**Parallel and Distributed Computing Assignment**

*Answer 1:*

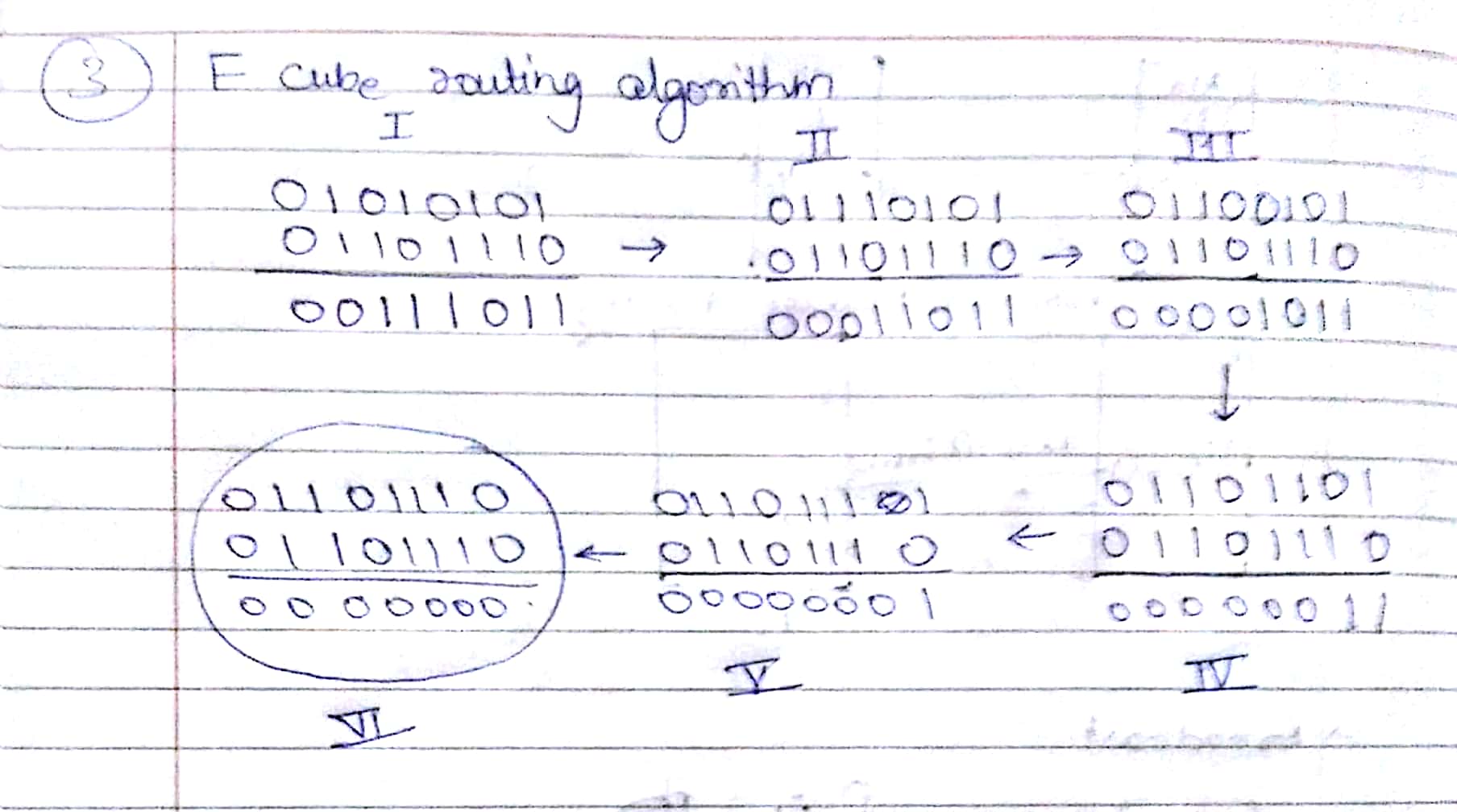
Since, the file splitting operation is sequential, and consumes 10% for every iteration to split the files, which affects the performance of the search algorithm.

Instead, the map and reduce phases of MapReduce split up the computing job into chunks that standard machines can process in a short time. First, the input data, in our case a collection of data is split into n size where the size of the split is chosen to ensure that the work can be distributed evenly (chunks should not be too large) and efficiently (the total number of chunks we need to manage should not be too large); 16 or 64 MB are good sizes in distributed indexing also, hadoop v2.0 supports up till 128 MB and hence it is more efficient . Splits are not preassigned to machines, but are instead assigned by the master node on an ongoing basis: As a machine finishes processing one split, it is assigned the next one. If a machine dies or becomes a laggard due to hardware problems, the split it is working on is simply reassigned to another machine.

*Answer 2:*

It is possible to integrate message passing and shared memory multicomputers by creating a communication channel between the shared memory block and message passing block.

But, generally, weather datasets are large-scale based datasets, where shared memory alone is efficient to process the rightful insights.

*Answer 3: *