map reduce:

input: a set of key/value pairs

output: a set of key/value pairs

Map: get input, produce the intermediate input pairs

Reduce: The intermediate values are supplied by an iterator.

The library is responsible for all the paralyzing and distribution

input(distributed system)->intermediate results(local disk)->output files(distributed systems)

no need for complete accuracy in large data system

-> tolerate some faults.