостоятельно конструировать свои знания в процессе решения практических задач и проблем, ориентироваться в информационном пространстве и уровень сформированности аналитических, исследовательских навыков, навыков практического и творческого мышления. Может выполняться в индивидуальном порядке или группой обучающихся.

Технология разработки проектов включает в себя следующие этапы:

разработка замысла проекта в соответствии с требованиями программы по следующей структуре:

- аудитория проекта (т.е. характеристика проблем целевой группы и лиц, непосредственно получающих пользу от проекта);
- цели и задачи проекта;
- содержание проекта;
- организация-исполнитель (или форма реализации проекта);
- планируемые результаты и критерии эффективности.

Проект считается выполненным полностью в случае

1. Предоставления полного объема учебных материалов по заранее утвержденной теме, полностью раскрывающих заявленную тему;
2. Предоставления материалов на электронном носителе и в печатном виде;
3. Соответствия представленных материалов требованиям по оформлению;
4. Наличия в материалах проекта описания методики использования ЦОР;
5. Успешной презентации и защиты проекта

Проект считается выполненным полностью в случае

1. Предоставления полного объема учебных материалов по заранее утвержденной теме, полностью раскрывающих заявленную тему;
2. Предоставления материалов на электронном носителе и в печатном виде;
3. Соответствия представленных материалов требованиям по оформлению;
4. Наличия в материалах проекта описания методики использования ЦОР;
5. Успешной презентации и защиты проекта

Задание на проектирование:

1. Создание правил и сопутствующих расчетов для игры;
2. Создание персонажей, предметов, загадок и миссий;
3. Создание уровней игры, включающей ландшафт карты и расположение на этой карте объектов;
4. Продумывание пространств, локаций, как они диктуют пользовательский опыт и увязываются с общей задумкой игры.

IX. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебные занятия по дисциплине 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
Мультимедийная аудитория: G467	<p>Проектор DLP, 3000 ANSI Lm, WXGA 1280x800, 2000:1 EW330U Mitsubishi; Моноблок HP ProOne 440 G3 23.8" All-in-One, диагональ экрана 23.8", разрешение экрана 1920x1080, Bluetooth, Wi-Fi, операционная система: Windows 10 Enterprise, оптический привод DVD, процессор: Intel Core i5-7500T, размер оперативной памяти: 8 ГБ, видеопроцессор: Intel HD Graphics 630, объем жесткого диска: 1Тб.</p> <p>Беспроводные ЛВС для обучающихся обеспечены системой на базе точек доступа 802.11a/b/g/n 2x2 MIMO(2SS). AfterEffects</p>	Techdesigner, MAX8, VVVV, Adobe Photoshop, Adobe Premier, Adobe
Мультимедийная аудитория: G469	<p>Проектор DLP, 4000 ANSI Lm, 1920x1080, 2000:1 FD630u Mitsubishi;</p> <p>Проектор DLP, 2800 ANSI Lm, 1920x1080, 2000:1 GT1080 Optoma; Проектор DLP, 3000 ANSI Lm, WXGA 1280x800, 2000:1 EW330U Mitsubishi;</p> <p>Беспроводные ЛВС для обучающихся обеспечены системой на базе точек доступа 802.11a/b/g/n 2x2 MIMO(2SS).</p> <p>Специализированное оборудование: Платформа Arduino UNO, Бесконтактный сенсорный Microsoft Kinect 2.0, Аудио система Dialog 2.0, MIDI контроллер Playtron, Одноплатный компьютер Raspberry PI</p>	Techdesigner, MAX8, VVVV, Adobe Photoshop, Adobe Premier, Adobe

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.

В целях обеспечения специальных условий обучения инвалидов и лиц с ограниченными возможностями здоровья в ДВФУ все здания оборудованы пандусами, лифтами, подъемниками, специализированными местами, оснащенными туалетными комнатами, табличками информационно-навигационной поддержки.

X. ФОНДЫ ОЦЕНОЧНЫХ СРЕДСТВ

Для текущей аттестации при изучении дисциплины «Администрирование хранилищ баз данных» используются следующие оценочные средства:

1) Устный опрос (УО):

Собеседование (консультация с преподавателем) (УО-1)

2) Письменные работы (ПР):

Конспект (ПР-7)

Проект (ПР-9)

№ п/п	Контролируемые разделы / темы дисциплины	Код и наименование индикатора достижения	Результаты обучения	Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Большие данные, что это такое, принцип трех V	ПК-2.1, ПК-2.2 ПК-4.1, ПК-4.2	знает	УО-1, ПР-7, ПР-9	Вопросы к зачету
			умеет		
			владеет		
2	Ухищрения, применяемые для хранения больших данных	ПК-2.1, ПК-2.2 ПК-4.1, ПК-4.2	знает	УО-1, ПР-7, ПР-9	Вопросы к зачету
			умеет		
			владеет		
3	Замена требований ACID на BASE в случае с хранением больших данных	ПК-2.1, ПК-2.2 ПК-4.1, ПК-4.2	знает	УО-1, ПР-7, ПР-9	Вопросы к зачету
			умеет		
			владеет		
4	Индексы, виды индексов, наиболее часто применимые в области больших данных	ПК-2.1, ПК-2.2 ПК-4.1, ПК-4.2	знает	УО-1, ПР-7, ПР-9	Вопросы к зачету
			умеет		
			владеет		
5	Инвертированные индексы, полнотекстовые поисковые системы, первая версия инвертированного индекса от Google, принцип бареллей для хранения больших данных	ПК-2.1, ПК-2.2 ПК-4.1, ПК-4.2	знает	УО-1, ПР-7, ПР-9	Вопросы к зачету

Текущая аттестация студентов

Текущая аттестация студентов по дисциплине «Администрирование хранилищ больших данных» проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

Текущая аттестация по дисциплине «Администрирование хранилищ больших данных» проводится в форме контрольных мероприятий (устного опроса, выступления с проектом, тестирования, конспекта первоисточника) по оцениванию фактических результатов обучения студентов и осуществляется ведущим преподавателем.

Объектами оценивания выступают:

- учебная дисциплина (активность на занятиях, своевременность выполнения различных видов заданий, посещаемость всех видов занятий по аттестуемой дисциплине);
- степень усвоения теоретических знаний;
- уровень овладения практическими умениями и навыками по всем видам учебной работы;
- результаты самостоятельной работы.

Каждому объекту оценивания присваивается конкретный балл. Составляется календарный план контрольных мероприятий по дисциплине и внесения данных в АРС. По окончании семестра студент набирает определенное количество баллов, которые переводятся в пятибалльную систему оценки.

Критерии оценки устного опроса:

Результат работы студента в ходе устного ответа в виде собеседования с преподавателем оценивается по следующим критериям: полнота раскрытия вопросов; степень самостоятельности выполнения задания; и его презентация; исполнение сроков предоставления выполненных заданий; способность отвечать на вопросы преподавателя и студентов в ходе устного опроса по заданной тематике.

✓ 100-86 баллов выставляется, если студент выразил своё мнение по сформулированной проблеме, аргументировал его, точно определив ее содержание и составляющие. Продемонстрировано знание и владение навыком самостоятельной исследовательской работы по теме вопроса. Фактических ошибок, связанных с пониманием проблемы, нет

✓ 85-76 баллов работа студента характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Для аргументации приводятся данные отечественных и зарубежных авторов.

Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет.

✓ 75-61 баллов проведен достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимание базовых основ и теоретического обоснования выбранной темы. Привлечены основные источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы

✓ 60-50 баллов если работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая составляющая темы. Допущено три или более трех ошибок смыслового содержания раскрываемой проблемы.

Критерии оценки конспекта:

✓ 100-85 баллов - выставляется студенту, если студент выразил своё мнение по сформулированной проблеме, аргументировал его, точно определив ее содержание и составляющие. Приведены данные отечественной и зарубежной литературы, статистические сведения, информация нормативно-правового характера. Студент знает и владеет навыком самостоятельной исследовательской работы по теме исследования; методами и приемами анализа теоретических и/или практических аспектов изучаемой области. Фактических ошибок, связанных с пониманием проблемы, нет; графически работа оформлена правильно

✓ 85-76 баллов - работа характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Для аргументации приводятся данные отечественных и зарубежных авторов. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет. Допущены одна-две ошибки в оформлении работы

✓ 75-61 балл – студент проводит достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимает базовые основы и теоретическое обоснование выбранной темы. Привлечены основные источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы, оформлении работы

✓ 60-50 баллов – если работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая

составляющая темы. Допущено три или более трех ошибок в смысловом содержании раскрываемой проблемы, в оформлении работы.

Оценочные средства для промежуточной аттестации

Промежуточная аттестация студентов по дисциплине «Администрирование хранилищ больших данных» проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной. Согласно учебному плану видом промежуточной аттестации по дисциплине «Администрирование хранилищ больших данных» предусмотрен зачет с оценкой, который выставляется по результатам работы в семестре. Зачет с оценкой проводится в устной форме по вопросам.

Вопросы к зачету

1. Какие принципы лежат в основе хранения и обработки больших данных? Какие существуют средства для хранения и обработки больших данных?
2. Что такое распределенное хранение данных и как оно используется в системах хранения больших данных? В чем отличие шардирования от ранжирования?
3. Что такое теорема о невозможности соблюдения ACID принципов в распределенных системах? Как замена требований ACID на BASE позволяет повысить производительность?
4. Какие виды индексов существуют в области больших данных? Какие из них являются наиболее часто применяемыми?
5. Что такое инвертированные индексы и полнотекстовые поисковые системы? Какой принцип используется для хранения больших данных в бареллеях? Какую первую версию инвертированного индекса разработала компания Google?
6. Что такое NoSQL СУБД и в чем их преимущества перед реляционными СУБД при работе с большими данными? Какие типы NoSQL СУБД существуют?
7. Какие алгоритмы используются для обработки больших данных? Как работают алгоритмы MapReduce и Spark?
8. Что такое машинное обучение и как оно используется в области больших данных? Какие методы машинного обучения наиболее часто применяются?
9. Что такое хранилище данных Hadoop и как оно используется для работы с большими данными? Какие компоненты входят в состав Hadoop?
10. Какие проблемы возникают при работе с большими данными? Какие методы используются для решения этих проблем?

**Критерии выставления оценки студенту на зачете
по дисциплине «Администрирование хранилищ баз данных»:**

Баллы (рейтинговой оценки)	Оценка экзамена (стандартная)	Требования к сформированным компетенциям
100-85	<i>«отлично»</i>	Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, обозначает особенности применения тех или иных методов работы в зависимости от ситуаций, свободно справляется с задачами, вопросами и другими видами применения знаний, владеет разносторонними навыками и приемами выполнения практических задач, способен использовать современные технические средства для оптимизации, унификации и модернизации работы.
85-76	<i>«хорошо»</i>	Оценка «хорошо» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения. Использует основной терминологический, -правовые акты, влияющие на способы и методы работы.
75-61	<i>«удовлетворительно»</i>	Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при выполнении практических работ, не владеет способами и методами работы не применяет их.