Abstract

In view of the fact that clusters used in large scale computing are on the rise, ensuring the wellbeing of these clusters is of paramount significance. This highlights the importance of supervising and monitoring the cluster. In this regard, many tools have been contributed that can efficiently monitor the Hadoop cluster. The majority of these tools congregates necessary information from each of the node in the cluster and takes it for processing. These diagnosis tools are mostly post execution analysis tools. This paper presents an exploratory assessment of the different log analyzers used for failure detection and monitoring in Hadoop.

References

- W. Tom, Hadoop:the definitive guide (O'reilly media, May 2009)
An Exploratory Survey of Hadoop Log Analysis Tools

- D. Borthakur, HDFS Architecture, http://hadoop.apache.org/common/docs/r0.20.0/hdfs_design.html, April 2009
- Scribe logfile aggregation system described by Facebook's Jeff Hammerbacher https://issues.apache.org/jira/browse/HADOOP-2206?focusedCommentId=12542775#action
12542775
- J. Boulon, A. Konwinski, R. Qi, A. Rabkin, E. Yang, and M. Yang, Chukwa, a large-scale monitoring system, In First Workshop on Cloud Computing and its Applications (CCA '08), Chicago, IL, 2008
- A. Rabkin, R Katz, Chukwa: a system for reliable large-scale log collection, In Proceedings of the 24th International Conference on Large Installation System Administration LISA'08, 10, USENIX Association Berkeley, CA, USA.

Index Terms

Computer Science

Artificial Intelligence
Keywords

Cloud computing  HDFS  Failure monitoring  Hadoop  Log analyzer