Abstract

MapReduce is an implementation for processing large scale data parallelly. Actual benefits of MapReduce occur when this framework is implemented in large scale, shared nothing cluster. MapReduce framework abstracts the complexity of running distributed data processing across multiple nodes in cluster. Hadoop is open source implementation of MapReduce framework, which processes the vast amount of data in parallel on large clusters. In Hadoop pluggable scheduler was implemented, because of this several algorithms have been developed till now. This paper presents the different schedulers used for Hadoop.

References

- Hadoop http://hadoop.apache.org/
- B. Thirumala Rao and Dr. L. S. S. Reddy, "Survey on Improved Scheduling in
Survey on Task Assignment Techniques in Hadoop

Hadoop MapReduce in Cloud Environments &quot;
net/blogs/hadoop/FairSharePres. ppt.
org/core/docs/current/capacity_scheduling.html.
- Matei Zaharia, Dhruba Borthakur, Joydeep Sen Sarma, Khaled Elmeleely, Scott
Shenker and Ion Stoica, &quot;Delay Scheduling: A Simple Technique For Achieving Locality
and Fairness in Cluster Scheduling&quot;.
- Thomas Sandholm and Kevin Lai. Dynamic proportional share scheduling in hadoop. In
JSSPP &apos;10: 15th Workshop on Job Scheduling Strategies for Parallel Processing,
April,2010
- K. Kc and K. Anyanwu, &quot;Scheduling Hadoop Jobs to Meet Deadlines&quot;, in
- Mark Yong, Nitin Garegrat, Shiwali Mohan: &quot;Towards a Resource Aware Scheduler
in Hadoop&quot; in Proc. ICWS, 2009, pp:102-109
- Jaideep Dhok and Vasudeva varma,&quot; Using Pattern Classification for Task
Assignment in MapReduce&quot;.
- Jaideep Dhok, Nitesh Maheshwari and Vasudeva Varma,&quot;Learning Based
Opportunistic Admission Control Algorithm for MapReduce as a Service&quot;

Index Terms
Computer Science Information Systems

Keywords
MapReduce Task Assignment Resource Management Scheduling