A Hybrid Data Deduplication Technique in Cloud Computing for the Data Storage

By
Omar Adnan Isaid
Supervisor
Dr. Hesham Abusaimeh

Abstract

The data deduplication is a recent compression method that work on removing the redundant data from data stream to meet the challenges that raised by exponential growth of data. On the other hand, the services of cloud computing are commonly used nowadays. Since, it provides a lot of services that play a role in reducing the expenses of storage and processing of data on the enterprises. Consequently, the data deduplication in cloud computing is a hot topic, and it receives attention by researchers and enterprises. This thesis proposes hybrid data deduplication technique in cloud computing for the data storage, that it combines two different hybrid subsystems at the same method. The proposed hybrid data deduplication method contains two different types of chunk level deduplication, that each type is combined with file level deduplication to increase its performance. Indeed, the file level deduplication is important. When, there are a lot of redundant copies of same file.

x

The proposed system is tested using a simulator that is developed using Java, and the results of hybrid subsystems at the hybrid system are compared using curve charts. The proposed system consists of two subsystems; the hybrid file variable size chunk level deduplication (FVCD) and hybrid file fix size chunk level deduplication (FFCD). And, the testing is conducted using two test cases, that each test case has three scenarios. The first scenario is comparison between FVCD and FFCD with chunk's size 500 bytes, at second scenario the first scenario is repeated and chunk' size increased of FFCD to became 1 KB. And, and at third scenario the chunk's size increased to 1 MB. After testing the simulator of proposed system, the results show that FVCD can provide better effectivity in reducing the data size, and FFCD can provide better execution time. Finally, the proposed system is a step toward working on improving the data deduplication techniques, depending on combining different types of data deduplication. And, let these types work together for covering wider range of user requirements, and for better employing the of data deduplication in the cloud for reducing the expenses and providing better cloud services for the users.