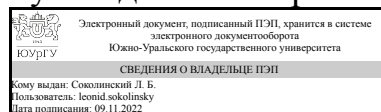


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Руководитель направления



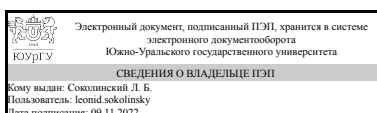
Л. Б. Соколинский

РАБОЧАЯ ПРОГРАММА

дисциплины 1.О.09 Теоретические основы разработки систем управления большими данными
для направления 02.04.02 Фундаментальная информатика и информационные технологии
уровень Магистратура
форма обучения очная
кафедра-разработчик Системное программирование

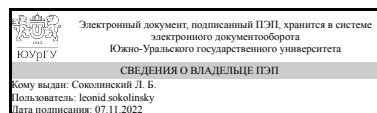
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 02.04.02 Фундаментальная информатика и информационные технологии, утверждённым приказом Минобрнауки от 23.08.2017 № 811

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 обработка 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
ОПК-1 Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	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
ОПК-2 Способен применять компьютерные/суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности	Знает: классификацию многопроцессорных систем, архитектуру систем управления большими данными Умеет: выбирать эффективную многопроцессорную систему для системы управления большими данными Имеет практический опыт: анализа эффективности многопроцессорной системы для системы управления большими данными
ОПК-3 Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования	Знает: методы, основные положения и концепции в области параллельной обработки запросов в системе управления большими данными Умеет: оценивать стоимость реляционных операций и реализовывать параллельные алгоритмы реляционных операций Имеет практический опыт: анализа эффективности решений в области параллельной обработки запросов в системе управления

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
1.О.04 Языки разметки, 1.О.18 Основы машинного обучения, 1.О.13 Язык Python для анализа данных, 1.О.03 Криптография и защита информации, ФД.01 Технологии интернета вещей, 1.О.21 Системы управления предприятием, ФД.02 Методы искусственного интеллекта, 1.О.10 Нейронные сети, 1.О.14 Программирование корпоративных информационных систем на языке Java	1.О.17 Квантовые вычисления, 1.О.11 Технологии параллельного программирования, 1.О.07 Современные методы DevOps, 1.О.08 Анализ информационных технологий

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

Дисциплина	Требования
ФД.01 Технологии интернета вещей	Знает: принципы организации киберфизических систем, существующие технологии в интернете вещей, отечественные и зарубежные достижения в области программно-аппаратных комплексов интернета вещей Умеет: анализировать существующие IoT-технологии и применять их в конкретных условиях, определять сервисы, функции и выбирать технологии их реализации при разработке киберфизических программно-аппаратных компонентов Имеет практический опыт: владения специальной терминологией, навыками программирования конечных устройств, навыками разработки моделей и алгоритмов для взаимодействия с программными и аппаратными компонентами, самостоятельного проектирования и реализации компонентов интернета вещей
1.О.04 Языки разметки	Знает: основы языков XSL, XSD и DTD, основные направления применения стандарта XML в управлении IT-проектами, корпоративными информационными системами и высоконагруженными Web-системами, способы разработки языков разметки на основе XML Умеет: создавать спецификацию XML-документа с помощью языков XSD и DTD, преобразовывать XML-документ в HTML с помощью XSL шаблона, осуществлять импорт-экспорт данных для XML-формата, на основе анализа исходных данных формировать XML-документ Имеет практический опыт: владения навыками по валидации и отладке XSD, DTD и XSL-документов, владения инструментами

	импорта-экспорта данных для XML-формата, владения методами валидации и отладки XML-документа
ФД.02 Методы искусственного интеллекта	<p>Знает: математические основы и технологии машинного обучения, современные интегрированные среды разработки ПО на языках высокого уровня и специализированные библиотеки искусственного интеллекта</p> <p>Умеет: применять современные методы машинного обучения на основе нейронных сетей, создавать и обучать глубокие и сверточные искусственные нейронные сети с применением специализированных библиотек</p> <p>Имеет практический опыт: анализа и оптимизации полученных решений на основе нейросетевого подхода, решения задач в области машинного обучения и компьютерного зрения</p>
1.О.21 Системы управления предприятием	<p>Знает: функциональные возможности систем класса ERP, среду разработки системы SAP ERP, методологию разработки, внедрения и эксплуатации систем управления предприятием, классификацию современных систем управления предприятием, задачи, решаемые с помощью систем управления предприятием</p> <p>Умеет: применять полученные знания для решения задач профессиональной деятельности с помощью систем управления предприятием, применять методологию разработки, внедрения и эксплуатации систем управления предприятием, находить, формулировать и решать актуальные проблемы с помощью систем управления предприятием</p> <p>Имеет практический опыт: решения задач профессиональной деятельности с помощью систем управления предприятием, управления проектами, связанными с разработкой, внедрением и эксплуатацией систем управления предприятием, решения актуальных проблем с помощью систем управления предприятием</p>
1.О.13 Язык Python для анализа данных	<p>Знает: программное обеспечение для решения задач анализа данных, основные инструменты (программные библиотеки и язык программирования) для обработки и анализа данных, основные инструменты (программные библиотеки и язык программирования) для выполнения операций обработки и анализа данных</p> <p>Умеет: устанавливать программное обеспечение (среды разработок, программные библиотеки, соответствующий backend), просматривать версию и состав используемого программного обеспечения, задавать требуемый backend для решения поставленной задачи, подбирать наиболее подходящие инструменты для анализа имеющихся данных и выявления закономерностей, применять специализированные библиотеки языка Python</p>

	<p>для сбора, обработки и анализа данных Имеет практический опыт: установки и инсталляции программного обеспечения, используемого для решения задач в области сбора, обработки и анализа данных, анализа готовых информационных наборов данных, сбора данных в различных форматах (csv, json, xml), предварительной подготовки данных (приведение типов/форматов, заполнение пропусков фильтрация, и т.п.); анализа и визуализации данных</p>
1.О.10 Нейронные сети	<p>Знает: технологии создания искусственных нейронных сетей с применением высокоуровневого языка программирования Python, математические основы и технологии создания и обучения искусственных нейронных сетей, математическую модель нейрона, основы линейной алгебры, технологии создания искусственных нейронных сетей, методы оптимизации, регуляризации и нормализации параметров нейронной сети и процесса ее обучения, методы оптимизации, регуляризации и нормализации параметров нейронной сети и процесса ее обучения Умеет: создавать и обучать глубокие и сверточные искусственные нейронные сети на Python, применять современные методы машинного обучения на основе нейронных сетей, осуществлять формализацию задачи, построение математической модели, подготовку обучающего набора данных, подбор топологии и создание искусственной нейронной сети в соответствии с поставленной задачей, создавать глубокие и сверточные искусственные нейронные сети с применением специализированных библиотек на разных вычислительных платформах (CPU/GPU/TPU) Имеет практический опыт: анализа и оптимизации полученных решений на основе нейросетевого подхода, формулирования и решения задач в области машинного обучения с использованием нейросетевого подхода, решения задач распознавания образов на разных вычислительных платформах (CPU/GPU/TPU)</p>
1.О.18 Основы машинного обучения	<p>Знает: математические основы, принципы создания, обучения и валидации моделей машинного обучения, технологию создания моделей машинного обучения с помощью библиотек языка Python, методы оптимизации, регуляризации, нормализации и валидации моделей машинного обучения Умеет: применять современные методы машинного обучения, создавать и обучать модели машинного обучения с помощью библиотек языка Python Имеет практический опыт: анализа и оптимизации полученных решений на основе машинного обучения, решения задач машинного обучения с</p>

	помощью библиотек языка Python
1.О.03 Криптография и защита информации	<p>Знает: основные подходы к математической формализации различных аспектов безопасности информационных систем и реализации средств защиты информации, основные требования информационной безопасности, основные алгоритмы шифрования данных, базовые понятия для математического обеспечения информационной безопасности</p> <p>Умеет: применять математические методы и алгоритмы защиты информации при решении профессиональных задач в области информационной безопасности, применять математические методы защиты информации, кодировать информацию с помощью основных алгоритмов шифрования</p> <p>Имеет практический опыт: самостоятельного формулирования задач и политик безопасности, построения систем защиты, использования основных алгоритмов шифрования для защиты данных и информационной безопасности</p>
1.О.14 Программирование корпоративных информационных систем на языке Java	<p>Знает: основы объектно-ориентированного языка, основные понятия, виды и характеристики современного программного обеспечения технологии Java, подходы к тестированию программ на Java, основные понятия, виды и характеристики современного программного обеспечения технологии Java</p> <p>Умеет: применять подходы объектно-ориентированного программирования при разработке программного обеспечения, проектировать и разрабатывать локальные приложения на языке Java, разрабатывать документацию с помощью Javadoc, использовать специализированные среды разработки Java</p> <p>Имеет практический опыт: проектирования классов, ООП-архитектуры, создания программных проектов в специализированных средах разработки Java, разработки тестов для веб-сайта с помощью библиотеки Selenium, создания программных проектов в специализированных средах разработки Java</p>

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 38,25 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>	32	32
Лекции (Л)	16	16
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	16	16
Лабораторные работы (ЛР)	0	0
<i>Самостоятельная работа (СРС)</i>	69,75	69,75
Изучение дополнительных материалов в области разработки параллельных баз данных	36	36
Подготовка к зачету	13,75	13,75
Подготовка к 4 промежуточным тестам	20	20
Консультации и промежуточная аттестация	6,25	6,25
Вид контроля (зачет, диф.зачет, экзамен)	-	зачет

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	32	16	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.	2
2	1	Разбор запроса. Синтаксический анализ запроса. Грамматика SQL. Построение дерева разбора. Разрешение представлений. Верификация запроса.	2
3	1	Конверсия запроса в логический план.	1
4	1	Логическая оптимизация. Алгебраические законы. Оптимизация операций выборки, проекции и удаления дубликатов. Оптимизация композиции операций выборки и прямого произведения.	2
5	1	Система хранения данных. Поля, записи, блоки. Использование буферного пула.	1
6	1	Индексы. В-деревья.	1
7	1	Оптимизация физического плана. Оценка стоимости реляционных операций. Статистические характеристики данных. Гистограммы. Выбор порядка соединений.	1
8	1	Исполнитель запросов. Итераторы. Алгоритмы соединений.	2
9	1	Модель параллельной обработки больших данных MapReduce.	2
10	1	Скетчинг данных. Фильтры Блума. Count-Min скетч. Определение числа уникальных элементов с помощью HyperLogLog.	2

5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1	1	1) Вычислить результат выражения реляционной алгебры - 5 примеров; 2)	4

		Написать выражение реляционной алгебры, вычисляющее ответ на запрос к базе данных "Поставки" - 8 примеров; 3) Написать выражение реляционной алгебры, соответствующее логическому плану - 5 примеров.	
2	1	1) Построить дерево разбора для SQL-запроса - 4 примера; 2) Конвертировать дерево разбора в логический план - 5 примеров; 3) Выполнить оптимизацию логического плана - 4 примера.	4
3	1	1) Вычислить количество блоков, необходимых для хранения отношения - 4 примера; 2) Вычислить количество блоков, необходимое для хранения индексного файла - 4 примера; 3) Вычислить оценку для размера результата реляционного выражения - 4 примера.	4
4	1	1) Вычислить оценку размера естественного соединения отношений с использованием гистограмм - 4 примера; 2) Вычислить оценку эффективности алгоритма соединения - 5 примеров; 3) С использованием метода динамического программирования определить оптимальный порядок соединений - 4 примера.	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
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	36
Подготовка к зачету	Основная литература. Презентации к лекциям	2	13,75
Подготовка к 4 промежуточным тестам	Основная литература. Презентации к лекциям. Практические работы.	2	20

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	зачет
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	зачет
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	зачет
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	зачет
5	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.	зачет

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
зачет	<p>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). Оценка за дисциплину формируется на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля. Зачтено: Величина рейтинга обучающегося по дисциплине 60...100 %. Незачтено: Величина рейтинга обучающегося по дисциплине 0...59 %. Если студент не согласен с оценкой, полученной по результатам текущего контроля, студент проходит мероприятие промежуточной аттестации в виде тестирования. Тестирование проводится в системе edu.susu.ru. Тест содержит 12 вопросов. На выполнение теста дается 45 минут. Дается две попытки. Вторая попытка не зависит от первой. В этом случае оценка за дисциплину рассчитывается на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и промежуточной аттестации. Фиксация результатов учебной деятельности по дисциплине проводится в день зачета при личном присутствии студента.</p>	В соответствии с пп. 2.5, 2.6 Положения

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ				
		1	2	3	4	5
ОПК-1	Знает: фундаментальные знания в области разработки систем управления большими данными	++				+
ОПК-1	Умеет: осуществлять первичный сбор и анализ материала в области разработки систем управления большими данными	++				+
ОПК-1	Имеет практический опыт: анализа и оптимизации найденных решений в области разработки систем управления большими данными	++				+
ОПК-2	Знает: классификацию многопроцессорных систем, архитектуру систем управления большими данными			+		+
ОПК-2	Умеет: выбирать эффективную многопроцессорную систему для системы управления большими данными			+		+
ОПК-2	Имеет практический опыт: анализа эффективности многопроцессорной системы для системы управления большими данными			+		+
ОПК-3	Знает: методы, основные положения и концепции в области параллельной обработки запросов в системе управления большими данными				++	
ОПК-3	Умеет: оценивать стоимость реляционных операций и реализовывать параллельные алгоритмы реляционных операций				++	
ОПК-3	Имеет практический опыт: анализа эффективности решений в области параллельной обработки запросов в системе 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:

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
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
Зачет, диф. 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.
Лекции	110 (3г)	Wi-Fi, MS Office, аудитория должна быть оборудована электрическими розетками, доской и проектором.