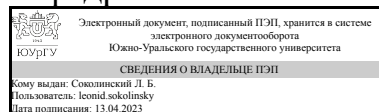


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Заведующий выпускающей
кафедрой



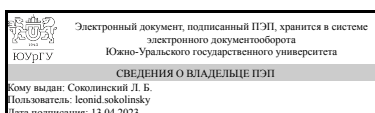
Л. Б. Соколинский

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.П0.04 Основы разработки систем управления большими данными
для направления 09.03.04 Программная инженерия
уровень Бакалавриат
профиль подготовки Инженерия информационных и интеллектуальных систем
форма обучения очная
кафедра-разработчик Системное программирование

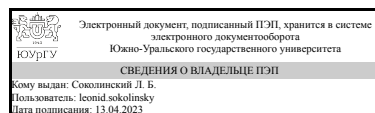
Рабочая программа составлена в соответствии с ФГОС ВО по направлению
подготовки 09.03.04 Программная инженерия, утверждённым приказом
Минобрнауки от 19.09.2017 № 920

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 позволят 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
ПК-6 (ПК-7 модели) Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	Знает: методы и технологии массово параллельной обработки и анализа данных Умеет: ПК-7.2. У-5. Умеет использовать методы и технологии массово параллельной обработки и анализа данных; Имеет практический опыт: применения методов и технологий массово параллельной обработки и анализа данных
ПК-7 (ПК-8 модели) Способен разрабатывать системы анализа больших данных	Знает: методы оперативной обработки потоков данных Умеет: ПК-8.1. У-3. Умеет выполнять потоковую обработку данных (data streaming, event processing); Имеет практический опыт: использования методов оперативной обработки потоков данных

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Глубокое обучение, Структуры и алгоритмы обработки данных, Базы данных, Технологии аналитической обработки информации, Основы интеллектуального анализа данных	Не предусмотрены

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

Дисциплина	Требования
Структуры и алгоритмы обработки данных	<p>Знает: виды представления данных, методы поиска и парсинга данных., базовые структуры данных и основные алгоритмы их обработки</p> <p>Умеет: выбирать оптимальные алгоритмы для решения задач предметной области и осуществлять их программную реализацию</p> <p>Имеет практический опыт: применения наиболее распространенных алгоритмов для решения задач с использованием сложных структур данных</p>
Глубокое обучение	<p>Знает: базовые архитектуры и модели искусственных нейронных сетей, базовые подходы к поиску и подготовке данных для моделей искусственных нейронных сетей, основы работы и построения моделей искусственных нейронных сетей</p> <p>Умеет: осуществлять оценку и отбор моделей искусственных нейронных сетей для решения задач, выделять входные и выходные переменные с целью использования предиктивных моделей, применять современные инструментальные средства и системы программирования для разработки моделей машинного обучения</p> <p>Имеет практический опыт: разработки систем искусственного интеллекта на основе моделей искусственных нейронных сетей, подготовки и разметки данных для моделей искусственных нейронных сетей, разработки моделей нейронных сетей для решения задач</p>
Технологии аналитической обработки информации	<p>Знает: методы и критерии оценки качества моделей машинного обучения, постановку базовых задач интеллектуального анализа данных (поиск шаблонов, классификация, кластеризация) и базовые методы их решения, общедоступные репозитории и специализированные библиотеки, содержащие наборы больших данных</p> <p>Умеет: сопоставить задачам предметной области классы задач машинного обучения, планировать и выполнять машинные эксперименты, оценивать точность и качество построенных моделей, настраивать и оптимизировать конфигурацию программного и аппаратного обеспечения с целью интеграции больших данных</p> <p>Имеет практический опыт: анализа требований и идентификации классов задач для реализации приложений машинного обучения, разработки приложений для аналитической обработки информации с помощью современных инструментальных</p>

	средств, разработки программных компонент для извлечения и подготовки больших данных для аналитической обработки информации
Базы данных	<p>Знает: устройство интерфейсов между реляционными SQL-хранилищами данных и нереляционными NoSQL-хранилищами данных, основы устройства систем баз данных, основы работы современных систем управления базами данных</p> <p>Умеет: 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), устанавливать и настраивать реляционные и нереляционные системы баз данных, создавать реляционные и нереляционные базы данных и запросы к ним</p> <p>Имеет практический опыт: написания запросов к реляционным и нереляционным большим базам данных, установки систем баз данных, разработки реляционных и нереляционных баз данных</p>
Основы интеллектуального анализа данных	<p>Знает: общедоступные репозитории и специализированные библиотеки, содержащие наборы больших данных, методы и критерии оценки качества моделей машинного обучения, постановку и методы решения основных задач интеллектуального анализа данных (поиск шаблонов, классификация, кластеризация)</p> <p>Умеет: настраивать и оптимизировать конфигурацию программного и аппаратного обеспечения с целью интеграции больших данных, сопоставить задачам предметной области классы задач машинного обучения, планировать и выполнять машинные эксперименты, оценивать точность и качество построенных моделей</p> <p>Имеет практический опыт: разработки программных компонент для извлечения и подготовки больших данных для интеллектуального анализа, анализа требований и определения необходимых классов задач для реализации приложений машинного обучения; определения метрик и критериев качества оценки моделей машинного обучения, разработки моделей машинного обучения для решения основных задач интеллектуального анализа данных (поиск шаблонов, классификация, кластеризация) и проведения вычислительных экспериментов по оценке точности и качества построенных моделей</p>

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 42,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		8	
Общая трудоёмкость дисциплины	108	108	
<i>Аудиторные занятия:</i>	36	36	
Лекции (Л)	24	24	
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	12	12	
Лабораторные работы (ЛР)	0	0	
<i>Самостоятельная работа (СРС)</i>	65,5	65,5	
Подготовка к 4 промежуточным тестам	18,5	18,5	
Изучение дополнительных материалов в области разработки параллельных баз данных	32	32	
Подготовка к дифференцированному зачету	15	15	
Консультации и промежуточная аттестация	6,5	6,5	
Вид контроля (зачет, диф.зачет, экзамен)	-	диф.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	36	24	12	0

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Введение. Общая схема обработки запроса в реляционной СУБД. Реляционная алгебра и реляционные операции. Представление реляционного выражения в виде логического плана. Общая схема компиляции запроса.	3
2	1	Разбор запроса. Синтаксический анализ запроса. Грамматика SQL. Построение дерева разбора. Разрешение представлений. Верификация запроса.	2
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.	2
5	1	Система хранения и индексные файлы. Поля, записи, блоки. Использование буферного пула. Индексы. В-деревья.	2
6	1	Оптимизация физического плана. Оценка стоимости реляционных операций. Статистические характеристики данных. Гистограммы. Выбор порядка соединений.	3
7	1	Исполнитель запросов. Итераторы. Алгоритмы соединений.	4
8	1	Исполнитель запросов. Итераторы. Алгоритмы соединений.	4
9	1	Модель параллельной обработки больших данных MapReduce. Структура вычислительного кластера. Распределенная файловая система. Схема работы MapReduce: функции высшего порядка Map и Reduce; процедура Shuffle. Программный стек Hadoop: набор инфраструктурных программных	2

		библиотек и утилит Hadoop Common, распределенная файловая система HDFS, система для планирования заданий и управления кластером YARN, платформа распределенных вычислений Hadoop MapReduce. Экосистема Hadoop: PIG - реализация реляционной алгебры на базе Hadoop, Hive - реализация SQL на базе Hadoop. Программная платформа Spark: Spark SQL - выполнение SQL-запросов над данными; Spark Streaming - обработка потоковых данных; Spark Mlib - набор библиотек машинного обучения; GraphX - распределенная обработка графов.	
--	--	--	--

5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1	1	Практическое задание 1. Реляционная алгебра	3
2	1	Практическое задание 2. Логический план запроса	3
3	1	Практическое задание 3. Физический план запроса	3
4	1	Практическое задание 4. Оптимизация запросов	3

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.	8	18,5
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	8	32
Подготовка к дифференцированному зачету	Основная литература. Презентации к лекциям	8	15

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	Учи-тыва-ется в ПА
1	8	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	дифференцированный зачет
2	8	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	дифференцированный зачет
3	8	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	дифференцированный зачет
4	8	Текущий контроль	Тест 4	1	3	Тест содержит три задания. Каждый правильный ответ оценивается 1 баллом.	дифференцированный зачет

						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	
5	8	Текущий контроль	Практическое задание 1. Реляционная алгебра	1	5	Решение 85% и больше примеров в задании оценивается в 5 баллов. Решение 75-84% примеров в задании оценивается в 4 балла. Решение 60-74% примеров в задании оценивается в 3 балла. Решение менее 60% примеров в задании оценивается в 0 баллов.	дифференцированный зачет
6	8	Текущий контроль	Практическое задание 2. Логический план запроса	1	5	Решение 85% и больше примеров в задании оценивается в 5 баллов. Решение 75-84% примеров в задании оценивается в 4 балла. Решение 60-74% примеров в задании оценивается в 3 балла. Решение менее 60% примеров в задании оценивается в 0 баллов.	дифференцированный зачет
7	8	Текущий контроль	Практическое задание 3. Физический план запроса	1	5	Решение 85% и больше примеров в задании оценивается в 5 баллов. Решение 75-84% примеров в задании оценивается в 4 балла. Решение 60-74% примеров в задании оценивается в 3 балла. Решение менее 60% примеров в задании оценивается в 0 баллов.	дифференцированный зачет
8	8	Текущий контроль	Практическое задание 4. Оптимизация запросов	1	5	Решение 85% и больше примеров в задании оценивается в 5 баллов. Решение 75-84% примеров в задании оценивается в 4 балла. Решение 60-74% примеров в задании оценивается в 3 балла. Решение менее 60%	дифференцированный зачет

		издательства Лань	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.
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.
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.