

# Query languages

## Assignment sheet 4

### 1 One pass $\gamma$ operation

Consider the following relation schemas:

```
Movie(title,year,length,studioName)
StarsIn(title,year,starName),
```

whose instances are, respectively:

```
Movie={ (a,b,c,d), (e,f,g,h), (i,j,k,l) }
StarsIn={ (a,b,x), (a,b,y), (e,f,x), (i,j,x) }.
```

Consider the following query:

```
SELECT starName, SUM(length)
FROM Movie NATURAL JOIN StarsIn
GROUP BY starName HAVING COUNT(*) >= 3;
```

Assuming that the memory is large enough and that it contains the result of the join, propose an in-memory search structure for the one pass  $\gamma$  operation and show how the structure changes when processing this operation over the instances above.

### 2 Operation costs and requirements

Suppose  $B(R) = B(S) = 10,000$ . What are the memory requirements and I/O costs of the following operations:

1. One pass  $\cup_S$
2. Two pass  $\cup_S$  based on sorting
3. Two pass  $\delta$  based on sorting
4. Two pass  $\gamma$  based on sorting

### 3 Two pass operations based on sorting (1)

Suppose tuples are integers, one block can hold at most 2 tuples and  $M = 4$ . Show how the following two pass operators behave in the following cases:

1.  $\delta$  on the sequence 0,1,2,3,4 repeated 6 times
2.  $\gamma_{a,avg(b)}(R)$ , where the instance of  $R(a,b)$  has 30 tuples  $t_0$  to  $t_{29}$ , with tuple  $t_i = (modulo(i,5),i)$

Answer the same questions for  $M = 3$ .

### 4 Two pass operations based on sorting (2)

Suppose that the second pass of an operator based on sorting does not use all of the  $M$  available memory blocks, because the instance to be sorted will take less than  $M$  lists. How to save I/O's by using the available blocks?

### 5 Two pass operations based on hashing (1)

Suppose tuples are integers, one block can hold at most 2 tuples and  $M = 4$ . Show how the two pass  $\delta$  based on hashing behaves on the sequence 0,1,2,3,4 repeated 6 times, assuming that the hash function is  $h(i) = i \bmod 5$ .

### 6 Two pass operations based on hashing (2)

How to modify the implementation of  $\gamma$  if on the one hand  $B(R) \leq M^2$ , and on the other hand the number of groups is so small that the size of the groups is greater than  $M$ ?

### 7 Hybrid hashing

Suppose that one or more buckets can be kept in memory. In the case of the  $\delta$  operation, show how this can help to save I/O's.