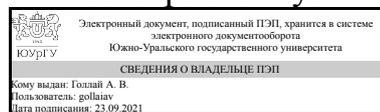


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Директор института
Высшая школа электроники и
компьютерных наук



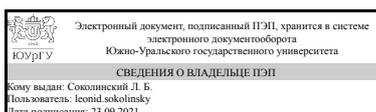
А. В. Голлой

РАБОЧАЯ ПРОГРАММА

дисциплины П.1.В.06.05 Обработка запросов в системах баз данных
для направления 09.06.01 Информатика и вычислительная техника
уровень аспирант тип программы
направленность программы
форма обучения очная
кафедра-разработчик Системное программирование

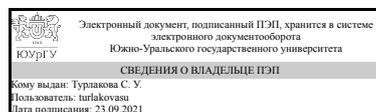
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 09.06.01 Информатика и вычислительная техника, утверждённым приказом Минобрнауки от 29.07.2014 № 875

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



Л. Б. Соколинский

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



С. У. Турлакова

1. Цели и задачи дисциплины

Целью курса является введение в проблематику обработки и оптимизации запросов в системах баз данных. При изучении этого курса должны быть решены следующие задачи: - знать этапы обработки запросов, преобразования дерева разбора; - иметь представление об алгоритмах оптимизации физических планов; - понимать методы генерации физического плана запроса.

Краткое содержание дисциплины

Компилятор запросов. Синтаксический анализ и деревья разбора. Грамматика для простого подмножества SQL. Препроцессор: разрешение представлений VIEW; семантический анализ (верификация имен отношений, верификация имен атрибутов и их квалификация, контроль типов). Алгебраические законы и планы запросов. Коммутативный и ассоциативный законы. Законы для операции выборки. Законы для операции проекции. Законы для соединения и прямого произведения. Преобразование дерева разбора в логический план запроса. Конверсия в реляционную алгебру. Удаление подзапросов из условий. Логическая оптимизация: продвижение операций выборки и проекции «вниз» по дереву запроса; замена сочетания выборки и прямого произведения на соединение. Пример, указывающий на целесообразность первоначального продвижения выборки вверх. Оценка стоимости операций. Оценка размеров промежуточных отношений. Оценка размеров проекции. Оценка размеров выборки. Оценка размеров соединения. Естественное соединение с несколькими атрибутами соединения. Соединение нескольких отношений. Оценка размеров для других операций. Использование статистик. Виды статистик (тотальные статистики и гистограммы). Гистограммы полные и частичные. Виды частичных гистограмм (с равной шириной интервалов, с равной высотой интервалов, с наиболее часто встречающимися значениями). Подходы к сбору статистик (периодичность сбора, малые выборки, инкрементный сбор, использование рудиментарных индексов). Физическая оптимизация. Общая схема работы оптимизатора физических планов. Подходы к перебору физических планов. Использование эвристик. Алгоритмы реализации операторов физической алгебры. Сканирование и сортировка. Стоимостная модель для физических операторов. Итераторная и потоковая модели для организации выполнения операторов физической алгебры. Однопроходные и многопроходные алгоритмы реализации операций над базой данных. Двухпроходные алгоритмы на базе сортировки и хеширования. Алгоритмы, базирующиеся на использовании индексов. Генерация физического плана запроса. Выбор методов для селекции. Выбор методов для соединения. Конвейеризация против материализации. Конвейеризация унарных операций. Конвейеризация бинарных операций. Нотация для физических планов запроса. Упорядочивание физических операций. Управление буферным пулом. Архитектура менеджера буферного пула. Стратегии вытеснения страниц. Связь между выбором физических операторов и управлением буферным пулом. Параллельные алгоритмы для реляционных операций. Распараллеливание покортежных операций. Распараллеливание полнотабличных операций. Производительность параллельных алгоритмов.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине (ЗУНы)
ПК-4.1 знанием задач развития теории программирования, создания и сопровождения программных средств различного назначения (для направленности 05.13.11)	Знать: математические основы функционирования компилятора запросов.
	Уметь: проводить оценку стоимости запроса на основе оценок стоимости реляционных операций.
	Владеть: методами управления буферным пулом в СУБД.
ПК-4.2 умением повышать эффективность и надежность процессов обработки и передачи данных и знаний в вычислительных машинах, комплексах и компьютерных сетях (для направленности 05.13.11)	Знать: принципы работы оптимизатора физических планов запросов.
	Уметь: применять алгоритмы реализации операторов физической алгебры.
	Владеть: параллельными алгоритмами реализации реляционных операций.

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Нет	Научно-исследовательская деятельность (3 семестр)

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 трудоёмкость дисциплины составляет 3 з.е., 108 ч.

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		2
Общая трудоёмкость дисциплины	108	108
<i>Аудиторные занятия:</i>	40	40
Лекции (Л)	40	40
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	0	0
Лабораторные работы (ЛР)	0	0
<i>Самостоятельная работа (СРС)</i>	68	68
Углубленное изучение многопроходных алгоритмов выполнения запросов	68	68
Вид итогового 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	Обзор этапов обработки запроса. Компилятор запросов.	2	2	0	0
2	Алгебраические законы и планы запросов	4	4	0	0
3	Преобразование дерева разбора в логический план запроса	4	4	0	0
4	Оценка стоимости операций	4	4	0	0
5	Использование статистик	4	4	0	0
6	Физическая оптимизация	4	4	0	0
7	Алгоритмы реализации операторов физической алгебры	6	6	0	0
8	Генерация физического плана запроса	4	4	0	0
9	Управление буферным пулом	4	4	0	0
10	Параллельные алгоритмы для реляционных операций	4	4	0	0

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Синтаксический анализ и деревья разбора. Грамматика для простого подмножества SQL. Препроцессор: разрешение представлений VIEW; семантический анализ (верификация имен отношений, верификация имен атрибутов и их квалификация, контроль типов).	2
2	2	Коммутативный и ассоциативный законы. Законы для операции выборки. Законы для операции проекции. Законы для соединения и прямого произведения.	4
3	3	Конверсия в реляционную алгебру. Удаление подзапросов из условий. Логическая оптимизация: продвижение операций выборки и проекции «вниз» по дереву запроса; замена сочетания выборки и прямого произведения на соединение. Пример, указывающий на целесообразность первоначального продвижения выборки вверх.	4
4	4	Оценка размеров промежуточных отношений. Оценка размеров проекции. Оценка размеров выборки. Оценка размеров соединения. Естественное соединение с несколькими атрибутами соединения. Соединение нескольких отношений. Оценка размеров для других операций.	4
5	5	Виды статистик (тотальные статистики и гистограммы). Гистограммы полные и частичные. Виды частичных гистограмм (с равной шириной интервалов, с равной высотой интервалов, с наиболее часто встречающимися значениями). Подходы к сбору статистик (периодичность сбора, малые выборки, инкрементный сбор, использование рудиментарных индексов).	4
6	6	Общая схема работы оптимизатора физических планов. Подходы к перебору физических планов. Использование эвристик.	4
7	7	Сканирование и сортировка. Стоимостная модель для физических операторов. Итераторная и потоковая модели для организации выполнения операторов физической алгебры. Однопроходные и многопроходные алгоритмы реализации операций над базой данных. Двухпроходные алгоритмы на базе сортировки и хеширования. Алгоритмы, базирующиеся на использовании индексов.	6
8	8	Выбор методов для селекции. Выбор методов для соединения.	4

		Конвейеризация против материализации. Конвейеризация унарных операций. Конвейеризация бинарных операций. Нотация для физических планов запроса. Упорядочивание физических операций.	
9	9	Архитектура менеджера буферного пула. Стратегии вытеснения страниц. Связь между выбором физических операторов и управлением буферным пулом.	4
10	10	Распараллеливание покортежных операций. Распараллеливание полнотабличных операций. Производительность параллельных алгоритмов.	4

5.2. Практические занятия, семинары

Не предусмотрены

5.3. Лабораторные работы

Не предусмотрены

5.4. Самостоятельная работа студента

Выполнение СРС		
Вид работы и содержание задания	Список литературы (с указанием разделов, глав, страниц)	Кол-во часов
Углубленное изучение многопроходных алгоритмов выполнения запросов	Гарсиа-Молина Г., Ульман Дж., Уидом Дж. Системы баз данных. Полный курс. М.: Издательский дом "Вильямс", 2004. 1088 с. - гл. 15, разд. 15.8.1-15.8.4 (стр. 738-741).	68

6. Инновационные образовательные технологии, используемые в учебном процессе

Инновационные формы учебных занятий	Вид работы (Л, ПЗ, ЛР)	Краткое описание	Кол-во ауд. часов
Интерактивная лекция	Лекции	Коллективное обсуждение проблематики обработки запросов к сверхбольшим базам данных	36

Собственные инновационные способы и методы, используемые в образовательном процессе

Не предусмотрены

Использование результатов научных исследований, проводимых университетом, в рамках данной дисциплины: Изучение достижений научных школ ЮУрГУ в области технологий обработки запросов к сверхбольшим базам данных

7. Фонд оценочных средств (ФОС) для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины

7.1. Паспорт фонда оценочных средств

Наименование разделов дисциплины	Контролируемая компетенция ЗУНы	Вид контроля (включая текущий)	№№ заданий
Обзор этапов обработки запроса. Компилятор запросов.	ПК-4.1 знанием задач развития теории программирования, создания и сопровождения программных средств различного назначения (для направленности 05.13.11)	промежуточный, текущий	1,2
Алгебраические законы и планы запросов	ПК-4.2 умением повышать эффективность и надежность процессов обработки и передачи данных и знаний в вычислительных машинах, комплексах и компьютерных сетях (для направленности 05.13.11)	промежуточный, текущий	3
Преобразование дерева разбора в логический план запроса	ПК-4.1 знанием задач развития теории программирования, создания и сопровождения программных средств различного назначения (для направленности 05.13.11)	промежуточный, текущий	4
Оценка стоимости операций	ПК-4.1 знанием задач развития теории программирования, создания и сопровождения программных средств различного назначения (для направленности 05.13.11)	промежуточный, текущий	6
Физическая оптимизация	ПК-4.2 умением повышать эффективность и надежность процессов обработки и передачи данных и знаний в вычислительных машинах, комплексах и компьютерных сетях (для направленности 05.13.11)	промежуточный, текущий	7
Алгоритмы реализации операторов физической алгебры	ПК-4.1 знанием задач развития теории программирования, создания и сопровождения программных средств различного назначения (для направленности 05.13.11)	промежуточный, текущий	8
Параллельные алгоритмы для реляционных операций	ПК-4.2 умением повышать эффективность и надежность процессов обработки и передачи данных и знаний в вычислительных машинах, комплексах и компьютерных сетях (для направленности 05.13.11)	промежуточный, текущий	9

7.2. Виды контроля, процедуры проведения, критерии оценивания

Вид контроля	Процедуры проведения и оценивания	Критерии оценивания
промежуточный, текущий	Компьютерный тест в системе Moodle (портал edu.susu.ru)	Отлично: Не менее 75% правильных ответов Хорошо: От 65% до 75% правильных ответов Удовлетворительно: От 50% до 65% правильных ответов Неудовлетворительно: Менее 50% правильных ответов

7.3. Типовые контрольные задания

Вид контроля	Типовые контрольные задания
промежуточный, текущий	1. Этапы обработки запроса. Препроцессор. 2. Разбор запроса. 3. Синтаксические категории и атомы. 4. Базисные синтаксические категории.

	5. Оценка размеров промежуточных отношений. 6. Статистические характеристики данных. 7. Использование гистограмм для оценки размеров промежуточных отношений. 8. Алгоритмы соединения. 9. Алгоритм вложенных циклов на диске и в оперативной памяти. 10. Алгоритм соединения с хешированием. 11. Оценка эффективности алгоритмов соединения. Тест.pdf
--	--

8. 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. и ред. К. А. Птицына. - 8-е изд. - М. и др.: Вильямс, 2005. - 1327 с. ил.
2. Кузнецов, С. Д. Базы данных Текст учебник для вузов по направлению "Приклад. математика и информатика" С. Д. Кузнецов. - М.: Академия, 2012. - 490, [1] с. ил.

б) дополнительная литература:

Не предусмотрена

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

1. Вестник ЮУрГУ. Серия: Вычислительная математика и информатика.

г) методические указания для студентов по освоению дисциплины:

1. Соколинский Л.Б. Параллельные системы баз данных. М.: Издательство Московского университета, 2013. 184 с. ISBN 978-5-211-06482-9.

из них: учебно-методическое обеспечение самостоятельной работы студента:

Электронная учебно-методическая документация

№	Вид литературы	Наименование разработки	Наименование ресурса в электронной форме	Доступность (сеть Интернет / локальная сеть; авторизованный / свободный доступ)
1	Методические пособия для самостоятельной работы студента	Библиографический каталог по информационным технологиям	Электронный каталог ЮУрГУ	Интернет / Свободный

9. Информационные технологии, используемые при осуществлении образовательного процесса

Перечень используемого программного обеспечения:

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

Перечень используемых информационных справочных систем:

1. EBSCO Information Services-EBSCOhost Research Databases(бессрочно)
2. -База данных ВИНТИ РАН(бессрочно)

10. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Лекции	110 (3г)	проектор