Summary of Academic Paper Number 2[1]

Andrew K Boles

University of Texas at San Antonio

Email: ckj771@my.utsa.edu

I. SUMMARY

In this paper, the authors propose an adaptation to the MapReduce framework to process real-time medical data. Perhaps the most popular example of MapReduce is the Hadoop[2] MapReduce framework. The addition proposed is a task-level adaptive MapReduce framework which will enable the MapReduce framework to assess the incoming flow of data and be able to allocate Map and Reduce tasks appropriately. For the Map task assignment, the authors suggest a process which would take advantage of the nodes present in the cluster: sending more tasks to computationally stronger nodes and less to those that are not as computationally strong. Additional Map tasks can be launched if needed as well as taken away if needed. This enables the program running to be able to be more efficient while handling the Map tasks. As for Reduce, the authors propose being able to add or subtract Reduce tasks as needed in order to more efficiently handle the data being processed and then output to the file system. Adding more Reduce tasks lowers the amount of data each task needs to process and also effectively makes the Reduce task as a whole more parallel. All of this is done by essentially implementing the MapReduce framework as a running daemon.

After suggesting these changes, the authors proposed some scheduling devices that would be able to implement the changes, in other words be able to allocate the number of MapReduce tasks. The two proposals would work together to form a new addition to the existing MapReduce framework.

REFERENCES

[1] F. Zhang, J. Cao, S. U. Khan, K. Li, K. Hwang. "A task-level adaptive MapReduce framework for real-time streaming data in healthcare applications", Future Generation Computer Systems, pp. 149-160, 2014.

[2] Apache, "Apache Hadoop". Available: *http://hadoop.apache.org/*