## Chapter 4.2 One Pass Algorithm

***Question:***

How could we execute each single operation of Logical Query Plan? (i.e. Join or Selection) Selecting an operator algorithm is a necessary part to convert Logical Query Plan to Physical Query Plan.

***Classification based on Algorithm:***

1. *Method based on Sort.*
2. *Method based on Hash.*
3. *Method based on Index.*

***Classification based on Difficulty and Cost:***

1. *One – Pass Algorithm:*

Reading data from disk for one time. This method requires that at least one Operation Object of the Operation can be stored into the main memory. (Especially for Projection and Selection)

1. *Two – Pass Algorithm:*

Data Size is too big to be put into the available main memory, but it is not the biggest data set. The feature of two – pass algorithm is to read data from disk and deal with them in one method and re-write them back to the disk, then in next pass, read the data again.

1. *Three – Pass (Multi – Pass) Algorithm:*

The Three – Pass (Multi - Pass) Algorithm has no limitation for the data size. They can finish work in three or multi - pass algorithm. They are also extension of Recursive Algorithm.

***Classification for Operator:***

1. *One Single Tuple, Unitary Operation:*

The Operation *(Selection and Projection Operator)* does not need to store the whole Relation or even the most part of the Relation into the main memory. The Operation reads one Block for one time, and generate the output by using the main memory.

1. *Whole Relation, Unitary Operation:*

The Operation *(Grouping and Deduplication Operator)* does need the whole or main part of tuples, then One – Pass Algorithm is limited to M size *(The buffer number that can be used in main memory.)* or smaller Relation.

1. *Whole Relation, Dual Operation:*

*(Union[并集], Intersection[交集], Difference[差集], Joint[连接], Product[积集] and Packet[包集])* If we want to use One – Pass Algorithm, then the size of Operator is limited to M size.

### Chapter 4.2.1 One Pass Algorithm for Single Tuple Operation

***Principle:***

No matter whether Relation R can be stored into main memory, Operations *(Selection and Projection)* have their obvious algorithms.

***Process:***

* Read one block of Relation R into the input buffer, operates on each tuple, and move the selected/projected tuple into the output buffer.

*(The output buffer maybe another operation input buffer, so the output buffer doesn’t count to the whole buffer.)*

* No matter how big the block B is, it requires that it must satisfies M >= 1.



***Cost:***

* The cost of operations *(Selection and Projection)* is decided by how Operation Objects in Relation R are provided. If Relation R is on the disk, then the cost is once execution of Table – Scan or Index – Scan.
* Normally, if Relation R is cluster, then the cost is *B(R)*; Otherwise, if Relation R is non – cluster, then cost is *T(R)*.

### Chapter 4.2.2 One Pass Algorithm for Single Tuple Operation in Whole Relation

Let’s consider One – Pass Algorithm for Single Tuple in whole Relation: *Deduplication and Grouping*.

***Deduplication:***

1. *Prerequisite:*
2. If the first time visiting the tuple, just duplicate it to the output buffer and output it.
3. If we had visited the tuple before, then no need to output it.

*(In order to support two prerequisites, then keep one copy for each tuple in main memory. Using one memory buffer to keep one block of one tuple of Relation while other M - 1 buffer area can be used to save every tuple that we saw before so far.)*

1. *Process:*

* Take the used data structure into consideration when we store the visited tuple. When considering the new tuple in the Relation R, we compare it with tuples that we have visited before.
* If the tuple doesn’t equal to any one in tuple collection, then copy it to output and add it into tuple list that we have visited before.

1. *Cost:*

***Grouping:***

1. Process:
2. Cost:

### Chapter 4.2.3 One Pass Algorithm for Two Tuples Operation