Workload Aware Replicated Datapartitioning for Twitter

Abstract

Most of the queries in twitter include multiuser operations. When a user login to twitter it requests the most recent tweets of whom he follows. These data may be present in different servers. The expense of these queries depends on how the data is partitioned. Existing solution for data partitioning involve hash or graph based partition. In this paper a new method for reducing the interaction between the servers are proposed. For this the data is partitioned such that most of the users that a user interacts are placed on the same partition. In addition to data partition selective replication is also implemented in the proposed approach. The data about the users that are requested most are replicated more than the other users. Experimental analysis indicates that the proposed technique provides significant improvements in the quality of the partitions, especially under low replication ratios.

References


Index Terms

Computer Science Information Sciences

Keywords

Data partitioning; Selective replication; Social network; Twitter; Cassandra.