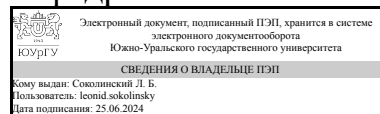


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Заведующий выпускающей
кафедрой



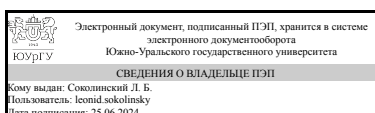
Л. Б. Соколинский

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.М0.02 Методы и системы обработки больших данных
для направления 09.04.04 Программная инженерия
уровень Магистратура
магистерская программа Искусственный интеллект и инженерия данных
форма обучения очная
кафедра-разработчик Системное программирование

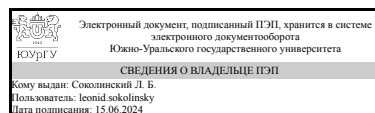
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 09.04.04 Программная инженерия, утверждённым приказом Минобрнауки от 19.09.2017 № 932

Зав.кафедрой разработчика,
д.физ.-мат.н., проф.



Л. Б. Соколинский

Разработчик программы,
д.физ.-мат.н., проф., заведующий
кафедрой



Л. Б. Соколинский

1. Цели и задачи дисциплины

Цель курса - дать представление о моделях, методах и алгоритмах, используемых в современных системах хранения и обработки больших данных. Эти знания позволять понять как работают СУБД и другие программные системы. Это позволит в будущем принимать правильные проектные решения в управлении центрами обработки данных в больших компаниях.

Краткое содержание дисциплины

Введение в обработку запросов в системах баз данных. Разбор запроса. Конверсия запроса. Логическая оптимизация. Организация системы баз данных. Индексы. Оценка стоимости операций. Статистические характеристики данных. Реализация исполнителя запросов. Алгоритмы соединения. Порядок соединений. Введение в параллельные системы баз данных. Формы параллельной обработки транзакций. Определение параллельной системы баз данных. Классификация многопроцессорных систем. Архитектура параллельных систем баз данных. Фрагментация. Организация межпроцессорных обменов. Балансировка загрузки в многопроцессорных иерархиях.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-8 Способен разрабатывать системы хранения и обработки больших данных	Знает: фундаментальные знания в области разработки систем управления большими данными Умеет: осуществлять первичный сбор и анализ материала в области разработки систем управления большими данными Имеет практический опыт: анализа и оптимизации найденных решений в области разработки систем управления большими данными

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Нет	Технологии распределенной обработки данных

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Нет

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 56,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		2
Общая трудоёмкость дисциплины	108	108
<i>Аудиторные занятия:</i>	48	48
Лекции (Л)	32	32
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	16	16
Лабораторные работы (ЛР)	0	0
<i>Самостоятельная работа (СРС)</i>	51,5	51,5
Подготовка к 4 промежуточным тестам	16	16
Подготовка к экзамену	10	10
Изучение дополнительных материалов в области разработки параллельных баз данных	25,5	25.5
Консультации и промежуточная аттестация	8,5	8,5
Вид контроля (зачет, диф.зачет, экзамен)	-	экзамен

5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Обработка запросов в системах баз данных	48	32	16	0

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Общая схема обработки запроса в реляционной СУБД. Реляционная алгебра и реляционные операции. Представление реляционного выражения в виде логического плана. Общая схема компиляции запроса.	4
2	1	Разбор запроса. Синтаксический анализ запроса. Грамматика SQL. Построение дерева разбора. Разрешение представлений. Верификация запроса.	4
3	1	Конверсия запроса в логический план.	2
4	1	Логическая оптимизация. Алгебраические законы. Оптимизация операций выборки, проекции и удаления дубликатов. Оптимизация композиции операций выборки и прямого произведения.	4
5	1	Система хранения данных. Поля, записи, блоки. Использование буферного пула.	2
6	1	Индексы. В-деревья.	2
7	1	Оптимизация физического плана. Оценка стоимости реляционных операций. Статистические характеристики данных. Гистограммы. Выбор порядка соединений.	2
8	1	Исполнитель запросов. Итераторы. Алгоритмы соединений.	4
9	1	Модель параллельной обработки больших данных MapReduce.	4
10	1	Скетчинг данных. Фильтры Блума. Count-Min скетч. Определение числа	4

		уникальных элементов с помощью HyperLogLog.	
--	--	---	--

5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1	1	Реляционная алгебра	4
2	1	Логический план запросов	4
3	1	Построение физического плана запросов	4
4	1	Оптимизация физического плана запросов	4

5.3. Лабораторные работы

Не предусмотрены

5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Подготовка к 4 промежуточным тестам	Основная литература. Презентации к лекциям. Практические работы.	2	16
Подготовка к экзамену	Основная литература. Презентации к лекциям	2	10
Изучение дополнительных материалов в области разработки параллельных баз данных	1. Пан К.С. - ПОДХОД К РАЗБИЕНИЮ СВЕРХБОЛЬШИХ ГРАФОВ С ПОМОЩЬЮ ПАРАЛЛЕЛЬНЫХ СУБД Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика - 2012г. №2 Электронно-библиотечная система издательства Лань https://e.lanbook.com/journal/issue/296731 2. Гавриш Е.В., Колтаков А.В., Медведев А.А., Соколинский Л.Б. - ПАРАЛЛЕЛЬНАЯ СУБД С ОТКРЫТЫМ ИСХОДНЫМ КОДОМ ДЛЯ КЛАСТЕРНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика - 2013г. №3 Электронно-библиотечная система издательства Лань https://e.lanbook.com/journal/issue/296735	2	25,5

6. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления баллов	Учитывается в ПА
1	2	Текущий контроль	Тест 1	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 30 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	экзамен
2	2	Текущий контроль	Тест 2	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 30 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	экзамен
3	2	Текущий контроль	Тест 3	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 30 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	экзамен
4	2	Текущий контроль	Тест 4	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 30 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	экзамен
5	2	Текущий контроль	Практическое задание 1. Реляционная алгебра и логический план	1	2	Наличие тетради с прорешанными на занятии задачами + 1 балл. Ответы на вопросы к заданию + 1 балл Максимальное количество баллов за задание - 2.	экзамен

			запросов				
6	2	Текущий контроль	Практическое задание 2. Физический план запросов	1	2	Наличие тетради с прорешанными на занятии задачами + 1 балл. Ответы на вопросы к заданию + 1 балл Максимальное количество баллов за задание - 2.	экзамен
7	2	Промежуточная аттестация	Итоговый тест	-	12	Тест содержит 12 равнозначных вопросов и рассчитан на 45 мин. Дается две попытки. Вторая попытка не зависит от первой. Количество баллов равно количеству правильных ответов на вопросы теста.	экзамен

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
экзамен	При оценивании результатов учебной деятельности обучающегося по дисциплине используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (Положение о БРС утверждено приказом ректора от 24.05.2019 г. № 179, в редакции приказа ректора от 10.03.2022 г. № 25-13/09). Процедура прохождения промежуточной аттестации осуществляется согласно Положению о текущем контроле успеваемости и промежуточной аттестации (приказ ректора от 27.02.2024 № 33-13/09). Оценка за дисциплину формируется на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля следующим образом: • Отлично: Величина рейтинга обучающегося по дисциплине 85...100 %. • Хорошо: Величина рейтинга обучающегося по дисциплине 75...84 %. • Удовлетворительно: Величина рейтинга обучающегося по дисциплине 60...74 %. • Неудовлетворительно: Величина рейтинга обучающегося по дисциплине 0...59 %. Если студент согласен с оценкой, полученной по результатам текущего контроля, то он может в день, предшествующий промежуточной аттестации дать свое согласие на автомат в личном кабинете. В случае явки студента на промежуточную аттестацию, давшего свое согласие на автомат в личном кабинете, студент имеет право пройти мероприятия текущего контроля по дисциплине на промежуточной аттестации для улучшения своего рейтинга в день ее проведения. Снижение оценки в этом случае запрещено. Если студент не дал согласия в личном кабинете, то он может согласиться с оценкой лично на промежуточной аттестации в день ее проведения. Если студент не согласен с оценкой, то он имеет право пройти контрольно-рейтинговые мероприятия на промежуточной аттестации для улучшения своего рейтинга в день ее проведения. Фиксация результатов учебной деятельности по дисциплине проводится в день промежуточной аттестации на основе согласия студента, данного им в личном кабинете. При отсутствии согласия в журнале дисциплины фиксация результатов происходит при личном присутствии студента. Если студент не дал согласие в личном кабинете и не явился на	В соответствии с пп. 2.5, 2.6 Положения

	промежуточную аттестацию – ему выставляется «неявка». Промежуточная аттестация проводится в форме тестирования. Тестирование проводится в системе edu.susu.ru. Тест содержит 12 вопросов. На выполнение теста дается 45 минут. В этом случае оценка за дисциплину рассчитывается на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и промежуточной аттестации.	
--	---	--

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ						
		1	2	3	4	5	6	7
ПК-8	Знает: фундаментальные знания в области разработки систем управления большими данными	+	+	+	+	+	+	+
ПК-8	Умеет: осуществлять первичный сбор и анализ материала в области разработки систем управления большими данными	+	+	+	+	+	+	+
ПК-8	Имеет практический опыт: анализа и оптимизации найденных решений в области разработки систем управления большими данными	+	+	+	+	+	+	+

Типовые контрольные задания по каждому мероприятию находятся в приложениях.

7. Учебно-методическое и информационное обеспечение дисциплины

Печатная учебно-методическая документация

а) основная литература:

1. Дейт, К. Д. Введение в системы баз данных К. Д. Дейт; Пер. с англ. Ю. Г. Гордиенко и др.; Под ред. А. В. Слепцова. - 7-е изд. - М. и др.: Вильямс, 2001. - 1071 с. табл.

б) дополнительная литература:

Не предусмотрена

в) отечественные и зарубежные журналы по дисциплине, имеющиеся в библиотеке:

1. Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. науч. журн. Южно-Уральский гос. ун-т
2. Программирование ,науч. журн. ,Рос. акад. наук, Отд-ние информатики, вычисл. техники и автоматизации, Моск. гос. ун-т

г) методические указания для студентов по освоению дисциплины:

1. Методические указания для студентов и преподавателей по освоению и организации самостоятельной работы студентов

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Методические указания для студентов и преподавателей по освоению и организации самостоятельной работы студентов

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной	Библиографическое описание
---	----------------	------------------------------------	----------------------------

		форме	
1	Дополнительная литература	Электронно-библиотечная система издательства Лань	Пан К.С. - ПОДХОД К РАЗБИЕНИЮ СВЕРХБОЛЬШИХ ГРАФОВ С ПОМОЩЬЮ ПАРАЛЛЕЛЬНЫХ СУБД Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика - 2012г. №2 https://e.lanbook.com/journal/issue/296731
2	Дополнительная литература	Электронно-библиотечная система издательства Лань	Гавриш Е.В., Колтаков А.В., Медведев А.А., Соколинский Л.Б. - ПАРАЛЛЕЛЬНАЯ СУБД С ОТКРЫТЫМ ИСХОДНЫМ КОДОМ ДЛЯ КЛАСТЕРНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика - 2013г. №3 https://e.lanbook.com/journal/issue/296735
3	Основная литература	Электронно-библиотечная система издательства Лань	Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с. — ISBN 978-5-97060-190-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/93571 (дата обращения: 10.10.2021). — Режим доступа: для авториз. пользователей.

Перечень используемого программного обеспечения:

1. РСК Технологии-Система "Персональный виртуальный компьютер" (ПВК) (MS Windows, MS Office, открытое ПО)(бессрочно)

Перечень используемых профессиональных баз данных и информационных справочных систем:

1. EBSCO Information Services-EBSCOhost Research Databases(28.02.2017)

8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Лекции	110 (3г)	Wi-Fi, MS Office, аудитория должна быть оборудована электрическими розетками, доской и проектором.
Зачет	110 (3г)	Wi-Fi, MS Office, аудитория должна быть оборудована электрическими розетками, доской и проектором, компьютеры должны быть подключены к локальной вычислительной сети и интернету.
Практические занятия и семинары	110 (3г)	Wi-Fi, MS Office, аудитория должна быть оборудована электрическими розетками, доской и проектором, компьютеры должны быть подключены к локальной вычислительной сети и интернету.