Paper Title: Improving HPC Application Performance in Cloud through Dynamic Load Balancing

Paper link:https://dl.acm.org/doi/10.1109/CCGrid.2013.65

1 Summary:
1.1 Motivation: The paper aims to improve load balancing in heterogeneous settings by introducing adaptability for resource optimisation, scalability improvement, and enhanced performance in dynamic environments in response to cloud-based HPC challenges.

1.2 Contribution: The paper addresses heterogeneity and interference in cloud-based HPC by offering an improved load balancing technique. Findings show a 45% reduction in execution time, which is essential for performance optimisation.

1.3 Methodology: The paper uses KVM for virtualization in an OpenStack cloud environment. To combat heterogeneity and interference, it uses an advanced load balancing algorithm, which improves performance in HPC applications.

1.4 Conclusion: The paper presents a customised load balancing method that tackles cloud environment heterogeneity and interference. The method's effectiveness in improving the execution of HPC applications in the cloud is highlighted by the significant performance gains observed in the results.

2 Limitations
2.1 First limitation:
Scalability Constraints:Restricted experiment scale as a result of limitations in cloud node availability, which prevents thorough assessment and extrapolation to larger cloud infrastructures.

2.2 Second limitation:
Sensitivity to Application Characteristics:Depending on the features and nature of particular HPC applications, the suggested load balancing method's efficacy may change.

3 Synthesis:
Adaptive load balancing for cloud-based HPC is synthesised in this paper, addressing interference and heterogeneity issues. The outcomes show that better load distribution and migration techniques lead to increased performance.