Week9 Paper Summary——MapReduce
By Yi Zhou

This week's paper introduces a programming model in order to process and generate large data sets. User specify a map function to process key/value pairs and then generate a set of intermediate key/value pairs. A reduce function is used to merge all intermediate values associated with the same key. This paper uses word frequencies counter example as we do in our homework to explain how map-reduce model works. After that, it shows some practical application of such model. It mainly focuses on the implementation of such model and tackle some essential problems in the real-world application such as fault tolerance, how to design task granularity and so on.

As far as I am concerned, MapReduce model is quite strong to tackle with the big data, based on which many convenient frame work have been invented such as Hadoop. Besides , MapReduce is useful in a wide range of applications, including distributed pattern-based searching, distributed sorting, web link-graph reversal, Singular Value Decomposition, web access log stats, inverted index construction, document clustering, machine learning and statistical machine translation. However, there are still some flaws in such a powerful model, for instance, MapReduce tasks must be written as acyclic dataflow programs, i.e. a stateless mapper followed by a stateless reducer, that are executed by a batch job scheduler. This paradigm makes repeated querying of datasets difficult and imposes limitations that are felt in fields such as machine learning, where iterative algorithms that revisit a single working set multiple times are the norm. Maybe, more refinement should be done according to the nee of practical task.