УТВЕРЖДАЮ

Зав. кафедрой СП ЮУрГУ

\_\_\_\_\_\_\_\_\_Л.Б. Соколинский

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2022 г.

Фонд оценочных средств ООП «Искусственный интеллект и инженерия данных» по направлению 09.04.04 – Программная инженерия

Дисциплина «Методы и системы обработки больших данных»

| **№ КМ** | **Вид КМ** | **Наименование КМ** | **Оценочные средства** |
| --- | --- | --- | --- |
|  | Текущий контроль | Тест 1 | Пример теста:  1) выполнить реляционную операцию    2) написать выражение реляционной алгебры, вычисляющее ответ на запрос к базе данных "Поставки"    3) написать выражение реляционной алгебры, соответствующее логическому плану    Вопросы для самоконтроля:  1. Реляционные операции  2. Реляционная алгебра и язык SQL  3. Этапы обработки запроса  4. Разбор запроса  5. Функции препроцессора  6. Логическая оптимизация  7. Организация системы базы данных  8. Индексы  9. Оценка стоимости операций  10. Статистические характеристики данных  11. Алгоритмы соединений  12. Порядок соединений |
|  | Текущий контроль | Тест 2 | Пример теста:  1) построить дерево разбора для SQL-запроса    2) конвертировать дерево разбора в логический план    3) выполнить оптимизацию логического плана |
|  | Текущий контроль | Тест 3 | Пример теста:  1) вычислить количество блоков, необходимых для хранения отношения    2) вычислить количество блоков, необходимое для хранения индексного файла    3) вычислить оценку для размера результата реляционного выражения. |
|  | Текущий контроль | Тест 4 | Пример теста:  1) вычислить оценку размера естественного соединения отношений с использованием гистограмм    2) вычислить оценку эффективности алгоритма соединения    3) с использованием метода динамического программирования определить оптимальный порядок соединений |
|  | Промежуточный контроль | Итоговый тест | **Вопросы для подготовки к зачету по дисциплине:**   1. Общая схема обработки запроса в реляционной СУБД. Реляционная алгебра и реляционные операции. Представление реляционного выражения в виде логического плана. Общая схема компиляции запроса. 2. Разбор запроса. Синтаксический анализ запроса. Грамматика SQL. Построение дерева разбора. Разрешение представлений. Верификация запроса. 3. Конверсия запроса в логический план. 4. Логическая оптимизация. Алгебраические законы. Оптимизация операций выборки, проекции и удаления дубликатов. Оптимизация композиции операций выборки и прямого произведения. 5. Система хранения данных. Поля, записи, блоки. Использование буферного пула. 6. Индексы. В-деревья. 7. Оптимизация физического плана. Оценка стоимости реляционных операций. Статистические характеристики данных. Гистограммы. Выбор порядка соединений. 8. Исполнитель запросов. Итераторы. Алгоритмы соединений. 9. Модель параллельной обработки больших данных MapReduce. 10. Скетчинг данных. Фильтры Блума. Count-Min скетч. Определение числа уникальных элементов с помощью HyperLogLog.   **Типовые вопросы теста**  1) Пронумеровать перечисленные артефакты (структуры) в порядке их появления при обработке запроса:   * приведенное дерево разбора * дерево разбора * физический план * текст запроса на SQL * логический план   2) На каком этапе строится приведенное дерево разбора?   * Лексический анализ * Разрешение представлений и верификация запроса * Построение логического плана * Оптимизация запроса * Построение физического плана   3) Заполнить терминальные символы в приведенном дереве разбора, соответствующем следующему запросу:  /\* Таблица "Красные детали" \*/  CREATE VIEW КД AS  SELECT Код\_Д  FROM Д  WHERE Цвет='Красный';  /\* Выдать коды поставщиков, поставляющих красные детали" \*/  SELECT Код\_П  FROM ПД  WHERE Код\_Д IN (КД);  4) Отметить буквой S синтаксические категории, не являющиеся базовыми;  Отметить буквой B базовые синтаксические ка-тегории;  Отметить буквой A атомы.  <Attribut>  <  SELECT  <Relat>  AND  T  <Condition>  5) Отметить выражения, допускаемые грамматикой:  <А> ::= <B>=<B>  <B> ::= (<B>)  <B> ::= <B> + <B>  <B> ::= <B> - <B>  <B> ::= 0 | 1 | 2 | 3 | 4  2  2=2  2=-2  1+2=3  1-3=-2  2-2=4  2+3=5  6) Пусть заданы следующие схемы отношений: R(A,B); S(B\*,С); U(B,D). Пусть отношение R может быть целиком размещено в оперативной памяти (ОП). Пусть отношения S и U слишком большие, чтобы поместиться в ОП. Известно, что от-ношения S и U отсортированы по убыванию значения атрибута B. Также известно, что отношение R не отсортировано по атрибуту B.  Отметить алгоритм соединения, который в данном случае будет оптимальным для выполнения следующего логического плана: R!S!U  \*Первичный ключ   * Соединение вложенными циклами * Соединение слиянием (с сортировкой R) * Соединение хешированием * Соединение по индексу   7)    8) |

Паспорт фонда оценочных средств приведен в п. 6.3 РПД.

Разработчик Л.Б. Соколинский