**ASSIGNMENT – 21.1**

Q1 .Explain the different methods of

● mapper class and

● reducer class

**1. Map method in mapper:**

map(inKey, inValue) -> list(intermediateKey, intermediateValue)

Maps are the individual tasks which transform input records into a intermediate records. The transformed intermediate records need not be of the same type as the input records. A given input pair may map to zero or many output pairs.

The purpose of the map phase is to organize the data in preparation for the processing done in the reduce phase. The input to the map function is in the form of key-value pairs, even though the input to a MapReduce program is a file or file(s). By default, the value is a data record and the key is generally the offset of the data record from the beginning of the data file.

The output consists of a collection of key-value pairs which are input for the reduce function. The content of the key-value pairs depends on the specific implementation.

**2. Reduce method in reducer:**

In this phase the reduce(WritableComparable, Iterator, OutputCollector, Reporter) method is called for each <key, (list of values)> pair in the grouped inputs.

The output of the reduce task is typically written to the FileSystem via OutputCollector.collect(WritableComparable, Writable).

Applications can use the Reporter to report progress, set application-level status messages and update Counters, or just indicate that they are alive.

The output of the Reducer is not sorted.

**3. Setup and cleanup methods in mapper and reducer :**

In hadoop Mapreduce program basically two methods are there Hadoop setup Method and hadoop cleanup method.We can write Hadoop setup method and cleanup method in both Mapper and reducer where we required in our mapreduce code.

**Hadoop setup method:**

During setup() we may read parameters from the configuration object to customize our processing logic.The lifecycle of a map/reduce task is (from a programmer’s point of view):

setup -> map -> cleanup

setup -> reduce -> cleanup

we can write hadoop setup method before Map task or before reducer task.

setup: Called once at the beginning of the task We can use hadoop setup method if you want to define user defined parameters to your map task or reducer task and you want to compare the words with mapreduce input file and if you want to use hadoop distributed cache in your mapreduce program at that time we will hadoop setup method.

**Hadoop Cleanup Method:**

cleanup: Called once at the end of the task.

During cleanup() we clean up any resources that we have allocated. There are other uses too, which is to flush out any accumulation of aggregate results.

As already mentioned, setup() and cleanup() are methods we can override, they are there for to initialize and clean up your map/reduce tasks. We actually don’t have access to any data from the input split directly during these phases. The lifecycle of a map/reduce task is (from a programmer’s point of view):

setup -> map -> cleanup

setup -> reduce -> cleanup

So, for each mapreduce first setup() method is called then map()/reduce() method is called and later cleanup() method is called before exiting the task.