**Title:** Distributed Hadoop Mapreduce Cluster on Android Devices

**Authors:** Dustin McAfee and Alan Grant

**Thesis:** New Android mobile devices are becoming more and more powerful with larger and larger amounts of memory and quad and even octa core processors. Hadoop Mapreduce is installed and optimized on top of a distributed cluster of Android devices. Utilizing multiple Android mobile devices for processing of big data with Hadoop Mapreduce is shown as feasible. Example algorithms such as WordCount is used on a three node distributed cluster of two Android devices as slaves and a linux laptop device as the master node.

**Approach:** Big data is termed here as a large number of data to be processed in a short period of time. Hadoop Mapreduce is one of the most popular algorithms for big data processing. Hadoop is a scalable, open source, fault-tolerant framework that allows for distributed processing of large data sets across clusters of computers using simple programming models. Mapreduce is the programming model used to process large data sets stored in Hadoop. Porting Hadoop Mapreduce to the mobile Android platform could have hardware advantages, since the cost of Android hardware is generally low. This would also enable a mobile workforce instead of a desktop workforce.

In order to port Hadoop to Android, a linux image is installed and run inside of the Android operating system; this requires root privileges. An open source package called Busybox is installed with root privileges, which provide replacements for many GNU utilities, including chroot [2]. Chroot is used to change root environment; it mounts a specific directory as root and makes a subprocess tree for specific system. In this case chroot is used with an Ubuntu Xenial image to boot Ubuntu on top of the Android operating system.

Open JDK 8, and Hadoop 2.7.5 is installed on a rooted Samsung Galaxy S7 running Android Nougat, a Hikey960 development board running AOSP 4.9, and a laptop running Ubuntu Xenial. Sample Mapreduce jobs are ran on slaves separately and the runtimes compared to the same jobs ran on the entire cluster. Different memory allocation configurations are benchmarked, and an optimal configuration is determined.

**Tentative List of References (bibtex citations included):**

[1] Jeffrey Dean and Sanjay Ghemawat. 2008. MapReduce: simplified data processing on large clusters. Commun. ACM 51, 1 (January 2008), 107-113. DOI: https://doi.org/10.1145/1327452.1327492

@article{Dean:2008:MSD:1327452.1327492,

author = {Dean, Jeffrey and Ghemawat, Sanjay},

title = {MapReduce: Simplified Data Processing on Large Clusters},

journal = {Commun. ACM},

issue\_date = {January 2008},

volume = {51},

number = {1},

month = jan,

year = {2008},

issn = {0001-0782},

pages = {107--113},

numpages = {7},

url = {http://doi.acm.org/10.1145/1327452.1327492},

doi = {10.1145/1327452.1327492},

acmid = {1327492},

publisher = {ACM},

address = {New York, NY, USA},

}

[2] Landley, Rob. Vlasenko, Denys. Reutner-Fisher, Bernhard (2018). BusyBox (1.28.2) [open source program]. [https://busybox.net/](https://busybox.net/about.html)

[3] Prachil S. Tambe, P. S. Y. R., B. B. R., “Implementation of Hadoop Pseudo-distributed Cluster on Android using ‘chroot’”, *International Journal Of Engineering And Computer Science*, vol. 4, no. 04, Apr. 2015.

@article{Prachil S. Tambe,

title={Implementation of Hadoop Pseudo-distributed Cluster on Android using ‘chroot’},

volume={4},

url={<http://www.ijecs.in/index.php/ijecs/article/view/1498>},

number={04},

journal={International Journal Of Engineering And Computer Science},

author={Prachil S. Tambe, Prof. S. Y. Raut, Balwant B. Raut, Mangesh S. Jondhale, Nilesh R. Jaware},

year={2015},

month={Apr.}

}

[4] Marinelli, E. Hyrax: Cloud Computing on Mobile Devices using MapReduce. Masters Thesis, Carnegie Mellon University, Sep. 2009.

@article{Marinelli,

title={Hyrax: Cloud Computing on Mobile Devices using MapReduce},

url={http://www.dtic.mil/dtic/tr/fulltext/u2/a512601.pdf},

author={Eugene Marinelli},

year={2009},

month={Sep.}

}

[5] Bothe B. Namrata, N. M. A., K. S. S. "*Migration of Hadoop To Android Platform Using ‘Chroot’*", International Journal of Innovative Research and Creative Technology, ISSN:2454-5988, Volume-1 Issue-5 page no.486-488, Apr. 2016. http://www.ijirct.org/papers/IJIRCT1201105.pdf

@article{Bothe,

title={Migration of Hadoop To Android Platform Using ‘Chroot’},

url={http://www.ijirct.org/papers/IJIRCT1201105.pdf},

author={Bothe B. Namrata, Mate N. Anagha, Karale S. Snehal, Kumbhar D. Nayan},

year={2016},

month={Apr.},

}