Overview of Replication Techniques on Distributed Database in Cloud Environment

Sharifah Hafizah Sy Ahmad Ubaidillah 1, A. Noraziah 1, 2
1Faculty of Computer Systems & Software Engineering, University Malaysia Pahang, 26300 Kuantan, Pahang, Malaysia
2IBM Centre of Excellence, University Malaysia Pahang 26300 Kuantan, Pahang, Malaysia

Replication is a topic of interest in the cloud computing, distributed systems, and database communities. It is also one of the practical methods in distributed systems for developing better reliability and availability. Replication is employed to reduce user waiting time, increasing data availability and decreasing cloud system bandwidth used by offering the user numerous replicas of a detailed service on multiple nodes in cloud computing. This paper presents an overview of replication techniques on distributed database in the cloud environment.

Keywords: Replication, Cloud Computing, Distributed System, Database, Availability

1. INTRODUCTION

Replication is one of the most extensively studied phenomena in a distributed environment 1. Replication is a method where the multiple copies of some data are stored at the multiple sites. Data will be retrieved from the nearest accessible replica when necessary needed, thus can avoid delays and increased the performance of the system. Availability needs to be immense in the cloud computing paradigm which makes replication of data in the cloud environment, a challenging task. Crisis in providing dynamic and precise wide area database replication is that it needs, incorporating numerous techniques from various fields, including distributed system, databases, network protocols and operating system 2. In the cloud, data replication is attained through data resource pool and the number of data replicas is statically set based on history and experience 3. Further, it is not needed to construct a replica for all data files, specifically for those seldom or recently used data files. Thus, it is required to adaptively replicate the regularly used data files, regulate the number of data replicas and the data node area where to place the new replicas according to the present cloud environments conditions.4

2. DISTRIBUTED DATABASE SYSTEM

A distributed database (DDB) is a set of numerous, logically interrelated databases distributed over a computer network. The software that regulates the DDB is called as the distributed database management system (DBMS). This software produces an access mechanism that generates this distribution transparent to the user 4. Distributed database system (DDBS) is the integration of DDB and DBMS. This integration is obtained through the combination of the database and networking technologies together. Or it can be expressed as a system that runs on a set of machines that do not have shared memory, but still looks to the user like a single machine. Figure 1 shows the example of DDBS.