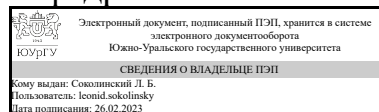


# ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:  
Заведующий выпускающей  
кафедрой



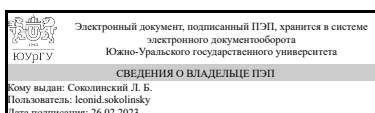
Л. Б. Соколинский

## 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

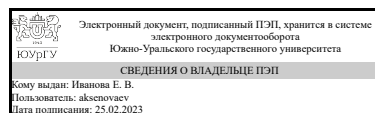
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 09.03.04 Программная инженерия, утверждённым приказом Минобрнауки от 19.09.2017 № 920

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



Л. Б. Соколинский

Разработчик программы,  
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 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
Основы интеллектуального анализа данных, Технологии аналитической обработки информации, Глубокое обучение, Структуры и алгоритмы обработки данных, Базы данных	Основы распределенной обработки данных

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., 54,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		6	
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>	53,5	53,5	
Изучение дополнительных материалов в области разработки параллельных баз данных	22,5	22,5	
Подготовка к 4 промежуточным тестам	16	16	
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	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	Практическое задание 1. Реляционная алгебра	4
2	1	Практическое задание 2. Логический план запроса	4
3	1	Практическое задание 3. Физический план запроса	4
4	1	Практическое задание 4. Оптимизация 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. 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>	6	22,5
Подготовка к 4 промежуточным тестам	Основная литература. Презентации к лекциям. Практические работы.	6	16
Подготовка к дифференцированному зачету	Основная литература. Презентации к лекциям	6	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	Учитывается в ПА
1	6	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	6	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	6	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	6	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	дифференцированный зачет

5	6	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	6	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	6	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	6	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% примеров в задании оценивается в 0 баллов.	дифференцированный зачет
10	6	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.	дифференцированный зачет



## 6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
дифференцированный 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).</p> <p>Оценка за дисциплину формируется на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля. Отлично: Величина рейтинга обучающегося по дисциплине 85...100 %.</p> <p>Хорошо: Величина рейтинга обучающегося по дисциплине 75...84 %. Удовлетворительно: Величина рейтинга обучающегося по дисциплине 60...74 %.</p> <p>Неудовлетворительно: 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 и 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 дифференцированного 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	6	7	8	10	
ПК-6	Знает: методы и технологии массово параллельной обработки и анализа данных	+	+	+	+	+	+	+	+	+	+
ПК-6	Умеет: ПК-7.2. У-5. Умеет использовать методы и технологии массово параллельной обработки и анализа данных;	+	+	+	+	+	+	+	+	+	+
ПК-6	Имеет практический опыт: применения методов и технологий массово параллельной обработки и анализа данных	+	+	+	+	+	+	+	+	+	+
ПК-7	Знает: методы оперативной обработки потоков данных										+
ПК-7	Умеет: ПК-8.1. У-3. Умеет выполнять потоковую обработку данных (data streaming, event processing);										+
ПК-7	Имеет практический опыт: использования методов оперативной обработки потоков данных										+

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 материалы по курсу Основы разработки систем управления большими данными
2. Методические указания для студентов и преподавателей по освоению и организации самостоятельной работы студентов

### из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Методические материалы по курсу Основы разработки систем управления большими данными
2. Методические указания для студентов и преподавателей по освоению и организации самостоятельной работы студентов

## 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 <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:

## 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.