Abstract

Cloud Computing is an emerging computing paradigm. It aims to share data, calculations, and service transparently over a scalable network of nodes. Cloud Computing is nothing but a collection of computing resources and services pooled together over internet and is provided to the users on pay-as-needed basis. In cloud computing, load balancing is required to distribute the dynamic local workload evenly across all the nodes. It helps to achieve a high user
satisfaction and resource utilization ratio by ensuring an efficient and fair allocation of every computing resource. Proper load balancing aids in minimizing resource consumption, implementing fail-over, enabling scalability, avoiding bottlenecks and over-provisioning etc. It helps in optimal utilization of resources and hence in enhancing the performance of the system. A few existing scheduling algorithms can maintain load balancing and provide better strategies through efficient job scheduling and resource allocation techniques as well. In order to gain maximum profits with optimized load balancing algorithms, it is necessary to utilize resources efficiently. This paper discusses some of the existing load balancing algorithms in cloud computing and also their challenges.

References

- Tanveer Ahmed, Yogendra Singh, Analytic study of load balancing techniques using tool cloud analyst
- Peter Mell, Timothy Grance, &quot;The NIST Definition of Cloud Computing&quot;, NIST Special Publication 800-145, September 2011.
- Meenakshi Sharma, Pankaj Shrama, Dr. Sandeep Sharma, Efficient Load
- Hardeep Uppal, dane Brandon, OpenFlow based load balancing.
- Hardeep Uppal, dane Brandon, OpenFlow based load balancing.
Load Balancing Techniques in Cloud Computing: A Study

Index Terms

Computer Science

Cloud Computing

Keywords

Load Balancing  Cloud Computing