Abstract

Cloud computing is a large-scale parallel and distributed computing paradigm. It is a collection of virtualized and interconnected computing resources which are managed to be one unified high performance computing power. Cloud environment being highly dynamic and heterogeneous, failures on data nodes are common. Therefore, to improve the availability and reliability of the system, the data is to be replicated to numerous suitable locations. By analyzing the content popularity, the popular data is replicated and the replicas are allocated to preferable data nodes. This phenomenon increases the data availability, speeding up of data access and minimizing cloud system bandwidth consumption. In this paper, various strategies
of data replication based on content popularity are studied and how the replication can be carried out effectively by considering a few parameters like number of replicas, replica placement and replica management is analyzed.

References


Index Terms

Computer Science Distributed Systems
Keywords
Cloud Computing  Content Popularity  Data Availability  Replication  Replica Allocation