

- Explain the functionality of `TextOutputFormat` with an example.
the `TextOutputFormat` writes records as lines of text. the keys and values may be any type, can suppress the key or the value from the output using a `NullWritable` type. for example, in task 1, it use the `TextOutputFormat` for output, and the key is the path in `Text` and the output is the count of the words in `IntWritable`.
- Explain the functionality of `MultipleOutputFormat` with an example
`MultipleOutputFormat` helps to produce multiple files per reducer. such as the problem of partitioning the weather dataset by weather station. it outputs one file per station, which contain all records for this station. it allows you to write data to files whose names are derived form the output keys and values. such as the map outputs name will be like name-m-nnnn, and the reducer outputs name will be like name-r-nnnn, the nnnn is the part number.
- Explain the idea behind distributed cache with an example.
The distributed cache is aim to reduce bandwidth for each cluster when doing MapReduce. The cluster load the file into cache to save time and bandwidth.

Applications specify the files, via urls (`hdfs://` or `http://`) to be cached via the `JobConf`. The `DistributedCache` assumes that the files specified via urls are already present on the `FileSystem` at the path specified by the url and are accessible by every machine in the cluster. The framework will copy the necessary files on to the slave node before any tasks for the job are executed on that node. Its efficiency stems from the fact that the files are only copied once per job and the ability to cache archives which are un-archived on the slaves.