Query Execution: Part 9

Review

The Measures:

- **I/O cost**: the number of disk accesses needed for the algorithm. This number *does not include* any disk accesses required to produce/store final output of the algorithm. It does include any disk write operations necessary to store intermediate information on disk.
- **Memory**: the largest number of memory buffers that can be occupied by the data during the execution of the algorithm.
- **Constraint**: the restriction (typically on the sizes of input relations) which guarantees that the algorithm is feasible/applicable.

The Parameters:

- M: size of the main memory buffer space.
- $B(R)$: number of disk blocks used to store relation R on disk.
- $T(R)$: number of tuples in relation R.
- $V(R, A_1, \ldots, A_k)$: number of unique value combinations for attributes $A_1, \ldots, A_k$ of relation R.

Selection

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, tuple-at-a-time	`OnePassSelection`	$B(R)$	$O(1)$	**none**	
Index-based		$\frac{B(R)}{V(R,A)}$	$O(1)$	**none**	clustered relation, index on selection attribute
		$\frac{T(R)}{V(R,A)}$	$O(1)$	**none**	unclustered relation, index on selection attribute

Projection

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, tuple-at-a-time	`OnePassProjection`	$B(R)$	$O(1)$	**none**	

Duplicate Elimination

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, full-relation	`OnePassDuplicateElimination`	$B(R)$	$O(M)$	$B(\delta(R)) \leq M$	
two-pass, sort-based		$3B(R)$	$O(M)$	$B(R) \leq M^2$	
two-pass, hash-based		$3B(R)$	$O(M)$	$B(R) \leq M^2$	
multipass, sort-based		$(2k-1)B(R)$	$O(M)$	$B(R) \leq M^k$	
multipass, hash-based		$(2k-1)B(R)$	$O(M)$	$B(R) \leq M^k$	

Grouping and Aggregation

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, full-relation	`OnePassGrouping`	$B(R)$	$O(M)$	$B(\gamma_L(R)) \leq M$	
two-pass, sort-based		$3B(R)$	$O(M)$	$B(R) \leq M^2$	
two-pass, hash-based		$3B(R)$	$O(M)$	$B(R) \leq M^2$	
multipass, sort-based		$(2k-1)B(R)$	$O(M)$	$B(R) \leq M^k$	
multipass, hash-based		$(2k-1)B(R)$	$O(M)$	$B(R) \leq M^k$	

Bag Union

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, full-relation	`OnePassBagUnion`	$B(R)+B(S)$	$O(1)$	**none**	

Set Union

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, full-relation	`OnePassSetUnion`	$B(R)+B(S)$	$O(M)$	$B(S) \leq M$	
two-pass, sort-based		$3B(R)$	$O(M)$	$B(S)+B(R) \leq M^2$	
two-pass, hash-based		$3B(R)$	$O(M)$	$\min(B(S), B(R)) \leq M^2$	
multipass, sort-based		$(2k-1)B(R)$	$O(M)$	$B(R)+B(S) \leq M^k$	
multipass, hash-based		$(2k-1)B(R)$	$O(M)$	$\min(B(R), B(S)) \leq M^k$	

Bag Intersection

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, full-relation	`BagIntersection`	$B(R)+B(S)$	$O(M)$	$B(S) \leq M$	
two-pass, sort-based		$3B(R)$	$O(M)$	$B(S)+B(R) \leq M^2$	
two-pass, hash-based		$3B(R)$	$O(M)$	$\min(B(S), B(R)) \leq M^2$	
multipass, sort-based		$(2k-1)B(R)$	$O(M)$	$B(R)+B(S) \leq M^k$	
multipass, hash-based		$(2k-1)B(R)$	$O(M)$	$\min(B(R), B(S)) \leq M^k$	

Set Intersection

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, full-relation	`SetIntersection`	$B(R) + B(S)$	$O(M)$	$B(S) \leq M$	
two-pass, sort-based		$3B(R)$	$O(M)$	$B(S) + B(R) \leq M^2$	
two-pass, hash-based		$3B(R)$	$O(M)$	$\min(B(S), B(R)) \leq M^2$	
multipass, sort-based		$(2k-1)B(R)$	$O(M)$	$B(R) + B(S) \leq M^k$	
multipass, hash-based		$(2k-1)B(R)$	$O(M)$	$\min(B(R), B(S)) \leq M^k$	

Bag Difference

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, full-relation	`BagDifference`	$B(R) + B(S)$	$O(M)$	$B(S) \leq M$	
two-pass, sort-based		$3B(R)$	$O(M)$	$B(S) + B(R) \leq M^2$	
two-pass, hash-based		$3B(R)$	$O(M)$	$\min(B(S), B(R)) \leq M^2$	
multipass, sort-based		$(2k-1)B(R)$	$O(M)$	$B(R) + B(S) \leq M^k$	
multipass, hash-based		$(2k-1)B(R)$	$O(M)$	$\min(B(R), B(S)) \leq M^k$	

Set Difference

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, full-relation	`SetDifference`	$B(R) + B(S)$	$O(M)$	$B(S) \leq M$	
two-pass, sort-based		$3B(R)$	$O(M)$	$B(S) + B(R) \leq M^2$	
two-pass, hash-based		$3B(R)$	$O(M)$	$\min(B(S), B(R)) \leq M^2$	
multipass, sort-based		$(2k-1)B(R)$	$O(M)$	$B(R) + B(S) \leq M^k$	
multipass, hash-based		$(2k-1)B(R)$	$O(M)$	$\min(B(R), B(S)) \leq M^k$	

Product

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, full-relation	`OnePassBagProduct`	$B(R) + B(S)$	$O(1)$	$B(S) \leq M$	
Nested loop		$\frac{B(R) \cdot B(S)}{M}$	M	**none**	

Join

Category	Algorithm	I/O cost	Memory	Constraint	Note
One-pass, full-relation	`OnePassJoin` `NaiveOnePassJoin`	$B(R) + B(S)$	$O(M)$	$B(S) \leq M$	
Nested loop	`TupleJoin` `BlockNestedLoopsJoin`	$T(R) \cdot T(S)$ $O\left(\frac{B(R) \cdot B(S)}{M}\right)$	$O(1)$ M	**none** **none**	
two-pass, sort-based		$3B(R)$	$O(M)$	$B(S) + B(R) \leq M^2$	
two-pass, hash-based		$3B(R)$	$O(M)$	$\min(B(S), B(R)) \leq M^2$	
multipass, sort-based		$(2k-1)B(R)$	$O(M)$	$B(R) + B(S) \leq M^k$	
multipass, hash-based		$(2k-1)B(R)$	$O(M)$	$\min(B(R), B(S)) \leq M^k$	
index-based	`zigzgagJoin`	$O\left(\frac{T(R) \cdot T(S)}{V(S,Y)}\right)$ $O\left(\frac{B(R) \cdot B(S)}{V(S,Y)}\right)$	M M	index index	**unclustered** **clustered**