

Abstract

Cloud computing has been used for business applications for quite some time with considerable success. The rapid rise of cloud computing can be attributed to the conveniences provided by it, such as infinite and dynamic resource availability, managing IT resources, limiting the cost to only the services used, etc. Cloud computing has been utilized by non-IT businesses which use computation to enhance their business, while the core business may be in a different domain. Traditional use of IT services would require over or under provisioning of computational and associated resources by these businesses; thereby increasing cost. In cloud computing, these costs are significantly reduced and the problems get eliminated. Traditionally, the scientific community has utilized the services of high performance computing (HPC) or Grid computing (HPC built by aggregating power of a distributed system). However, a similar problem that businesses face is also equally applicable to the scientific computing community. Generally, scientists are experts in their own domain of science and are expected not to be proficient in maintaining IT resources. Also, resource requirements are dynamic in nature and varies between each experiment or project. Considering these factors, a use case for cloud computing in scientific computing can be made. Traditional data center based cloud computing had distributed computing embedded in it in terms of aggregating computation powers of multiple devices and creating each virtual device. Recently, this decentralization has been expedited with the use of concepts such as Volunteer cloud, Edge computing, etc. Hence, apart from the typical issues associated with cloud computing like Virtual Machine management, issues arising out of distributed systems also need attention. In this research work, issues arising out of data placement in such distributed systems and associated reliability factors are studied and some novel solutions to existing problems are proposed.