Mining Logs to Predict Failures in Clusters and Grids

Aditya Pakki, Harshitha Venkata Parnandi
School of Computing
University of Utah, Salt Lake City, UT, 84112, USA
{pakkiadi,pvharsh}@cs.utah.edu

1. TEAM MEMBERS

- Aditya Pakki - $u0922291$
- Harshitha Venkata Parnandi - $u0927529$

2. PROBLEM STATEMENT

There is a rapid growth in the number of compute devices per cluster or grid to keep up with the Moore's law. However, the hardware vendors have not been able to provide a corresponding growth in the reliability characteristics of these newer devices, resulting in higher incidence of system failures. We intend to use an analytical approach to predict the incidence of failures, using event co-relation from the past system logs generated by similar clusters.

3. METHODOLOLGY

To the best of our knowledge, [1] is the first publicly available repository providing logs of failures in clusters. Most of the previous papers used [1] or private data as their basis for understanding resiliency. These logs are available only upto 2007. [4], [5] and [3] are few relevant papers, which tried to address the problem using data mining approaches. We have identified [2] as a source for cleaned as well as raw supercomputing logs upto 2015. We intend to use the concepts learned in class, the latest available datasets, and our own mining rules to perform event co-relation and identify failures.

Researchers in the area of High Performance Computing(HPC) consider system resiliency to be one of the important challenges in reaching Exascale computing and the field relies solely on checkpointing and restart as its fault tolerance mechanism. We believe, an analytical approach might provide an alternative viewpoint to the ongoing research in solving this challenge.

4. REFERENCES

[1] The computer failure data repository (cfdr). https://www.usenix.org/cfdr.

[2] Logs of real parallel workloads from production systems. http://www.cs.huji.ac.il/labs/parallel/workload/logs.html.

[3] A. Gainaru, F. Cappello, S. Trausan-Matu, and B. Kramer. Event log mining tool for large scale hpc systems. In *Euro-Par 2011 Parallel Processing*, pages 52–64. Springer, 2011.

[4] R. Ren, X. Fu, J. Zhan, and W. Zhou. Logmaster: Mining event correlations in logs of large scale cluster systems. *arXiv preprint arXiv:1003.0951*, 2010.

[5] B. Schroeder and G. A. Gibson. Understanding failures in petascale computers. In *Journal of Physics: Conference Series*, volume 78, page 012022. IOP Publishing, 2007.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235