Abstract

MapReduce paradigm is highly suitable for large scale data intensive applications in the cloud environment. The scale of these applications necessitates minimization of cluster power consumption to reduce operational costs and carbon footprint. Energy consumption can be reduced by selective power down of nodes during periods of low utilization. Hadoop is basically used for batch processing of huge jobs. Before jobs are submitted, the files used them are uploaded into the cluster. A file is split up into a number of chunks and distributed across the
Hadoop cluster. This paper addresses the problem of block allocation in distributed file system to improve reliability and energy efficiency. A framework to reduce power requirements of a cluster by identifying the number of replicas and their placement for reliable completion of the job has been designed. This will address the issues like block allocation, reliable job submission and minimization of cluster nodes to reduce power consumption. This framework is integrated with hadoop’s namenode. The scheduler component in Hadoop has also been modified to enable submission of jobs to active data node containing data to be operated on. A greedy approach and an evolutionary approach using Particle Swarm Optimization (PSO) has been designed to identify suitable nodes to be activated in a cluster. Experimental results demonstrate the performance of these approaches.

References

- Willis Lang and Jignesh M. Patel, Energy Management for MapReduce Clusters, Computer Sciences Department, University of Wisconsin Madison, USA.
- Jacob Leverich, Christos Kozyrakis, On the Energy (In)efficiency of Hadoop Clusters, Computer Systems Laboratory, Stanford University.
Energy Efficient and Reliable Job Submission in Hadoop Clusters

- Jan Stoess, Christoph Klee, Stefan Domthera, Frank Bellosa, Transparent, Power-Aware Migration in Virtualized Systems.
- Live Data Center Migration across WANs: A Robust Cooperative Context Aware Approach K. K. Ramakrishnan, Prashant Shenoy, Jacobus Van der Merwe AT&T Labs-Research / ?? University of Massachusetts
- R. Jeyarani, R. Vasanth Ram, N. Nagaveni, Design and implementation of adaptive power-aware virtual machine provisioner (APA-VMP) using swarm
- Power-aware linear programming based scheduling for heterogeneous computer clusters. Rini T Kaushik, Milind Bhandarkar, GreenHDFS: Towards an energy-conserving, storage-efficient, hybrid Hadoop compute cluster.
- PSO reference - http://www.swarmintelligence.org – Study about PSO.

Index Terms

Computer Science  Information Science

Keywords

Energy Efficiency  Hadoop  Reliability  PSO