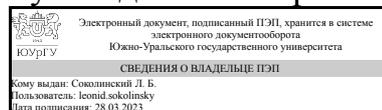


УТВЕРЖДАЮ:  
Руководитель направления



Л. Б. Соколинский

## РАБОЧАЯ ПРОГРАММА

дисциплины 1.О.08 Методы и системы обработки больших данных  
для направления 02.04.02 Фундаментальная информатика и информационные  
технологии

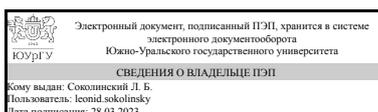
уровень Магистратура

форма обучения очная

кафедра-разработчик Системное программирование

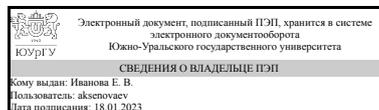
Рабочая программа составлена в соответствии с ФГОС ВО по направлению  
подготовки 02.04.02 Фундаментальная информатика и информационные технологии,  
утверждённым приказом Минобрнауки от 23.08.2017 № 811

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



Л. Б. Соколинский

Разработчик программы,  
к.физ.-мат.н., доцент



Е. В. Иванова

## 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 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-1 Способен разрабатывать системы хранения и обработки больших данных, в том числе на основе методов искусственного интеллекта	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
1.О.09 Глубокие нейронные сети	Не предусмотрены

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

Дисциплина	Требования
1.О.09 Глубокие нейронные сети	Знает: классы задач обработки больших данных на основе методов искусственных нейронных сетей, специализированные библиотеки для создания искусственных нейронных сетей,

	<p>математическую модель нейрона, технологии создания искусственных нейронных сетей, методы оптимизации, регуляризации и нормализации параметров нейронной сети и процесса ее обучения Умеет: применять современные инструментальные средства и системы программирования для разработки и обучения моделей искусственных нейронных сетей, осуществлять формализацию задачи, построение математической модели, подготовку обучающего набора данных, подбор топологии и создание искусственной нейронной сети в соответствии с поставленной задачей Имеет практический опыт: создания и обучения искусственных нейронных сетей с применением специализированных библиотек, формулирования и решения задач в области машинного обучения с использованием нейросетевого подхода</p>
--	---

#### 4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 54,25 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		1	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>	53,75	53,75	
Изучение дополнительных материалов в области разработки параллельных баз данных	28	28	
Подготовка к зачету	10,75	10.75	
Подготовка к 4 промежуточным тестам	15	15	
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	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 уникальных элементов с помощью HyperLogLog.	4

## 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
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 <a href="https://e.lanbook.com/journal/issue/296731">https://e.lanbook.com/journal/issue/296731</a>	2	28

	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 <a href="https://e.lanbook.com/journal/issue/296735">https://e.lanbook.com/journal/issue/296735</a>		
Подготовка к зачету	Основная литература. Презентации к лекциям	2	10,75
Подготовка к 4 промежуточным тестам	Основная литература. Презентации к лекциям. Практические работы.	2	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-мestr	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.	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	
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	Текущий контроль	Практическое задание 1. Реляционная алгебра и логический 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.	зачет
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.	зачет
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.	зачет

## 6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
зачет	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	В соответствии с пп. 2.5, 2.6 Положения

	мероприятие промежуточной аттестации в виде 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. Фиксация результатов учебной деятельности по дисциплине проводится в день зачета при личном присутствии студента.	
--	--	--

### 6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ						
		1	2	3	4	5	6	7
ПК-1	Знает: фундаментальные знания в области разработки систем управления большими данными	+	+	+	+	+	+	+
ПК-1	Умеет: осуществлять первичный сбор и анализ материала в области разработки систем управления большими данными	+	+	+	+	+	+	+
ПК-1	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	Библиографическое описание
---	----------------	------------------------	----------------------------

		электронной 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 <a href="https://e.lanbook.com/journal/issue/296731">https://e.lanbook.com/journal/issue/296731</a>
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 <a href="https://e.lanbook.com/journal/issue/296735">https://e.lanbook.com/journal/issue/296735</a>
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: <a href="https://e.lanbook.com/book/93571">https://e.lanbook.com/book/93571</a> (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
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.
Лекции	110 (3г)	Wi-Fi, MS Office, аудитория должна быть оборудована электрическими розетками, доской и проектором.