



КГЭУ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГЭУ»)

УТВЕРЖДАЮ

Директор Института цифровых
технологий и экономики

_____ Э.И. Беляев

«29» ноября 2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.09 Подсистемы хранения и извлечения данных

(Код и наименование дисциплины в соответствии с РУП)

Направление подготовки _____ **09.03.03 Прикладная информатика** _____
(Код и наименование направления подготовки)

Направленность(и)
(профиль(и)) _____ **Прикладной искусственный интеллект** _____
(Наименование направленности (профиля) образовательной программы)

Квалификация _____ **Бакалавр** _____
(Бакалавр / Магистр)

г. Казань, 2023

Программу разработал(и):

Наименование кафедры	Должность, уч.степень, уч.звание	ФИО разработчика
УРФУ кафедра информационных технологий и систем управления	без ученой степени, без ученого звания, Ассистент	Аксёнов Александр Сергеевич
УРФУ информационных технологий и систем управления	к.т.н., Доцент	Ронкин Михаил Владимирович
УРФУ информационных технологий и систем управления	к.т.н., Доцент	Созыкин Андрей Владимирович
КГЭУ Кафедра ИТИС	Доцент, к.т.н, доцент	Хамитов Р.М.

Согласование	Наименование подразделения	Дата	№ протокола	Подпись
Одобрена	кафедра ИТИС	27.11.2023	11	_____ И.о. зав.каф.,к.ф.-м.н., доц. Соловьев С. А.
Согласована	Учебно-методический совет института ИЦТЭ	27.11.2023	3	_____ Директор, к.т.н., доц. Беляев Э.И.
Одобрена	Ученый совет института ИЦТЭ	28.11.2023	3	_____ Директор, к.т.н., доц. Беляев Э.И.

1. Цель, задачи и планируемые результаты обучения по дисциплине

(Цель и задачи освоения дисциплины, соответствующие цели ОП)

Целью освоения дисциплины Подсистемы хранения и извлечения данных является формирование системного мышления в сфере хранения данных как записей некоторой системы – базы данных. Студентам на уровне алгоритмов и структур данных раскрываются особенности работы стоимостных оптимизаторов СУБД различной природы. В результате изучения модуля студенты способны объяснить и реализовывать на языке программирования индексы (хэш, балансированные деревья b-tree, LSM-tree), а также базовые СУБД типа ключ-значение и алгоритмы объединения двух множеств по предикату, такие как вложенные циклы, сортированное слияние, объединение по хэшу.

Компетенции и индикаторы, формируемые у обучающихся:

Код и наименование компетенции	Код и наименование индикатора
ПК-7 Способен разрабатывать системы анализа больших данных	ПК-7.1 Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных ПК-7.2 Разрабатывает программные компоненты обработки, удаленной, распределенной и объединенной аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших

2. Место дисциплины в структуре ОП

Предшествующие дисциплины (модули), практики, НИР, др. Информационные технологии, Программирование, Компьютерные сети, Информационная безопасность, Объектно-ориентированное программирование, Базы данных, Проектирование информационных систем, Сбор и верификация данных.

Последующие дисциплины (модули), практики, НИР, др. Рекомендательные системы, Прикладные и наукоемкие задачи искусственного интеллекта, Инструменты решения задач искусственного интеллекта, Платформы бизнес-приложений и языки программирования, Платформы бизнес-анализа, Приложения искусственного интеллекта, Производственная практика (преддипломная), Выполнение и защита выпускной квалификационной работы.

3. Структура и содержание дисциплины

3.1. Структура дисциплины

Для очной формы обучения

Вид учебной работы	Всего ЗЕ	Всего часов	Семестр(ы)
			7
ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ	3	108	108
КОНТАКТНАЯ РАБОТА*	-	61	61
АУДИТОРНАЯ РАБОТА	1,39	50	50
Лекции	0,5	18	18
Практические (семинарские) занятия	-	-	-
Лабораторные работы	0,89	32	32
САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ	1,61	58	58
Проработка учебного материала	0,61	22	22
Курсовой проект	-	-	-
Курсовая работа	-	-	-
Подготовка к промежуточной аттестации	1	36	36
Промежуточная аттестация:			Э
			-

3.2. Содержание дисциплины, структурированное по разделам и видам занятий

Разделы дисциплины	Всего часов	Распределение трудоемкости по видам учебной работы				Формы и вид контроля	Индексы индикаторов формируемых компетенций
		лекции	лаб. раб.	пр. зан.	сам. раб.		
Раздел 1. Алгоритмы и структуры данных.	18	6	10		7	ТК1	ПК-7.3, ПК-7.У
Раздел 2. Хранение и извлечение данных.	18	6	12		8	ТК2 ТК3	ПК-7.3, ПК-7.У
Раздел 3. Ускорение доступа к данным.	18	6	10		7		ПК-7.3, ПК-7.У
Экзамен	36				36	ОМ	ПК-7.3, ПК-7.У
Итого за 7 семестр	108	18	32		58		
ИТОГО	108	18	32		58		

3.3. Содержание дисциплины

Раздел 1. Алгоритмы и структуры данных.

1.1 Введение в алгоритмы и структуры данных. Использование массивов и связанных списков

Определение и обсуждение алгоритмов и структур данных. Понятие сложности алгоритмов и его значимость. Массивы: основные характеристики, преимущества и недостатки. Связные списки: основные характеристики, преимущества и недостатки. Обзор и сравнение

использования массивов и связанных списков

1.2 Использование хэш-таблиц. Алгоритм и примеры

Определение хэш-таблицы. Обзор основных концепций хэширования. Основы хэш-функций. Свойства хороших хэш-функций. Работа хэш-таблиц. Коллизии и способы их устранения. Предложения использования хэш-таблиц

1.3 Введение в балансированные деревья. Что такое B-деревья и как они работают.

Введение в балансированные деревья. Определение и основные свойства B-деревьев. Принципы работы B-деревьев. Использование B-деревьев на практике

1.4 Введение в LSM деревья. Что такое LSM-деревья и как они работают

Введение в структуры данных. LSM-деревья: определение и основные свойства. Принцип работы LSM-деревьев.

Раздел 2. Хранение и извлечение данных.

2.1 Введение в хранение и извлечение данных

Типы баз данных: реляционные, NoSQL, In-memory, текстовые. Структура баз данных. Хранение и извлечение данных. Использование баз данных в анализе данных, для веб- и мобильных приложений.

2.2 Базовые СУБД типа ключ-значение. Как с ними работать, примеры применения

Введение в СУБД типа ключ-значение. Принципы работы. Обзор популярных СУБД типа ключ-значение. Примеры применения в реальных задачах.

2.3 Способы объединения двух множеств по предикату указанными способами (вложенные циклы, сортированное слияние, объединение по хэшу)

Определение понятий "множество", "предикат" и "способ объединения". Вложенные циклы. Сортированное слияние. Объединение по хэшу. Преимущества и ограничения каждого метода объединения по хэшу

Раздел 3. Ускорение доступа к данным.

3.1 Понятие и виды индексов в базах данных

Введение в индексы в базах данных. Виды индексов в базах данных. Использование индексов. Примеры работы с индексами.

3.2 Стоимостные оптимизаторы в СУБД

Введение в оптимизацию запросов. Понятие стоимостных оптимизаторов. Принципы работы стоимостных оптимизаторов. Примеры работы стоимостных оптимизаторов. Анализ влияния оптимизации на производительность СУБД

3.4. Тематический план практических занятий

«Данный вид работы не предусмотрен учебным планом».

3.5. Тематический план лабораторных работ

Раздел 1. Алгоритмы и структуры данных.

1. Создание и использование массивов и связанных списков в python
2. Создание и использование хэш-таблиц в python
3. Работа с балансированными деревьями в python
4. Работа с LSM деревьями в python

Раздел 2. Хранение и извлечение данных.

5. Знакомство с базами данных в python
6. Работа с базовыми СУБД типа ключ-значение в python
7. Реализация алгоритмов слияния в python

Раздел 3. Ускорение доступа к данным.

8. Работа с индексами в python
9. Оптимизация запросов

3.6. Курсовой проект /курсовая работа

«Данный вид работы не предусмотрен учебным планом».

4. Оценивание результатов обучения

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля и промежуточной аттестации, проводимых по балльно-рейтинговой системе (БРС).

Шкала оценки результатов обучения по дисциплине:

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности индикатора компетенции			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено			не зачтено
ПК-7	7.1 Разрабатывает программные компоненты извлечения, хранения, подготовки	знать:				
		общедоступные репозитории и специализированные библиотеки, содержащие наборы	Уровень знаний в объеме, соответствует в программе подготовки, без	Уровень знаний в объеме, соответствует в программе, имеет место несколько	Минимально допустимый уровень знаний, имеет место	Уровень знаний ниже минимальных требований, имеют место грубые

<p>больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных</p>	<p>больших данных</p>	<p>ошибок</p>	<p>негрубых ошибок</p>	<p>много негрубых ошибок</p>	<p>ошибки</p>	
	<p>принципы работы экосистемы Hadoop, фреймворка SPARK</p>	<p>Уровень знаний в объеме, соответствующем программе подготовки, без ошибок</p>	<p>Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок</p>	<p>Минимально допустимый уровень знаний, имеет место много негрубых ошибок</p>	<p>Уровень знаний ниже минимальных требований, имеют место грубые ошибки</p>	
	<p>устройство интерфейсов между реляционным и SQL-хранилищами данных и нереляционными NoSQL-хранилищами данных</p>	<p>Уровень знаний в объеме, соответствующем программе подготовки, без ошибок</p>	<p>Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок</p>	<p>Минимально допустимый уровень знаний, имеет место много негрубых ошибок</p>	<p>Уровень знаний ниже минимальных требований, имеют место грубые ошибки</p>	
	<p>предметно-ориентированные языки</p>	<p>Уровень знаний в объеме, соответствующем программе подготовки, без ошибок</p>	<p>Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок</p>	<p>Минимально допустимый уровень знаний, имеет место много негрубых ошибок</p>	<p>Уровень знаний ниже минимальных требований, имеют место грубые ошибки</p>	
	<p>уметь:</p>					
	<p>настраивать и оптимизировать конфигурацию программного и аппаратного обеспечения с целью интеграции больших данных</p>	<p>Сформированы умения</p>	<p>Частично сформированы умения</p>	<p>Сформированы базовые умения</p>	<p>Не сформированы базовые умения</p>	
	<p>разрабатывать программное обеспечение</p>	<p>Сформированы умения</p>	<p>Частично сформированы умения</p>	<p>Сформированы базовые умения</p>	<p>Не сформированы умения</p>	

		для очистки и валидации наборов больших данных		умения	умения	базовые умения
		выполнять потоковую обработку данных (data streaming, event processing)	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		использовать шины данных (Apache Kafka)	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		использовать языки запросов, в том числе нереляционных, для поддержки различных типов данных (например, XML, RDF, JSON, мультимедиа) и операций с большими данными (например, матричные операции)	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
	ПК-7.2	знать:				
	Разрабатывает программные компоненты обработки, удаленной, распределенной и объединенной аналитики, использования результатов анализа,	принципы и методы анализа больших данных, включая спецификации и стандартизацию метаданных	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
		устройство и принципы работы систем обработки и	Уровень знаний в объеме, соответствует	Уровень знаний в объеме, соответствует	Минимально допустимый	Уровень знаний ниже минималь

описания и управления качеством и достоверностью больших данных	анализа больших массивов данных (SQL, NoSQL, Hadoop, ETL)	вующем программ е подготовк и, без ошибок	вующем программ е, имеет место несколько негрубых ошибок	уровень знаний, имеет место много негрубых ошибок	ных требовани й, имеют место грубые ошибки
	архитектуру и принципы работы промышленны х решений, созданных на основе искусственног о интеллекта	Уровень знаний в объеме, соответст вующем программ е подготовк и, без ошибок	Уровень знаний в объеме, соответст вующем программ е, имеет место несколько негрубых ошибок	Минимал ьно допустим ый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минималь ных требовани й, имеют место грубые ошибки
	методы и технологии машинного обучения на больших данных	Уровень знаний в объеме, соответст вующем программ е подготовк и, без ошибок	Уровень знаний в объеме, соответст вующем программ е, имеет место несколько негрубых ошибок	Минимал ьно допустим ый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минималь ных требовани й, имеют место грубые ошибки
	уметь:				
	разрабатывать программное обеспечение для анализа больших данных	Сформир ованы умения	Частично сформиро ваны умения	Сформир ованы базовые умения	Не сформиро ваны базовые умения
	разрабатывать программные и технические средства визуализации больших данных и результатов их анализа	Сформир ованы умения	Частично сформиро ваны умения	Сформир ованы базовые умения	Не сформиро ваны базовые умения
	Умеет использовать системы обработки и анализа больших массивов данных (SQL,	Сформир ованы умения	Частично сформиро ваны умения	Сформир ованы базовые умения	Не сформиро ваны базовые умения

		NoSQL, Hadoop, ETL, процессы и инструменты)				
		использовать технологии науки о данных и больших данных в разработке для решения практических задач промышленности	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		описывать и управлять качеством и достоверностью больших данных	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения

Оценочные материалы для проведения текущего контроля и промежуточной аттестации приведены в Приложении к рабочей программе дисциплины.

Полный комплект заданий и материалов, необходимых для оценивания результатов обучения по дисциплине, хранится на кафедре разработчика.

5. Учебно-методическое и информационное обеспечение дисциплины

5.1. Учебно-методическое обеспечение

5.1.1. Основная литература

1. Павлов, Л. А. Структуры и алгоритмы обработки данных : учебник для вузов / Л. А. Павлов, Н. В. Первова. — 3-е изд., стер. — Санкт-Петербург : Лань, 2021. — 256 с. — ISBN 978-5-8114-7259-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/156929>
2. Мясникова, Н. А., Алгоритмы и структуры данных : учебное пособие / Н. А. Мясникова. — Москва : КноРус, 2023. — 185 с. — ISBN 978-5-406-10688-4. — URL: <https://book.ru/book/946265> — Текст : электронный.
3. Тюкачев, Н. А. С#. Алгоритмы и структуры данных / Н. А. Тюкачев, В. Г. Хлебостроев. — 6-е изд., стер. — Санкт-Петербург : Лань, 2023. — 232 с. — ISBN 978-5-507-47248-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/346067>
4. Никитина, Т. П. Программирование. Основы Python для инженеров / Т. П. Никитина, Л. В. Королев. — Санкт-Петербург : Лань, 2023. — 156 с. — ISBN 978-5-507-45284-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/302720>
5. Наместников, А. М. Базы данных. Практический курс : учебное пособие : в 2 частях / А. М. Наместников. — Ульяновск : УЛГТУ, 2017 — Часть 1 : Объектно-реляционные базы данных на примере PostgreSQL 9.5 — 2017. — 113 с. — ISBN 978-5-9795-1743-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165100>.
6. Дадян, Эдуард Григорьевич. Данные: хранение и обработка : учебник / Э. Г. Дадян. - Москва : ИНФРА-М, 2021. - 205 с. - URL: <http://new.ibooks.ru/bookshelf/377999>. - ISBN 978-5-16-016447-2 (print). - ISBN 978-5-16-107936-2 (online). - Текст : электронный.
7. Волк, В. К. Базы данных. Проектирование, программирование, управление и администрирование / В. К. Волк. — 4-е изд., стер. — Санкт-Петербург : Лань, 2023. — 244 с. — ISBN 978-5-507-47243-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/346439>
8. Тарасов С.В. СУБД для программиста. Базы данных изнутри / С.В. Тарасов. - Москва : СОЛОН-ПРЕСС, 2018. - 320 с. - ISBN 978-2-7466-7383-0. - URL: <https://ibooks.ru/bookshelf/344900/reading>. - Текст: электронный.
9. Цехановский, В. В. Управление данными : учебник / В. В. Цехановский, В. Д. Чертовской. — Санкт-Петербург : Лань, 2022. — 432 с. — ISBN 978-5-8114-1853-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/212084>

5.1.2. Дополнительная литература

1. Масленникова О.Е. Основы искусственного интеллекта. Учебное пособие / О.Е. Масленникова, И.В. Гаврилова. - Москва : Флинта, 2019. - 283 с. - ISBN 978-5-9765-1602-1. - URL: <https://ibooks.ru/bookshelf/337972>
2. Остроух, А. В. Системы искусственного интеллекта : монография / А. В. Остроух, Н. Е. Суркова. — 3-е изд., стер. — Санкт-Петербург : Лань, 2023. — 228 с. — ISBN 978-5-507-46441-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/310199>

5.2. Информационное обеспечение

5.2.1. Электронные и интернет-ресурсы

№ п/п	Наименование электронных и интернет-ресурсов	Ссылка
1	Электронно-библиотечная система «Лань»	https://e.lanbook.com/
2	Электронно-библиотечная система «ibooks.ru»	https://ibooks.ru/
3	Электронно-библиотечная система «book.ru»	https://www.book.ru/
4	Портал «Открытое образование»	https://npoed.ru
5	Российская национальная библиотека	https://nlr.ru/
6	КиберЛенинка	https://cyberleninka.ru
7	Техническая библиотека	https://techlibrary.ru
8	Национальная электронная библиотека (НЭБ)	https://rusneb.ru/
9	Веб-сервис для хостинга IT-проектов и их совместной разработки Github	http://www.github.ru

5.2.2. Профессиональные базы данных / Информационно-справочные системы

№ п/п	Наименование профессиональных баз данных	Адрес	Режим доступа
1	Официальный интернет-портал правовой информации	http://pravo.gov.ru	http://pravo.gov.ru
2	Справочная правовая система «Консультант Плюс»	http://consultant.ru	http://consultant.ru
3	Справочно-правовая система по законодательству РФ	http://garant.ru	http://garant.ru

5.2.3. Лицензионное и свободно распространяемое программное обеспечение дисциплины

№ п/п	Наименование программного обеспечения	Описание	Реквизиты подтверждающих документов
1	Операционная система Microsoft Windows 10	Пользовательская операционная система	Договор №133/2021 от 12.10.2021, лицензиар - ЗАО «Софт Лайн Трейд», тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно
2	Microsoft Office 2019	Пакет офисных приложений	Договор №133/2021 от 12.10.2021, лицензиар - ЗАО «Софт Лайн Трейд», тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно

3	LMS Moodle	ПО для эффективного онлайн- взаимодействия преподавателя и студента	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
4	Браузер Chrome	Система поиска информации в сети интернет	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
5	СУБД PostgreSQL	Система управления базами данных https://www.postgresql.org/	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
6	pgAdmin для PostgreSQL	Система администрирования и разработки pgAdmin для PostgreSQL https://www.pgadmin.org/	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
7	DBeaver Community	Free Universal Database Tool https://dbeaver.io/	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.

6. Материально-техническое обеспечение дисциплины

Наименование вида учебной работы	Наименование учебной аудитории, специализированной лаборатории	Перечень необходимого оборудования и технических средств обучения
Лекции	Учебная аудитория для проведения занятий лекционного типа	Специализированная учебная мебель, технические средства обучения, служащие для представления учебной информации большой аудитории (мультимедийный проектор, компьютер (ноутбук), экран), демонстрационное оборудование, учебно-наглядные пособия
Лабораторные работы	Учебная лаборатория программной инженерии, ауд. В-608	Специализированное лабораторное оборудование по профилю лаборатории программной инженерии, специализированная учебная мебель на 50 посадочных мест, 24 компьютера с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-610	Специализированная учебная мебель на 42 посадочных места, 17 компьютеров с

		возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, экран для проектора, моноблок), необходимое лицензионное программное обеспечение
	Учебная лаборатория информационной безопасности, ауд. В-615	Специализированное лабораторное оборудование по профилю лаборатории информационной безопасности, специализированная учебная мебель на 35 посадочных мест, 15 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-617	Специализированная учебная мебель на 24 посадочных места, 21 компьютер с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-619	Специализированная учебная мебель на 26 посадочных мест, 21 компьютер с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-621	Специализированная учебная мебель на 35 посадочных мест, 13 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, экран для проектора, моноблок), необходимое лицензионное программное обеспечение
	Учебная лаборатория реинжиниринга и управления бизнес-процессами, ауд. В-623	Специализированное лабораторное оборудование по профилю лаборатории реинжиниринга и управления бизнес-процессами, специализированная учебная мебель на 34 посадочных места, 13 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, В-600	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения

		(мультимедийный проектор, ноутбук, экран), видеорекамеры, необходимое лицензионное программное обеспечение
Самостоятельная работа	Компьютерный класс с возможностью выхода в Интернет и обеспечением доступа в ЭИОС В-600	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, технические средства обучения (мультимедийный проектор, компьютер (ноутбук), экран), видеорекамеры, программное обеспечение
	Читальный зал библиотеки	Специализированная мебель, компьютерная техника с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, экран, мультимедийный проектор, программное обеспечение

7. Особенности организации образовательной деятельности для лиц с ограниченными возможностями здоровья и инвалидов

Лица с ограниченными возможностями здоровья (ОВЗ) и инвалиды имеют возможность беспрепятственно перемещаться из одного учебно-лабораторного корпуса в другой, подняться на все этажи учебно-лабораторных корпусов, заниматься в учебных и иных помещениях с учетом особенностей психофизического развития и состояния здоровья.

Для обучения лиц с ОВЗ и инвалидов, имеющих нарушения опорно-двигательного аппарата, обеспечены условия беспрепятственного доступа во все учебные помещения. Информация о специальных условиях, созданных для обучающихся с ОВЗ и инвалидов, размещена на сайте университета www/kgeu.ru. Имеется возможность оказания технической помощи ассистентом, а также услуг сурдопереводчиков и тифлосурдопереводчиков.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушенным слухом справочного, учебного материала по дисциплине обеспечиваются следующие условия:

- для лучшей ориентации в аудитории, применяются сигналы оповещения о начале и конце занятия (слово «звонок» пишется на доске);
- внимание слабослышащего обучающегося привлекается педагогом жестом (на плечо кладется рука, осуществляется нерезкое похлопывание);
- разговаривая с обучающимся, педагогический работник смотрит на него, говорит ясно, короткими предложениями, обеспечивая возможность чтения по губам.

Компенсация затруднений речевого и интеллектуального развития слабослышащих обучающихся проводится путем:

- использования схем, диаграмм, рисунков, компьютерных презентаций с гиперссылками, комментирующими отдельные компоненты изображения;
- регулярного применения упражнений на графическое выделение существенных признаков предметов и явлений;
- обеспечения возможности для обучающегося получить адресную

консультацию по электронной почте по мере необходимости.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушениями зрения справочного, учебного, просветительского материала, предусмотренного образовательной программой по выбранному направлению подготовки, обеспечиваются следующие условия:

- ведется адаптация официального сайта в сети Интернет с учетом особых потребностей инвалидов по зрению, обеспечивается наличие крупношрифтовой справочной информации о расписании учебных занятий;

- педагогический работник, его собеседник (при необходимости), присутствующие на занятии, представляются обучающимся, при этом каждый раз называется тот, к кому педагогический работник обращается;

- действия, жесты, перемещения педагогического работника коротко и ясно комментируются;

- печатная информация предоставляется крупным шрифтом (от 18 пунктов), тотально озвучивается;

- обеспечивается необходимый уровень освещенности помещений;

- предоставляется возможность использовать компьютеры во время занятий и право записи объяснений на диктофон (по желанию обучающихся).

Форма проведения текущей и промежуточной аттестации для обучающихся с ОВЗ и инвалидов определяется педагогическим работником в соответствии с учебным планом. При необходимости обучающемуся с ОВЗ, инвалиду с учетом их индивидуальных психофизических особенностей дается возможность пройти промежуточную аттестацию устно, письменно на бумаге, письменно на компьютере, в форме тестирования и т.п., либо предоставляется дополнительное время для подготовки ответа.

8. Методические рекомендации для преподавателей по организации воспитательной работы с обучающимися.

Методическое обеспечение процесса воспитания обучающихся выступает одним из определяющих факторов высокого качества образования. Преподаватель вуза, демонстрируя высокий профессионализм, эрудицию, четкую гражданскую позицию, самодисциплину, творческий подход в решении профессиональных задач, в ходе образовательного процесса способствует формированию гармоничной личности.

При реализации дисциплины преподаватель может использовать следующие методы воспитательной работы:

- методы формирования сознания личности (беседа, диспут, внушение, инструктаж, контроль, объяснение, пример, самоконтроль, рассказ, совет, убеждение и др.);

- методы организации деятельности и формирования опыта поведения (задание, общественное мнение, педагогическое требование, поручение, приучение, создание воспитывающих ситуаций, тренинг, упражнение, и др.);

- методы мотивации деятельности и поведения (одобрение, поощрение социальной активности, порицание, создание ситуаций успеха, создание ситуаций для эмоционально-нравственных переживаний, соревнование и др.)

При реализации дисциплины преподаватель должен учитывать следующие направления воспитательной деятельности:

Гражданское и патриотическое воспитание:

- формирование у обучающихся целостного мировоззрения, российской идентичности, уважения к своей семье, обществу, государству, принятым в семье и обществе духовно-нравственным и социокультурным ценностям, к национальному, культурному и историческому наследию, формирование стремления к его сохранению и развитию;

- формирование у обучающихся активной гражданской позиции, основанной на традиционных культурных, духовных и нравственных ценностях российского общества, для повышения способности ответственно реализовывать свои конституционные права и обязанности;

- развитие правовой и политической культуры обучающихся, расширение конструктивного участия в принятии решений, затрагивающих их права и интересы, в том числе в различных формах самоорганизации, самоуправления, общественно-значимой деятельности;

- формирование мотивов, нравственных и смысловых установок личности, позволяющих противостоять экстремизму, ксенофобии, дискриминации по социальным, религиозным, расовым, национальным признакам, межэтнической и межконфессиональной нетерпимости, другим негативным социальным явлениям.

Духовно-нравственное воспитание:

- воспитание чувства достоинства, чести и честности, совестливости, уважения к родителям, учителям, людям старшего поколения;

- формирование принципов коллективизма и солидарности, духа милосердия и сострадания, привычки заботиться о людях, находящихся в трудной жизненной ситуации;

- формирование солидарности и чувства социальной ответственности по отношению к людям с ограниченными возможностями здоровья, преодоление психологических барьеров по отношению к людям с ограниченными возможностями;

- формирование эмоционально насыщенного и духовно возвышенного отношения к миру, способности и умения передавать другим свой эстетический опыт.

Культурно-просветительское воспитание:

- формирование эстетической картины мира;

- формирование уважения к культурным ценностям родного города, края, страны;

- повышение познавательной активности обучающихся.

Научно-образовательное воспитание:

- формирование у обучающихся научного мировоззрения;

- формирование умения получать знания;

- формирование навыков анализа и синтеза информации, в том числе в профессиональной области.

Вносимые изменения и утверждения на новый учебный год

№ п/п	№ раздела внесения изменений	Дата внесения изменений	Содержание изменений	«Согласовано» Зав. каф. реализующей дисциплину	«Согласовано» председатель УМК института (факультета), в состав которого входит выпускающая
1	2	3	4	5	6
1					
2					
3					



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГУУ»)

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
по дисциплине

Б1.В.09 Подсистемы хранения и извлечения данных

(Наименование дисциплины в соответствии с учебным планом)

Направление подготовки 09.03.03 Прикладная информатика
(Код и наименование направления подготовки)

Направленность(и)
) (профиль(и)) Прикладной искусственный интеллект

Квалификация Бакалавр

(Бакалавр / Магистр)

2. Оценочные материалы текущего контроля и промежуточной аттестации

Шкала оценки результатов обучения по дисциплине:

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности индикатора компетенции			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено		не зачтено	
ПК-7	7.1 Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных	знать:				
		общедоступные репозитории и специализированные библиотеки, содержащие наборы больших данных	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
		принципы работы экосистемы Hadoop, фреймворка SPARK	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
		устройство интерфейсов между реляционным и SQL-хранилищами данных и нереляционными NoSQL-хранилищами данных	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
		предметно-ориентированные языки	Уровень знаний в объеме, соответствует	Уровень знаний в объеме, соответствует	Минимально допустимый	Уровень знаний ниже минимальных

			вующем программ е подготовк и, без ошибок	вующем программ е, имеет место несколько негрубых ошибок	уровень знаний, имеет место много негрубых ошибок	ных требовани й, имеют место грубые ошибки
		уметь:				
		настраивать и оптимизировать конфигурацию программного и аппаратного обеспечения с целью интеграции больших данных	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		разрабатывать программное обеспечение для очистки и валидации наборов больших данных	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		выполнять потоковую обработку данных (data streaming, event processing)	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		использовать шины данных (Apache Kafka)	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		использовать языки запросов, в том числе нереляционных, для поддержки различных типов данных (например, XML, RDF, JSON,	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения

		мультимедиа) и операций с большими данными (например, матричные операции)				
		знать:				
	ПК-7.2 Разрабатывает программные компоненты обработки, удаленной, распределенной и объединенной аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных	принципы и методы анализа больших данных, включая спецификации и стандартизацию метаданных	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки
устройство и принципы работы систем обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, ETL)		Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки	
архитектуру и принципы работы промышленных решений, созданных на основе искусственного интеллекта		Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки	
методы и технологии машинного обучения на больших данных		Уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок	Минимально допустимый уровень знаний, имеет место много негрубых ошибок	Уровень знаний ниже минимальных требований, имеют место грубые ошибки	

		уметь:				
		разрабатывать программное обеспечение для анализа больших данных	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		разрабатывать программные и технические средства визуализации больших данных и результатов их анализа	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		Умеет использовать системы обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, ETL, процессы и инструменты)	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		использовать технологии науки о данных и больших данных в разработке для решения практических задач промышленности	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения
		описывать и управлять качеством и достоверностью больших данных	Сформированы умения	Частично сформированы умения	Сформированы базовые умения	Не сформированы базовые умения

Оценка **«отлично»** выставляется за выполнение лабораторных работ в семестре; тестовых заданий; глубокое понимание методов сбора и обобщения информации о проблемной области путем опроса экспертов, исходных данных о функционировании проблемной и предметной областей, документированных источников знания, а также формирования требований к системе

искусственного интеллекта; знание инструментов и методов подготовки и разметки структурированных и неструктурированных данных

Оценка «хорошо» выставляется за выполнение лабораторных работ в семестре; тестовых заданий; понимание глубокое понимание методов сбора и обобщения информации о проблемной области путем опроса экспертов, исходных данных о функционировании проблемной и предметной областей, документированных источников знания, а также формирования требований к системе искусственного интеллекта; знание инструментов и методов подготовки и разметки структурированных и неструктурированных данных;

Оценка «удовлетворительно» выставляется за выполнение лабораторных работ в семестре и тестовых заданий;

Оценка «неудовлетворительно» выставляется за слабое и неполное выполнение лабораторных работ в семестре и тестовых заданий.

3. Перечень оценочных средств

Краткая характеристика оценочных средств, используемых при текущем контроле успеваемости и промежуточной аттестации обучающегося по дисциплине:

Наименование оценочного средства	Краткая характеристика оценочного средства	Описание оценочного средства
Контрольная работа (КнТР)	Средство проверки умений применять полученные знания для решения задач определенного типа по теме или разделу	Комплект контрольных заданий по вариантам
Отчет по лабораторной работе (ОЛР)	Выполнение лабораторной работы, обработка результатов испытаний, измерений, эксперимента. Оформление отчета, защита результатов лабораторной работы по отчету	Перечень заданий и вопросов для защиты лабораторной работы, перечень требований к отчету
Тест (Тест)	Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося	Комплект тестовых заданий

4. Перечень контрольных заданий или иные материалы, необходимые для оценки знаний, умений и навыков, характеризующих этапы формирования компетенций в процессе освоения дисциплины

Пример задания

Для текущего контроля ТК1:

Проверяемая компетенция: ПК-7. Способен разрабатывать системы анализа больших данных. ПК-7.1. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных. ПК-7.2. Разрабатывает программные

компоненты обработки, удаленной, распределенной и объединенной аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных.

Отчет по лабораторной работе

Данный вид контроля за учебной деятельностью осуществляется в течение семестра. Посещение лабораторных занятий и выполнения заданий является допуском к зачету (промежуточной аттестации) по дисциплине. Обучающийся не допускается к зачету (промежуточной аттестации), если не сданы отчеты (в виде разработанных тестовых наборов или программ) по всем учебным модулям, а также в случае недобора баллов согласно бально-рейтинговой системы (менее 55). Для каждого раздела предусмотрено от 1 до 3 лабораторных заданий. Каждому студенту выдается индивидуальное задание.

Текущий контроль проводится перед началом каждого лабораторного занятия. Обучающиеся предоставляют отчет по лабораторной работе в электронном виде в виртуальной образовательной среде Moodle.

Перечень примерных заданий:

Раздел 1. Алгоритмы и структуры данных.

Дополнительные типы данных:

1. Создание и использование массивов и связанных списков в python

Дополнительные структуры данных:

1. Изучите основные концепции массивов и связанных списков, включая различия между ними. Какие есть преимущества и недостатки каждой структуры данных?

2. Ознакомьтесь с основами работы с массивами в Python, используя встроенные типы данных, такие как список (list). Попрактикуйтесь в создании массивов, доступе по индексу, вставке и удалении элементов. Какие есть различия в списке и массиве?

3. Реализуйте собственный класс массива, чтобы лучше понять внутреннюю работу. Добавьте методы для вставки, удаления и доступа к элементам массива.

4. Изучите основы работы со связанными списками в Python, используя классы Node для представления узлов и создания списка. Попрактикуйтесь в вставке, удалении и поиске элементов в связанных списках.

5. Реализуйте свой собственный класс для двусвязного списка. Добавьте методы для вставки, удаления и доступа к элементам.

6. Сделайте вывод и напишите, в чём удобство изученных структур данных, для каких задач применимы эти структуры?

2. Создание и использование хэш-таблиц в python

Реализация хэш-таблиц:

1. Изучите базовые концепции хэш-таблиц, включая хэширование, коллизии и методы их разрешения (например, цепочки, открытое адресное хэширование). Какая роль хэш-функций в создании уникальных индексов для

данных?

2. Изучите встроенные структуры данных для хранения хэш-таблиц в Python, такие как словари (dict) и множества (set). Попрактикуйтесь в создании, добавлении, поиске и удалении элементов из словарей и множеств.

3. Создайте свою собственную простую хэш-таблицу, используя массив и функцию хэширования. Реализуйте методы вставки, поиска и удаления элементов. Обработайте коллизии, выбрав подходящий метод разрешения.

4. Изучите различные хэш-функции и их свойства. Какие сильные и слабые стороны различных хэш-функций? В чём важность выбора эффективной хэш-функции для предотвращения коллизий?

3. Работа с балансированными деревьями в python

Реализация балансированных деревьев.

1. Изучите основные концепции балансированных деревьев, таких как AVL-деревья. Какие принципы балансировки и вращений, обеспечивающих сбалансированность дерева?

2. Изучите понятие высоты дерева и его влияние на балансировку. Рассмотрите как высота дерева влияет на сложность операций вставки, удаления и поиска?

3. Изучите встроенные структуры данных в Python, такие как set и sortedcontainers, которые могут использовать балансированные деревья. Попрактикуйтесь в выполнении операций вставки, удаления и поиска с использованием встроенных средств.

4. Создайте свою собственную реализацию балансированного дерева в Python. Реализуйте операции вставки, удаления и поиска, обеспечивая при этом соблюдение баланса дерева. Какие вращения, необходимые для коррекции баланса?

4. Работа с LSM деревьями в python

Реализация LSM деревьев:

1. Изучите основные концепции LSM-деревьев (Log-Structured Merge Trees) и их предназначение для эффективного управления большими объемами данных. Понимание структуры LSM-дерева, включая уровни сортированных файлов и журнал операций.

2. Изучите возможности работы с файлами и бинарными данными в Python. Понимание базовых операций записи, чтения и слияния файлов.

3. Создайте базовую реализацию LSM-дерева в Python. Реализуйте методы вставки, поиска и слияния для уровней дерева. Рассмотрите, какие компоненты потребуются для хранения данных и управления операциями.

4. Изучите методы оптимизации LSM-деревьев, такие как компактизация и управление уровнями. Какова важность выбора параметров и настройки LSM-дерева для конкретных требований?

Критерием оценивания результатов является правильность выполнения

задания согласно заявленным требованиям

При оценке выполненной лабораторной работы учитываются следующие критерии:

Отчет составлен, верно, даны подробные описания определений и понятий, верно приведены примеры – 5 баллов;

Отчет выполнен в целом верно, имеются незначительные ошибки при оформлении отчетов – 4 балла;

Отчет выполнен со значительным количеством ошибок, не соответствует заявленному стандарту, часто демонстрируются ошибки - 3 балла;

Отчет составлен неверно, большое количество ошибок при оформлении – 2 балла.

Максимальное количество баллов за каждое выполненное лабораторное задание – 5 баллов.

Максимальное количество баллов за все задания раздела – 20 баллов.

Для текущего контроля ТК2:

Проверяемая компетенция: ПК-7. Способен разрабатывать системы анализа больших данных. ПК-7.1. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных. ПК-7.2. Разрабатывает программные компоненты обработки, удаленной, распределенной и объединенной аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных.

Отчет по лабораторной работе

Данный вид контроля за учебной деятельностью осуществляется в течение семестра. Посещение лабораторных занятий и выполнения заданий является допуском к зачету (промежуточной аттестации) по дисциплине. Обучающийся не допускается к зачету (промежуточной аттестации), если не сданы отчеты (в виде разработанных тестовых наборов или программ) по всем учебным модулям, а также в случае недобора баллов согласно бально-рейтинговой системы (менее 55). Для каждого раздела предусмотрено от 1 до 3 лабораторных заданий. Каждому студенту выдается индивидуальное задание.

Текущий контроль проводится перед началом каждого лабораторного занятия. Обучающиеся предоставляют отчет по лабораторной работе в электронном виде в виртуальной образовательной среде Moodle.

Перечень примерных заданий:

Раздел 2. Хранение и извлечение данных.

3. Знакомство с базами данных в python

БД Python

1. Изучите основные понятия баз данных, такие как таблицы, строки, столбцы, SQL (Structured Query Language), индексы и т.д. Какие различия между реляционными и нереляционными базами данных?

2. Ознакомьтесь с основами SQL, языком запросов для работы с реляционными базами данных. Попрактикуйтесь в написании простых SQL-запросов для создания таблиц, вставки, обновления и выборки данных.

3. Изучите доступные библиотеки для взаимодействия с базами данных в Python, такие как SQLAlchemy, SQLite3, psycopg2 (для PostgreSQL), pymysql (для MySQL) и др. Выберите библиотеку, соответствующую вашим потребностям и типу базы данных.

4. Создайте базу данных для хранения данных пассажиров самолёта в SQLite, создайте таблицу и проведите основные операции (вставка, выборка, обновление, удаление).

4. Работа с базовыми СУБД типа ключ-значение в python

Разработка СУБД типа ключ-значение.

1. Изучите основные понятия баз данных типа ключ-значение, такие как ключ, значение, операции вставки, обновления и удаления. Какие преимущества и недостатки данного типа баз данных?

2. Рассмотрите различные базы данных типа ключ-значение для Python, такие как Redis, Berkeley DB, LevelDB и др. Выберите ту, которая соответствует вашим требованиям и предпочтениям.

3. Установите выбранную базу данных на своем компьютере. Познакомьтесь с основными настройками и параметрами конфигурации.

4. Разработайте СУБД типа ключ-значение, в которой будут записаны текста статей по теме поездов. Используйте уникальный ключ, связанный с запросом, который возвращает популярные статьи (например, "popular_articles_cache").

5. Перед выполнением запроса к базе данных проверьте наличие данных в кэше для данного ключа.

6. Если данные присутствуют в кэше, верните их без выполнения запроса к базе данных.

7. Если данные отсутствуют в кэше, выполните запрос к базе данных для получения списка популярных статей. Сохраните результаты запроса в кэше, связав их с уникальным ключом "popular_articles_cache".

8. Измерьте в миллисекундах а) обращение по ключу к не популярным статьям, б) обращение к популярной статье, но её не было в кеше, в) обращение к популярной статье, которая лежит в кэше. Сравните получившиеся измерения и сделайте вывод?

7. Реализация алгоритмов слияния в python.

Операторы слияния SQL

1. Изучите основы баз данных, включая реляционные и нереляционные модели. Познакомьтесь с языком SQL для работы с реляционными базами данных.

2. Подробно изучите алгоритмы слияния данных, такие как сортировка слиянием. Как алгоритмы слияния могут быть применены к базам данных для эффективного объединения данных из разных источников?

3. Разработайте базу данных пассажиров поезда и базу данных с информацией о личных данных пользователей (записей должно быть больше, чем значений пассажиров).

4. На разных примерах рассмотрите операции CROSS JOIN, [INNER] JOIN, [OUTER] JOIN, [OUTER] JOIN, FULL [OUTER] JOIN, UNION, INTERSECT, EXCEPT.

5. Разберите что делает каждый оператор, приведите пример, применяя данные о пассажирах и информации о них. Распишите для чего реализуется каждая операция.

Критерием оценивания результатов является правильность выполнения задания согласно заявленным требованиям

При оценке выполненной лабораторной работы учитываются следующие критерии:

Отчет составлен, верно, даны подробные описания определений и понятий, верно приведены примеры – 5 баллов;

Отчет выполнен в целом верно, имеются незначительные ошибки при оформлении отчетов – 4 балла;

Отчет выполнен со значительным количеством ошибок, не соответствует заявленному стандарту, часто демонстрируются ошибки - 3 балла;

Отчет составлен неверно, большое количество ошибок при оформлении – 2 балла.

Максимальное количество баллов за каждое выполненное лабораторное задание – 5 баллов.

Максимальное количество баллов за все задания раздела – 15 баллов.

Для текущего контроля ТКЗ:

Проверяемая компетенция: ПК-7. Способен разрабатывать системы анализа больших данных. ПК-7.1. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных. ПК-7.2. Разрабатывает программные компоненты обработки, удаленной, распределенной и объединенной аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных.

Отчет по лабораторной работе

Данный вид контроля за учебной деятельностью осуществляется в течение семестра. Посещение лабораторных занятий и выполнения заданий является допуском к зачету (промежуточной аттестации) по дисциплине. Обучающийся не допускается к зачету (промежуточной аттестации), если не сданы отчеты (в виде разработанных тестовых наборов или программ) по всем учебным модулям, а также в случае недобора баллов согласно бально-рейтинговой системы (менее 55). Для каждого раздела предусмотрено от 1 до 3 лабораторных заданий. Каждому студенту выдается индивидуальное задание.

Текущий контроль проводится перед началом каждого лабораторного занятия. Обучающиеся предоставляют отчет по лабораторной работе в электронном виде в виртуальной образовательной среде Moodle.

Перечень примерных заданий:

Раздел 3. Ускорение доступа к данным.

8. Работа с индексами в python.

Статистика:

1. Создайте индекс на таблице билетов (tickets) по имени пассажира (passenger_name).

2. Какая статистика имеется для этой таблицы?

3. Объясните оценку кардинальности и выбор плана выполнения следующих запросов:

а) выборка всех билетов,

б) выборка билетов на имя ALEKSANDR IVANOV,

в) выборка билетов на имя ANNA VASILEVA,

г) выборка билета с идентификатором 0005432000284

9. Оптимизация запросов.

Профилирование:

1. Выполните первую версию отчета, показанного в демонстрации, включив вывод текста и времени

выполнения запросов в журнал сообщений.

2. Посмотрите, какая информация попала в журнал сообщений.

3. Повторите предыдущие пункты, включив расширение auto_explain с выводом вложенных запросов.

Оптимизация запросов

1. Оптимизируйте запрос, выводящий контактную информацию пассажиров, летевших бизнес-классом, рейсы которых были задержаны более чем на 5 часов.

Критерием оценивания результатов является правильность выполнения задания согласно заявленным требованиям

При оценке выполненной лабораторной работы учитываются следующие критерии:

Отчет составлен, верно, даны подробные описания определений и понятий, верно приведены примеры – 5 баллов;

Отчет выполнен в целом верно, имеются незначительные ошибки при оформлении отчетов – 4 балла;

Отчет выполнен со значительным количеством ошибок, не соответствует заявленному стандарту, часто демонстрируются ошибки – 3 балла;

Отчет составлен неверно, большое количество ошибок при оформлении – 2 балла.

Максимальное количество баллов за каждое выполненное лабораторное задание – 5 баллов.

Отчет по контрольной работе

Цель работы:

В результате такого проекта студенты приобретут практические навыки работы с базами данных и СУБД, научатся использовать различные алгоритмы и структуры данных для эффективного хранения и извлечения данных, а также улучшат свои навыки программирования на Python.

Задание и требования:

Студенты могут реализовать алгоритмы объединения множеств, такие как сортированное слияние и объединение по хэшу, для оптимизации поиска и извлечения информации из базы данных. Они могут также разработать систему кэширования для ускорения доступа к данным, используя LRU-кэш или другие алгоритмы.

Работа должна быть оформлена в виде python скрипта и отправлена в форму для приема работы. Задание индивидуально.

Обучающиеся предоставляют отчет по контрольной работе в электронном виде в виртуальной образовательной среде Moodle.

Примерный перечень тем

1. Разработка базы данных для хранения и извлечения информации о книгах

Примерные задания

Практический проект на Python, который студенты могут создать в рамках данного модуля, может быть связан с разработкой базы данных для хранения и извлечения информации о книгах. Студенты могут использовать алгоритмы и структуры данных, изученные в данном модуле, для реализации индексов, таких как хэш-индекс и LSM-дерево, а также базовых СУБД типа “ключ-значение”, таких как Redis и MongoDB.

Студенты могут разработать веб-приложение, которое позволит пользователям добавлять книги в базу данных, искать их по ключевым словам и категориям, а также просматривать информацию о каждой книге, включая ее название, автора, жанр и количество страниц.

Критерием оценивания результатов является правильность выполнения задания согласно заявленным требованиям

При оценке выполненной контрольной работы учитываются следующие критерии:

Отчет составлен, верно, даны подробные описания определений и понятий, верно приведены примеры – 10 баллов;

Отчет выполнен в целом верно, имеются незначительные ошибки при оформлении отчетов – 8 балла;

Отчет выполнен со значительным количеством ошибок, не соответствует заявленному стандарту, часто демонстрируются ошибки – 6 балла;

Отчет составлен неверно, большое количество ошибок при оформлении – 4 балла.

Максимальное количество баллов за контрольную работу – 10 баллов.

Максимальное количество баллов за все задания раздела – 20 баллов.

Для промежуточной аттестации:

Тест

Вопрос	Варианты ответа
Имеется некоторая структура данных, в которую заносятся упорядоченные по убыванию символы. Считывание данных из этой структуры даёт результат: F, E, D, C, B, A. Чем является эта структура данных?	<input type="checkbox"/> Дерево
	<input type="checkbox"/> Граф
	<input type="checkbox"/> Очередь
	<input type="checkbox"/> Стек
Имеется упорядоченный массив целых чисел из 9 элементов. Сколько операций сравнения потребуется при двоичном поиске для нахождения искомого ключа, если он находится в точно в середине массива?	<input type="checkbox"/> 0
	<input type="checkbox"/> 9
	<input type="checkbox"/> 1
	<input type="checkbox"/> 5
	<input type="checkbox"/> 8
Имеется двоичное дерево (не являющееся деревом поиска), содержащее целые числа. Восходящий просмотр дерева даёт следующий результат: 2, 4, 6, 8, 10, 12, 14. Какой узел является корнем дерева?	<input type="checkbox"/> 14
	<input type="checkbox"/> 2
	<input type="checkbox"/> 8
	<input type="checkbox"/> 6
	<input type="checkbox"/> 10
Какая сортировка из следующих является самой неэффективной?	<input type="checkbox"/> Пузырьковая
	<input type="checkbox"/> Отбором
	<input type="checkbox"/> Быстрая
	<input type="checkbox"/> Шелла
Имеется неупорядоченный массив целых чисел. Для нахождения ключа используется двоичный поиск. Гарантируется ли в этом случае истинность результата поиска?	<input type="checkbox"/> Да
	<input type="checkbox"/> Гарантируется при условии, что значение ключа не превышает размера массива
	<input type="checkbox"/> Гарантируется при условии, что в процедуре поиска используется цикл for
	<input type="checkbox"/> Гарантируется при условии, что в процедуре поиска используется цикл while
	<input type="checkbox"/> Нет
Какие операции над элементами характерны для очередей и стеков?	<input type="checkbox"/> Поиск элемента и сортировка
	<input type="checkbox"/> Занесение элемента, извлечение элемента и просмотр
	<input type="checkbox"/> Занесение элемента и извлечение элемента
	<input type="checkbox"/> Занесение элемента, извлечение элемента и очистка
	<input type="checkbox"/> Занесение элемента, извлечение элемента, просмотр, сортировка и удаление текущего элемента
Какая структура данных используется для сохранения и восстановления содержимого регистров общего назначения центрального процессора при вызове процедур?	<input type="checkbox"/> Стек
	<input type="checkbox"/> Двоичное дерево
	<input type="checkbox"/> Список
	<input type="checkbox"/> Таблица
	<input type="checkbox"/> Очередь

Полный комплект заданий и материалов, необходимых для оценивания

результатов обучения по дисциплине, хранится на кафедре разработчика и содержит 50 тестовых вопросов на каждую компетенцию, из них 20% - закрытого типа, 80% - открытого типа.

Вопросы к экзамену

1. Различия, составляющие реляционные СУБД: компоненты и процесс обработки запроса, основные этапы. Основные элементы оптимизатора.

2. Применение операций выборки и проекции. Различные подходы к осуществлению операции соединения. Исполнение операции агрегации.

3. Понятие переписывания запроса. Основные типы преобразований. Понимание оптимизации запросов. Примеры классических систем в оптимизации запросов. Понятие канонического представления и его пример. Логическая и физическая алгебра. Трансформационный оптимизатор.

4. Задачи оптимизатора. Операторные деревья: линейные, лево-линейные, и кустистые деревья. Граф соединений и его смысл. Алгоритм построения кустистых деревьев. Оптимизация сложных запросов.

5. История гистограмм. Значение использования гистограмм в СУБД. Определение и примеры гистограмм. Основные понятия гистограммы: спред, площадь и частота. Гистограммы с постоянной шириной. Примеры серийных гистограмм. Экспериментальное исследование различных типов гистограмм.

6. Классификация гистограмм по параметру источника и параметру сортировки. Гистограммы равной глубины: алгоритм построения и примеры. Класс серийных гистограмм: определение, описание v-optimal и spline-based.

7. Распределенные СУБД: история и цель. Прозрачность в распределенных СУБД. Фрагментация и репликация в распределенных СУБД. Аспекты распределенных СУБД: автономность, распределенность, гетерогенность.

8. Основные типы распределенных СУБД: клиент-серверные, P2P СУБД, гетерогенные мультибазы.

9. Оптимизация в общем для распределенных СУБД. Два типа стоимостных моделей для оптимизации запросов в распределенных средах, с примером. Выполнение запросов в распределенных СУБД.

10. Выполнение запросов в клиент-серверных распределенных СУБД, применяемые стратегии. Оптимизация в клиент-серверной системе. Двухэтапный подход. Выполнение запросов в гетерогенных распределенных СУБД. Статистическая информация в гетерогенных системах.

11. Представление данных на диске: слотированная страница. История колоночных систем. Два фактора, стимулирующие популярность колоночного подхода. Предшественники колоночных систем. Два подхода к построению колоночных систем. Общая архитектура колоночных систем.

12. Сжатие в колоночных системах: алгоритмы и преимущества легковесных схем. Операции над сжатыми данными. Отложенная и прямая материализация в агрегированных запросах без соединений.

13. OLAP и OLTP: схемы "звезда" и "снежинка". Выполнение соединений с помощью стратегий отложенной и прямой материализации. Проблемы

прямой материализации и доступа к диску. Невидимое соединение: алгоритм и преимущества.

14. Критика реляционной модели. Объектно-ориентированные БД и объектно-реляционные БД. Структура данных в ООБД и ОРБД. Запросы в ООБД и ОРБД. Архитектура объектных систем и основные вопросы. Особенности оптимизации в объектных системах.

15. Объектные серверы и страничные серверы: факты и рассуждения. Управление буфером. Управление OID: стратегии POID и LOID. Типы LOID, генерация LOID, перевод LOID в POID. Хардварное и программное Pointer Swizzling. Понятие выражения пути и соображения по оптимизации.

16. Задача настройки СУБД. Автоматическая настройка СУБД: история и актуальность. Подходы к настройке физического уровня. Формальная постановка задачи и ее три компонента. Решение и классификация методов решения. Процедурные и стоимостные методы.

17. Горизонтальное фрагментирование данных.

18. Для чего применяется кластеризация записей.

19. Что такое B-дерево и B+-дерево. Многомерное индексирование.

20. Задача расщепления в R-дереве.

21. KD-дерево. Построение KD-дерева.

22. Графы в базах данных.