

## Week 2 Paper 2 Summary by Wenhan Zhu

### MapReduce as a General Framework to Support Research in Mining Software Repositories (MSR)

This paper demonstrates using Hadoop, an open source implementation of MapReduce, to support large scale analysis of software systems. The authors migrated J-REX a in house tool for analyzing java code to Hadoop. They used distributed methods on different part of the migrated J-REX and concluded that when using distributed solutions on all phases of J-REX would provide the best result. Using Hadoop introduces flexibility and scalability to the process. Hadoop runs on all major platforms and allow machines to enter and leave without hassle.

3 Things I want to discuss:

- Performance Evaluation  
The performance evaluation part of the paper contains many inconsistencies, I would like to discuss more about the result of the evaluation.
- Running time of different flavors of migrated J-REX  
The reason of the result is not quite clear to me, more discussion would help understanding the differences.
- How distributed system in general helps  
The paper mentions how distributed make things run faster but never went into much detail.