**Программа учебной дисциплины «Перспективные системы   
управления базами данных»**

Утверждена

Академическим советом ООП

Протокол № от «\_\_»\_\_\_\_\_20\_\_ г.

|  |  |
| --- | --- |
| Автор | Кузнецов Сергей Дмитриевич |
| Число кредитов |  |
| Контактная работа (час.) |  |
| Самостоятельная работа (час.) |  |
| Курс |  |
| Формат изучения дисциплины |  |

1. **ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ**

Целями освоения дисциплины «Перcпективные системы управления базами данных» являются изучение основных направлений развития технологии систем управления данными, методов и алгоритмов организации перспективных систем управления базами данных.

В результате освоения дисциплины студент должен:

**знать:**

- основные виды используемых в настоящее время и ожидаемых в будущем систем управления базами данных (СУБД);

- основные структуры данных, используемые в SQL-ориентированных СУБД;

- принципы, методы и средства управления транзакциями;

- методы выполнения и оптимизации запросов;

- принципы организации параллельных СУБД.

**уметь:**

- квалифицированно выбирать и применять СУБД в качестве среды приложений баз данных;

- настраивать приложения в соответствии со спецификой конкретной базы данных и СУБД;

- при необходимости расширять и/или дорабатывать СУБД;

- участвовать в проектах по разработке новых систем управления данными.

**владеть:**

- фундаментальной подготовкой в области управления данными;

- методами организации структур данных, требуемыми в СУБД;

- методами управления транзакциями;

- методами выполнения и оптимизации запросов.

Изучение дисциплины «Физика» базируется на следующих дисциплинах:

- математическая логика и дискретная математика;

- программирование;

- архитектура вычислительных систем;

- операционные системы;

- основы организации баз данных.

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

* знать основы математической логики;
* уметь программировать;
* знать основные архитектуры компьютеров;
* знать принципы построения операционных систем;
* знать основы организации баз данных.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин:

1. Параллельные СУБД.
2. Организация информационных систем.
3. **СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ**

**Тема 1. Классификация СУБД. Роль и место SQL-ориентированных СУБД**

Разные подходы к классификации СУБД: по используемой модели данных (реляционные, SQL-ориентированные, объектно-ориентированные и т.д.); транзакционные и аналитические; универсальные и специализированные; файл-серверные, клиент-серверные и встраиваемые; дисковые и сохраняющие данные в основной памяти (in-memory); однопроцессорные, параллельные с общей памятью (shared everything), параллельные с общими дисками (shared disks), параллельные без общих ресурсов (shared nothing); проприетарные и с открытыми кодами. Актуальность SQL-ориентированных СУБД, общая архитектура и основные функции таких СУБД.

**Тема 2. Структуры данных в SQL-ориентированной базе данных**

Способы хранения таблиц SQL-ориентированных СУБД по строкам и столбцам, преимущества и недостатки разных способов; наиболее распространенные индексные структуры.

**Тема 3. Управление транзакциями**

Понятия ACID-транзакции и сериализации транзакций. Основные подходы к сериализации ACID-транзакций: двухфазный протокол синхронизационных блокировок и метод временных меток. Базовые версионные алгоритмы. Понятие распределенной транзакции и методы фиксации таких транзакций.

**Тема 4. Методы выполнения «реляционных» операций, компиляция и построение планов выполнения операций SQL**

Общие подходы к обработке запросов и операций обновления баз данных: методы выполнения «реляционных» операций соединения, группировки и т.д., компиляция и построение планов выполнения операций SQL.

**Тема 5. Оптимизация запросов в SQL-ориентированных СУБД**

Методы перезаписи запросов, методы оценки стоимости планов выполнения запросов, методы оценки селективности предикатов, оценки распределений значений в столбцах таблиц, методы оптимизации запросов с многочисленными соединениями и т.д.

**Тема 6. Параллельные СУБД категории shared everything**

Аппаратно-программные архитектуры СУБД, обеспечивающие распараллеливание операций над базами данных. Параллельные СУБД категории shared everything, имеющиеся решения (например, в Oracle), особенности компиляции и оптимизации запросов, идеи, позволяющие добиваться горизонтальной масштабируемости при росте числа ядер.

**Тема 7. ПараллельныеСУБД категории shared disks**

Подходы к построению параллельных СУБД категории shared disks, имеющиеся решения (Oracle RAC, Tibero), специальные требования к аппаратуре.

**Тема 8. Параллельные СУБД категории shared nothing**

Параллельные СУБД категории shared nothing, в том числе, аналитические СУБД (Vertica, Greenplum, Asterdata), особенности компиляции и оптимизации запросов, методы выполнения распределенных запросов, методы хранения и разделения данных, методы параллельного серверного программирования, а также транзакционные СУБД, ориентация на высокий уровень доступности или на высокий темп обработки транзакций, методы уменьшения числа распределенных транзакций при заданной рабочей нагрузке. Подходы к организации систем обработки данных на основе совместного использования технологий баз данных и распределенных систем: аналитика и MapReduce, Hadoop и OpenStack, «транзакционные» СУБД категории NoSQL, «теорема» CAP и ее следствие – транзакции BASE (Basically Available, Soft-state, Eventual consistency – доступность в большинстве случаев), неустойчивое состояние, согласованность в конечном счете и другие ослабленные модели транзакций, NoSQL в мире SQL: применение теоремы CAP при реализации параллельных СУБД без использования общих ресурсов, поддерживающих высокий уровень доступности.

**Тема 9. Влияние новых видов аппаратных средств хранения**

Влияние на общую архитектуру СУБД, структуры данных и алгоритмы выполнения операций и оптимизацию запросов потенциального перехода к использованию новых видов аппаратных средств хранения данных – флеш-памяти (твердотельных дисков) и энергонезависимой основной памяти.

1. **ОЦЕНИВАНИЕ**

**3.1 Критерии оценки знаний, навыков**

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

Порядок формирования оценок по дисциплине

Преподаватель оценивает работу студентов на практических занятиях и выполненные ими домашние задания. Оценка за работу студента на практических занятиях и в рамках домашних заданий выставляется в рабочую ведомость. Полученная оценка по 10-ти балльной шкале определяется перед итоговым контролем - Оауд.

Преподаватель оценивает контрольную работу студентов. Оценка за контрольную работу студента выставляется в рабочую ведомость. Оценка по 10-ти балльной шкале за контрольную работу – Оконтр. работа.

Накопленная оценка за текущий контроль учитывает результаты студента по текущему контролю следующим образом:

Онакопл. = 0,5\* Оауд + 0,5\*Оконтр.работа

В диплом выставляет результирующая оценка по учебной дисциплине, которая формируется по следующей формуле:

Орезульт = 0,5\*Онакопл + 0,5\*Оэкз

1. **ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ**

**4.1 Тематика заданий текущего контроля**

1 Области применения встраиваемых систем управления базами данных

2. Методы организации мультидоступа к B-деревьям

3. Способы избежать появления распределенных транзакций

4 Алгоритм построения n-мерного куба

5. Суть метода перезаписи на основе «магических множеств»

6. Особенности оптимизации запросов в системах с общей памятью

7. Построение распределенного буфера в системах с общими дисками

8. Методы разделения данных в системах без общих ресурсов.

**4.2 Примеры контрольных вопросов для экзамена**

1. Общая архитектура СУБД, интерфейсы, связь с моделями данных
2. Файл-серверные, клиент-серверные и встраиваемые СУБД
3. Дисковые и “in memory” СУБД
4. Архитектуры shared-everything, shared-disks и shared-nothing
5. Понятие модели данных
6. Дореляционные модели данных
7. Классическая реляционная модель данных Кодда
8. Реляционная алгебра Кодда
9. Объектно-ориентированная модель данных: литеральные и объектные типы
10. Объектно-ориентированная модель данных: связи, экстенты, ключи, наследование
11. Объектно-ориентированная модель данных: объекты и OID. Что такое «литеральное множество объектов»? Как реализуются связи? Откуда вся эта путаница?
12. Объектно-ориентированная модель данных: язык OQL
13. Модель данных SQL: встраиваемые, параметризуемые и определяемые типы данных
14. Модель данных SQL: неопределенные значения и особенности булевского типа данных
15. Модель данных SQL: типы коллекций. Что дает тип множества в совокупности с анонимным строчным типом?
16. Модель данных SQL: индивидуальные типы данных, структурные типы, определяемые пользователями, ссылочные типы, наследование типов
17. Модель данных SQL: определение базовых хранимых таблиц. Типизированные таблицы, наследование типизированных таблиц
18. Чем похожи и чем различаются экстенты объектно-ориентированной модели данных и типизированные таблицы модели данных SQL? Как в этих моделях работает разыменование ссылок?
19. Модель данных SQL: общий вид и семантика оператора выборки
20. Модель данных SQL: логические выражения раздела WHERE оператора SELECT. Предикаты сравнения, BETWEEN, IS NULL, IN
21. Модель данных SQL: предикаты LIKE, SIMILAR, EXISTS, квантифицированного сравнения
22. Модель данных SQL: ограничения целостности, немедленно проверяемые и откладываемые ограничения, особенности ссылочных ограничений
23. Истинно реляционная модель данных: скалярные типы данных, реальное и возможные представления, операция выбора значения
24. Истинно реляционная модель данных: кортежные типа и типы отношений
25. Истинно реляционная модель данных: представление отсутствующей информации
26. Истинно реляционная модель данных: наследование
27. Хранение таблиц по строкам
28. Хранение таблиц по столбцам
29. Организация индексов на основе B+-деревьев, балансировка
30. Упреждающее расщепление, откладываемое слияние, переливание, B-деревья с составными ключами
31. Мультидоступ к B-деревьям
32. Интерфейс RSS
33. Общие принципы хэширования
34. Расширяемое хэширование
35. Линейное хэширование
36. Хранение журнальной и служебной информации
37. Свойства ACID-транзакций
38. Виды конфликтов транзакций и уровни изолированности
39. Сериализация транзакций
40. Синхронизационные блокировки, протокол 2PL
41. Гранулированные блокировки
42. Предикатные блокировки
43. Синхронизационные тупики, их обнаружение и разрушение
44. Метод временных меток
45. Версионные алгоритмы сериализации транзакций: версионные вариант метода временных меток
46. Версионные алгоритмы сериализации транзакций: версионные вариант методп 2PL
47. Версионные алгоритмы сериализации транзакций: алгоритм для поддержки только читающих транзакций
48. Управление буферным пулом в основной памяти
49. Физическая синхронизация
50. Протокол WAL и его связь с буферизацией
51. Индивидуальные откат транзакций
52. Общая схема восстановления после мягкого сбоя
53. Восстановление физической согласованности баз данных
54. Общая схема восстановления после жесткого сбоя
55. Архивация базы данных, сжатие логического журнала
56. Путь обработки запроса в SQL-ориентированной СУБД
57. Логическая оптимизация запросов: приведение запросов к каноническому представлению
58. Логическая оптимизация запросов: изменение порядка выполнения реляционных операций
59. Логическая оптимизация запросов: приведение запроса к алгебраическому представлению
60. Семантическая оптимизация запросов: подстановка представлений и ограничений целостности
61. Семантическая оптимизация запросов: чем отличается от простой подстановки?
62. Выбор и оценка альтернативных планов выполнения запросов: алгоритмы выполнения соединений
63. Выбор и оценка альтернативных планов выполнения запросов: степень селективности простого предиката
64. Предположения System R, оценка селективности предикатов и стоимости планов запросов
65. Способы определения истинных распределений данных, гистограммы
66. Определение селективности предикатов и стоимости планов при наличии гистограмм
67. Оптимизаторы с фиксированными стратегиями
68. Оптимизаторы гибкой структурой, пример Starburst

**5. РЕСУРСЫ**

**5.1 Базовый учебник**

С.Д. Кузнецов. Базы данных. Академия, Серия: Университетский учебник, 2012 г., 496 стр.ISBN 978-5-7695-8430-5

**5.2 Основная литература**

Гектор Гарсиа-Молина, Джеффри Д. Ульман, Дженнифер Уидом. Системы баз данных. Полный курс, изд-во "Вильямс", 2003

К. Дж. Дейт. Введение в системы баз данных. Восьмое издание. изд-во "Вильямс", 2008

**5.3 Дополнительная литература**

Abraham Silberschatz, Henry F. Korth, S. Sudarshan. Database System Concepts. McGraw-Hill Education; 6 edition, 2010

Jim Gray, Andreas Reuter. Transaction Processing: Concepts and Techniques. Morgan Kaufmann; 1 edition (September 15, 1992)

V. Markl, G. M. Lohman, V. Raman. LEO: An autonomic query optimizer for DB2. IBM SYSTEMS JOURNAL, VOL 42, NO 1, 2003

Matthias Jarke, Jurgen Koch. Query Optimization in Database Systems. Computing Surveys, Vol. 16, No. 2, June 1984

M. Stonebraker, D. Abadi, A. Batkin, X. Chen, M. Cherniack, M. Ferreira, E. Lau, A. Lin, S. Madden, E. O’Neil, P. O’Neil, A. Rasin, N. Tran, and S. Zdonik. [C-Store: A Column-oriented DBMS.](http://web.mit.edu/dna/www/vldb.pdf) In Proc. VLDB, 2005.

M. Stonebraker and U. Cetintemel. [One Size Fits All: An Idea whose Time has Come and Gone.](http://www.cs.brown.edu/%7Eugur/fits_all.pdf) In Proc. ICDE, 2005.

[Michael Stonebraker](mailto:stonebraker@csail.mit.edu), [Samuel Madden](mailto:madden@csail.mit.edu), [Daniel J. Abadi](mailto:dnastavros@csail.mit.edu), [Stavros Harizopoulos](mailto:stavros@csail.mit.edu), [Nabil Hachem](mailto:nhachem@agdba.com), [Pat Helland](mailto:phelland@microsoft.com). [The End of an Architectural Era (It's Time for a Complete Rewrite)](http://web.mit.edu/dna/www/abadirdf.pdf). Proceedings of VLDB, 2007, Vienna, Austria

Ippokratis Pandis, Ryan Johnson, Nikos Hardavellas, Anastasia Ailamaki. Data-Oriented Transaction Execution. 36th International Conference on Very Large Data Bases, September 13-17, 2010, Singapore. Proceedings of the VLDB Endowment, Vol. 3, No. 1, 2010, pp. 928-939

Сергей Кузнецов. MapReduce: внутри, снаружи или сбоку от параллельных СУБД?, Труды Института системного программирования, т. 19, М., ИСП РАН, 2010, стр. 35-40.

Сергей Кузнецов. Транзакционные параллельные СУБД: новая волна. Труды Института системного программирования, т. 20, М., ИСП РАН, 2011, стр. 189-251.

С.Д. Кузнецов, А.А. Прохоров. Алгоритмы управления буферным пулом СУБД при работе с флэш-накопителями. Труды Института системного программирования, т. 23, М., ИСП РАН, 2012, стр. 173-194.

**5.4 Программное обеспечение**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Наименование** | **Условия доступа** |
| 1. | Microsoft Windows 7 Professional RUS  Microsoft Windows 10  Microsoft Windows 8.1 Professional RUS | *Из внутренней сети университета (договор)* |
| 2. | Microsoft Office Professional Plus 2010 | *Из внутренней сети университета (договор)* |

* 1. **Профессиональные базы данных, информационные справочные системы,   
     интернет-ресурсы (электронные образовательные ресурсы)**

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Наименование** | **Условия доступа** |
|  | ***Профессиональные базы данных, информационно-справочные системы*** | |
| 1. | Консультант Плюс | *Из внутренней сети университета (договор)* |
| 2. | Электронно-библиотечная система Юрайт | URL: https://biblio-online.ru/ |
|  | ***Интернет-ресурсы (электронные образовательные ресурсы)*** | |
| 1. | Открытое образование | URL: https://openedu.ru/ |

* 1. **Материально-техническое обеспечение дисциплины**

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);

мультимедийный проектор с дистанционным управлением.

Учебные аудитории для лабораторных и самостоятельных занятий по дисциплине оснащены ­­­­­­­­­­­­­­­­­­­­­­­­ \_\_\_\_\_\_\_\_\_\_\_\_\_\_, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.