Implementation of deduplication process in cloud computing systems

fatemeh zargar*,hamidreza ahmadifar,

Abstract Implementation of deduplication process in cloud computing systems In recent years, most business processes have been digitalized, i.e., information such as communication data, accounts, contracts, advertising material, construction or business plans only exists in digital form. The data is often of great value and its irrecoverable loss or damage could be a total disaster for its owner. Additionally, companies are legally obliged to preserve tax records for a certain period (6 or 10 years), and to leave them available to the fiscal authorities. As a result, the world is moving towards using cloud storages. The volume of data storage in cloud computing has increased and storage providers need to consider reducing the amount of data storage in the cloud. To do this, you need to store only one copy of data in the cloud. Storage companies are looking for a way to recognize duplicate information in a variety of ways to reduce storage space in the clouds. This study attempts to explain the implement of 3 process of reducing redundancy. The first process divides the information into equal blocks and identifies duplicated data by Comparing that blocks which stored in the cloud. The second process comparing new and existing data in a cloud storage space and in case of similarity, replaces them with a reference to existing data. Therefor this process prevents duplicate data storing in cloud computing. The third method is a special type for storing scanned image data into the cloud, and in this process, only different points of the input images are stored in the cloud.