

Санкт-Петербургский государственный университет

РАБОЧАЯ ПРОГРАММА

учебной дисциплины

Оптимизация баз данных

Database optimization

Язык(и) обучения

русский

Трудоёмкость (границы трудоёмкости) в зачетных единицах: 3

Регистрационный номер рабочей программы: 018997

Санкт-Петербург

2016

Раздел 1. Характеристики учебных занятий

1.1. Цели и задачи учебных занятий

Формирование у обучающихся устойчивого понимания основных принципов разработки программного обеспечения связанного с хранением и манипуляциями данными. Задачами курса являются развитие теоретических знаний, навыков и подходов к оптимизации баз данных.

1.2. Требования к подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)

Для изучения необходимы знания в области общей теории баз данных.

1.3. Перечень результатов обучения (learning outcomes)

После изучения курса обучающиеся должны:

- усвоить основные принципы организации баз данных,
- иметь представление о механизмах функционирования систем управления базами данных (СУБД),
- уметь выявлять причины снижения производительности и/или увеличение объёма хранимых данных и выбирать и применять методы устранения обнаруженных проблем,
- уметь выбирать подходящий инструментарий и средства.

1.4. Перечень активных и интерактивных форм учебных занятий

В процессе преподавания данной дисциплины используются как классические методы обучения (лекции), так и различные виды самостоятельной работы студентов по заданию преподавателя, которые направлены на развитие творческих качеств студентов.

Основой успешного изучения дисциплины является самостоятельная работа магистрантов.

В рамках данного курса используются такие интерактивные формы обучения (14 часов):

- интерактивные лекции и лекции-конференции;
- кейс-метод для овладения критериями выбора системы управления базами данных;
- метод групповой работы при решении задач во время практических занятий;
- экспертная оценка другими обучающимися выбранных методов и решений, направленных на оптимизацию процессов хранения и обработки данных в контексте заданной задачи.

При изложении части тем, по желанию лектора, применяется мультимедиа-проектор для проведения презентаций.

Раздел 2. Организация, структура и содержание учебных занятий

2.1. Организация учебных занятий

2.1.1. Основной курс

| Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся | | | | | | | | | | | | | | | | | | |
|---|--|----------|--------------|----------------------|---------------------|--------------------|-------------|------------------|--------------------------|---------------------|--------------------------------|-----------------------------|---|-----------------------------|---|--------------|-------------------------------------|--------------------------------|
| Период обучения (модуль) | Контактная работа обучающихся с преподавателем | | | | | | | | | | | Самостоятельная работа | | | Объём активных и интерактивных форм учебных занятий | Трудоёмкость | | |
| | лекции | семинары | консультации | практические занятия | лабораторные работы | контрольные работы | коллоквиумы | текущий контроль | промежуточная аттестация | итоговая аттестация | под руководством преподавателя | в присутствии преподавателя | сам.раб. с использованием методических материалов | текущий контроль (сам.раб.) | | | промежуточная аттестация (сам.раб.) | итоговая аттестация (сам.раб.) |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | | | | | | | | | | | | | | | | |
| очная форма обучения | | | | | | | | | | | | | | | | | | |
| Семестр 2 | 12 | | 2 | 14 | | | | 6 | 2 | | | | 42 | 6 | 24 | | 14 | 3 |
| ИТОГО | 12 | | 2 | 14 | | | | 6 | 2 | | | | 42 | 6 | 24 | | 14 | 3 |

| Формы текущего контроля успеваемости, виды промежуточной и итоговой аттестации | | | |
|--|--------------------------------------|-------------------------------|--|
| Период обучения (модуль) | Формы текущего контроля успеваемости | Виды промежуточной аттестации | Виды итоговой аттестации (только для программ итоговой аттестации и дополнительных образовательных программ) |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | |
| очная форма обучения | | | |
| Семестр 2 | собеседование | экзамен | |

2.2. Структура и содержание учебных занятий

Период обучения (модуль): **Семестр 2**

| № п/п | Наименование темы (раздела, части) | Вид учебных занятий | Количество часов |
|-------|--|----------------------------|------------------|
| 1 | Модуль 1. Классификация и обзор причин деградации производительности или препятствий к масштабированию | лекции | 4 |
| | | практические занятия | 4 |
| | | по методическим материалам | 12 |
| 2 | Модуль 2. Средства диагностики и тестирования запросов и окружения СУБД | лекции | 4 |
| | | практические занятия | 4 |
| | | по методическим материалам | 12 |
| 3 | Модуль 3. Методы и инструменты оптимизации | лекции | 4 |
| | | практические занятия | 6 |
| | | по методическим материалам | 18 |

Содержание учебных занятий

Модуль 1. Введение, планирование и проектирование баз данных

Темы лекций и практических занятий:

1.1 Роль и место баз данных в современных приложениях, функциональные возможности и использование различных СУБД для решения прикладных задач. Системы реляционных баз данных, системы клиент-сервер, интеграция информации. Краткий обзор структуры СУБД, механизмы транзакций и обработки запросов. Значимость вопросов эффективности и оптимизации баз данных.

1.2 Классификация аппаратных, программных и организационных проблем, приводящим к деградации производительности или препятствий к масштабированию. Обзор примеров и кейсов, характерных для одной или нескольких категорий.

Модуль 2. Средства диагностики и тестирования запросов и окружения СУБД

Темы лекций и практических занятий:

2.1 Обзор средств диагностики, работы СХД, СУБД, средства профилирования. Статистика СУБД, статистика и план исполнения запросов, анализ ресурсных затрат при выполнении запроса.

2.2 Разбор кейсов, упомянутых в модуле 1 с применением продемонстрированных инструментов диагностики и анализа. Обзор методов тестирования для приложений баз данных и запросов: юнит-тестирование, нагрузочное тестирование. Генерация тестовых данных.

Модуль 3. Управление транзакциями, обработка запросов

Темы лекций и практических занятий:

3.1 Цели нормализации. Избыточность данных и аномалии обновления. Влияние нормализации и избыточности на скорость доступа к данным и объёмам их хранения. Избыточность как метод оптимизации. Шардинг, партицирование и репликация.

3.2 Значимость вопросов управления транзакциями. Управление параллельным доступом. Методы упорядочивания. Теоретические и практические аспекты соблюдения требований ACID и влияния ограничений, описанных в теореме CAP на примере ряда современных СУБД.

3.2 Общие методы обработки запросов, декомпозиция запросов. Оценка стоимости операций реляционной алгебры. Вопросы использования курсоров. Обработка запросов в распределенных системах. Приложения баз данных, обзор топологий, классификация проблем, возникающих при разработке и эксплуатации приложений баз данных, задач, решаемых приложениями баз данных и их влияние на производительность и масштабируемость. Классификация видов и архитектур приложений. Обзор компромиссных решений – кэш, OLTP и OLAP, степень зависимости от схемы.

Раздел 3. Обеспечение учебных занятий

3.1. Методическое обеспечение

3.1.1. Методические указания по освоению дисциплины

Самостоятельная работа студентов в основном заключается в установке, настройке и апробации различных современных СУБД, инструментария для диагностики и профилирования, тестирования и оптимизации, ознакомление с рекомендуемыми статьями и материалами, изучение лекционного материала, учебников и учебных пособий.

Время и место самостоятельной работы выбираются студентами по своему усмотрению с учетом рекомендаций преподавателя.

Самостоятельную работу над дисциплиной следует начинать с изучения учебно-методического комплекса, который содержит основные требования к знаниям, умениям, навыкам. Необходимо также вспомнить рекомендации преподавателя, данные в ходе лекционных занятий или консультаций, затем приступить к изучению отдельных разделов и тем.

Подготовка к лекции заключается в следующем:

- внимательно изучить материал предыдущей лекции;
- целесообразно составить краткий конспект или схему, отображающую смысл и связи основных понятий данного раздела, включенных в него тем, а затем, полезно изучить выдержки из литературы;
- узнать тему предстоящей лекции (по тематическому плану, по материалам, размещенным в системе дистанционного обучения Blackboard);
- ознакомиться с учебным материалом по учебнику и учебным пособиям;
- записать возможные вопросы, которые вы зададите лектору на лекции.

Подготовка к практическим занятиям:

- выполнить практические задания домашней работы;
- внимательно изучить материал лекций, относящихся к данному семинарскому занятию, ознакомиться с учебным материалом по учебнику и учебным пособиям;
- выписать основные термины;
- уяснить, какие учебные элементы остались неясными и сформулировать вопросы, которые необходимо задать преподавателю на занятии или консультации;

готовиться можно индивидуально, парами или в составе малой группы, последние являются эффективными формами работы.

3.1.2. Методическое обеспечение самостоятельной работы

Презентации и набор рекомендуемых ресурсов сети Интернет, образцы баз данных.

3.1.3. Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания

Для прохождения промежуточной аттестации необходимо пройти собеседование по программе курса в части пройденного материала, отчётность по выполнению самостоятельных работ в виде доклада на практических занятиях.

Для получения итоговой оценки необходимо сдать экзамен по курсу.

Итоговая оценка "отлично" ставится при условии своевременной сдачи практических работ в течении семестра, а также безупречного ответа на экзаменационный билет и дополнительные вопросы.

Итоговая оценка "хорошо" ставится при условии сдачи в течении семестра более половины практических заданий, а также ответа не менее чем на половину вопросов экзаменационного билета и дополнительные вопросы.

Итоговая оценка "удовлетворительно" ставится при условии сдачи не менее трех практических работ за семестр, а также ответа на один вопрос из экзаменационного билета и некоторые дополнительные вопросы.

Преподаватель предоставляет информацию о задолженностях студента в аттестационную комиссию.

3.1.4. Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)

ВОПРОСЫ ТЕСТИРОВАНИЯ

1. Как данные хранятся в реляционной базе данных?
 - а) в виде отдельных сущностей с определенными атрибутами,
 - б) в виде плоских таблиц,
 - в) в виде неструктурированного массива.
2. Что такое Система Управления Базами Данных (СУБД)?
 - а) программа, либо комплекс программ, предназначенных для управления и хранения данными.
 - б) набор формальных требований определяющий последовательность доступа к данным,
 - в) типичная схема данных, созданная с использованием средств проектирования.
3. Зачем нужны транзакции?
 - а) для обеспечения атомарности изменений,
 - б) для контроля прав доступа,
 - в) для проведения оптимизации.
4. Что такое нормализация?
 - а) возрастающая сложность,
 - б) процесс выявления независимых сущностей и вынесение их в отдельные таблицы,
 - в) контроль времени.
5. Что такое блокировки?
 - а) механизм защиты разделяемых записей от конкурентного доступа,
 - б) отказ в обслуживании в следствии технической неисправности,
 - в) ограничение прав пользователя.
6. Что такое представление?
 - а) столбец таблицы определяющий первичный ключ,
 - б) механизм оптимизации запросов для реляционных БД,
 - в) запрос на выборку, хранящийся на сервере как отдельный объект.
7. Когда возникает дублирование данных?
 - а) для обеспечения избыточности,
 - б) в случае недостаточной нормализации,
 - в) для разрешения вопросов блокировки.
8. Возможно ли избежать дублирования данных в реляционной модели?
 - а) возможно,

- б) невозможно,
 - в) возможно с потерей производительности.
9. Что такое взаимная блокировка?
- а) наличие устаревших данных в таблице,
 - б) одновременное изменение связанных записей,
 - в) отсутствие результатов поиска.
10. Возможно ли обеспечить отсутствие взаимной блокировки?
- а) возможно,
 - б) невозможно,
 - в) возможно с обеспечением дублирования данных.
11. Что такое план выполнения запроса?
- а) Последовательность действий программиста для написания и исполнения запроса.
 - б) Последовательность действий, описанных на языке SQL.
 - в) Последовательность действий СУБД для выполнения SQL запроса.

ПРИМЕР САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Начальные действия

1. Ознакомиться с документом, описывающим формулы расчёта баланса и примером его изменения
2. Ознакомиться со скриптом создания базы данных
3. Ознакомиться с программными компонентами (функции, триггера)
4. Разработать и применить генератор тестовых данных
5. Разработать простой тест на корректность расчёта баланса

Ожидаемый результат - доступная для оптимизации БД с тестовыми данными

Задачи I уровня

1. Реализовать индексы, повышающие производительность операций вставки и изменения платежей без модификации программных компонент

Задачи II уровня

В исходной версии балансы обновляются в рамках транзакции целевого действия над платежом - создания или изменения. Предположим, что требования к производительности не выполняются при такой реализации.

Введём две роли пользователей: бухгалтер-оператор и бухгалтер-аналитик.

Оператор - занимается вводом и корректировкой платежей, не имеет доступа к данным о балансах.

Бухгалтер-аналитик - отслеживает балансы и заведённые платежи для принятия решения о предпринимаемых финансовых действиях (какие счета использовать, какие образовались долги и т.д.).

1. Дать оценку затрат на выполнения операций расчёта балансов в рамках транзакций создания и изменения платежа. Желательно представить количественную оценку, но допустимо и относительную (к примеру, "90% ресурсов и времени уходит на расчёт баланса"). Чем детальнее, тем лучше.
2. Предложить сценарий оптимизации механизмов расчёта. Сценарий должен допускать максимизацию скорости целевых изменений и допускать отложенное вычисление балансов (балансы и данные платежей должны быть согласованы в конечном счёте).
3. Оценить недостатки предлагаемого сценария с точки зрения потенциальных пользователей.

Задачи III уровня

1. Реализовать предложенный сценарий отложенного вычисления.
2. Дать оценку затрат на выполнение операций расчёта балансов в рамках транзакций создания и изменения платежа. Дать заключение о полученном росте производительности, если таковое будет наблюдаться.
3. Дать оценку возникшим проблемам и недостаткам, сравнить их с оценками, сделанными при проектировании сценария оптимизации (указать, что сбылось, что неожиданно проявилось и т.д.).
4. Дать заключение о преимуществах и недостатках выполненной оптимизации (превосходят ли полученные преимущества те недостатки, которые возникли после оптимизации).

3.1.5. Методические материалы для оценки обучающимися содержания и качества учебного процесса

Для оценки обучающимися содержания и качества учебного процесса используется анкета-отзыв установленная локальными актами СПбГУ.

3.2. Кадровое обеспечение

3.2.1. Образование и (или) квалификация преподавателей и иных лиц, допущенных к проведению учебных занятий

К преподаванию привлекаются преподаватели, имеющие ученую степень, а также главные и ведущие специалисты в этой области. Допускается проведение семинарских занятий обучающимися в аспирантуре (под руководством научного руководителя) для прохождения педагогической практики.

3.2.2. Обеспечение учебно-вспомогательным и (или) иным персоналом

Не требуется.

3.3. Материально-техническое обеспечение

3.3.1. Характеристики аудиторий (помещений, мест) для проведения занятий

Наличие компьютеризированных рабочих мест для проведения совместных лабораторных работ и демонстрации материалов курса (образцы баз данных).

3.3.2. Характеристики аудиторного оборудования, в том числе неспециализированного компьютерного оборудования и программного обеспечения общего пользования

Универсальные компьютеры, объединенные в локальную сеть, мультимедийное оборудование (проектор, экран). Системное программное обеспечение общего назначения (MS Windows или Linux любой современной версии).

3.3.3. Характеристики специализированного оборудования

Отсутствуют.

3.3.4. Характеристики специализированного программного обеспечения

Отсутствуют.

3.3.5. Перечень и объёмы требуемых расходных материалов

Фломастеры цветные, губки, бумага формата А4, канцелярские товары, картриджи для принтеров в объеме, необходимом для организации и проведения занятий, по заявкам преподавателей, подаваемым в установленные сроки.

3.4. Информационное обеспечение

3.4.1. Список обязательной литературы

1. Введение в системы баз данных: учебное пособие / К. Дж. Дейт; пер. с англ. и ред.: К. А. Птицын. – 8-е изд. – М.; СПб.; Киев: Издательский дом "Вильямс", 2005. – 1328 с.

3.4.2. Список дополнительной литературы

1. Семь баз данных за семь недель. Введение в современные базы данных и идеологию NoSQL / Э. Редмонд, Д. Уилсон; ред. Ж. Картер. – М.: ДМК Пресс, 2015. – 384 с.
2. Высоконагруженные приложения. Программирование, масштабирование, поддержка / Мартин Клепман – М.: Питер, 2018. – 640 с.
3. NoSQL: новая методология разработки нереляционных баз данных / М. Фаулер, Прамодкумар Дж. Садаладж; Пер. с англ. – М.: ООО "И.Д. Вильямс", 2017. – 192 с.: ил.
4. Настройка приложений баз данных / Новиков Б.А., Домбровская Г.Р. – СПб.: БХВ-Петербург, 2012. – 240 с.
5. СУБД для программиста. Базы данных изнутри / С. Тарасов, Издательство Соломон, 2015. – 320 с.
6. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени / Натан Марц, Джеймс Уоррен – М.: Вильямс, 2017. – 368 с.

7. Cassandra. Полное руководство. Распределенные данные в масштабе веба / Джефф Карпентер, Эбен Хьюитт. – М.: ДМК Пресс, 2017. – 400 с.
8. MongoDB в действии / Кайл Бэнкер. – М.: ДМК Пресс, 2017. – 394 с.
9. Системы баз данных. Полный курс / Гектор Гарсиа-Молина, Джеффри Д. Ульман, Дженнифер Уидом – М.: Вильямс, 2014. – 1088 с.
10. Основы будущих систем баз данных. Третий манифест: детальное исследование влияния теории типов на реляционную модель данных, включая полную модель наследования типов / К. Дж. Дейт, Х. Дарвен; пер. : С. Д. Кузнецов, Т. А. Кузнецова ; ред. : С. Д. Кузнецов. - 2-е изд. – М.: Янус-К, 2004. – 656 с.
11. Разработка приложений на основе Microsoft SQL Server 2008/ учебное пособие / Л. Лобел, Э. Дж. Браст, С. Форте; пер. с англ. А. Лашкевич. – М.: Русская редакция; СПб.: БХВ-Петербург, 2010. – 1024 с.
12. Теория и практика построения баз данных: учебное пособие / Д. Кренке; пер. с англ.: А. Вахитов. – 9-е изд. – СПб.: Питер, 2005. – 859 с.

3.4.3. Перечень иных информационных источников

1. <https://db-engines.com/en/>
2. <http://nosql-database.org/>

Нагрузочное тестирование для определения вертикальной масштабируемости

1. [Тестирование производительности баз данных при помощи tSQLt и SQLQueryStress](#)
2. [The SQL Server Sqlosim Utility](#)
3. [Нагрузочное тестирование базы данных. ContiPerf + DBUnit](#)

Индексы

1. [A site explaining SQL indexing to developers](#)
2. [Вопросы про индексы, которые вам не надо будет задавать](#)
3. [14 вопросов об индексах в SQL Server, которые вы стеснялись задать](#)

Index fragmentation

1. <http://www.sql-server-performance.com/2008/analyze-and-fix-index-fragmentation-in-sql-server-2008/>
2. <http://www.community.terrasoft.ru/blogs/8318>
3. <http://technet.microsoft.com/ru-ru/library/ms189858.aspx>
4. <https://ola.hallengren.com/sql-server-index-and-statistics-maintenance.html>

DMV для работы с индексами:

1. <http://blogs.msdn.com/b/bartd/archive/2007/07/19/are-you-using-sql-s-missing-index-dmvs.aspx>
2. <http://sqlserverplanet.com/dmvs/find-index-usage>

Индексированные представления

1. [SQL Server Indexed Views: The Basics](#)

Влияние сортировки значений в индексе (ASC | DESC)

1. [Descending indexes](#)
2. [Specifying ascending or descending order for index keys](#)
3. [Choosing between an ascending or descending index](#)
4. [Building SQL Server Indexes in Ascending vs Descending Order](#)

In-Memory database

1. [За счет чего Tarantool такой оптимальный](#)
2. [In-Memory OLTP в SQL Server 2016: Архитектура, реализация и возможности](#)
3. [In-Memory OLTP в SQL Server 2016: Практическое использование](#)
4. [Сравнение аналитических in-memory баз данных](#)
5. [Как выбрать In-memory NoSQL базу данных с умом. Тестируем производительность](#)

Диагностика

1. [Basic SQL Server Performance Troubleshooting For Developers](#)
2. [Managing SQL Server Statistics](#)
3. [Мониторинг Microsoft SQL Server «на коленке»](#)
4. [Реализация индикатора производительности запросов, хранимых процедур и триггеров в MS SQL Server. Автотрассировка](#)
5. [Using Optimizer WhatIf and StatsStream to Simulate a Production Environment](#)

Общие вопросы

1. [Improving the Quality of SQL Server Database Connections in the Cloud](#)
2. [Lightweight Tables или практические советы при проектировании БД](#)
3. [Applying The Scientific Method to SQL Server Performance Tuning](#)
4. [SQL Server Internals Party: Storage, Memory and the Query Life-cycle](#)
5. [Практическая оптимизация MySQL: измерять, чтобы ускорять](#)

Статьи по теме

1. [Uber — причины перехода с Postgres на MySQL](#)
2. [DevConf: переход Uber с PostgreSQL на MySQL](#)

Практикум "Исследование объёмов данных"

Генерация данных

Generating Test Data in TSQL

<http://habrahabr.ru/post/220185/>

Полезные инструменты

1. [SQL Swiss Army Knife Series](#)
2. [sp_BlitzIndex® – SQL Server's Index Sanity Test](#)
3. [How to analyse SQL Server performance](#)

4. [Exploring Your SQL Server Databases with T-SQL](#)
5. <http://whoisactive.com/>
6. [sys.dm_exec_sessions](#)

Раздел 4. Разработчики программы

| Фамилия, имя, отчество | Учёная степень | Учёное звание | Должность | Контактная информация (служебный адрес электронной почты, служебный телефон) |
|-------------------------|----------------|---------------|-----------------------|---|
| Севрюков Сергей Юрьевич | | | старший преподаватель | s.sevryukov@spbu.ru |