

**CHENNAI – PONDICHERRY**

**An Algorithm for Finding the Minimum Cost of Storing and Regenerating Datasets in Multiple Clouds**

**Abstract:**

The proliferation of cloud computing allows users to flexibly store, re-compute or transfer large generated datasets with multiple cloud service providers. However, due to the pay-as-you-go model, the total cost of using cloud services depends on the consumption of storage, computation and bandwidth resources which are three key factors for the cost of IaaS-based cloud resources. In order to reduce the total cost for data, given cloud service providers with different pricing models on their resources, users can flexibly choose a cloud service to store a generated dataset, or delete it and choose a cloud service to regenerate it whenever reused. However, finding the minimum cost is a complicated yet unsolved problem. In this paper, we propose a novel algorithm that can calculate the minimum cost for storing and regenerating datasets in clouds, i.e. whether datasets should be stored or deleted, and furthermore where to store or to regenerate whenever they are reused. This minimum cost also achieves the best trade-off among computation, storage and bandwidth costs in multiple clouds. Comprehensive analysis and rigid theorems guarantee the theoretical soundness of the paper, and general (random) simulations conducted with popular cloud service providers’ pricing models demonstrate the excellent performance of our approach.

**Existing System:**

Especially, nowadays applications are getting more and more data intensive [4], where the generated data are often gigabytes, terabytes, or even petabytes in size. These generated data contain important intermediate or final results of computation, which may need to be stored for reuse. Hence, cutting the cost of cloud-based data management in a pay-as-you-go fashion becomes a big concern for deploying applications in cloud computing environment.

Cloud computing has such a fast growing market, more and more cloud service providers appear with different prices of computation, storage and bandwidth resources. As unlimited storage and processing power can be easily obtained on-demand from different commercial service providers like utilities, users have multiple options to cope with the large generated application data, e.g., datasets d1, d2 … d8

Specifically, users can store all data in the cloud and simply pay for the storage cost, and alternatively, they can delete some data to save the storage cost and pay for the computation cost to regenerate them whenever they are reused, e.g,.

Further-more, users can also change to cheaper service providers to store or to regenerate data with paying for the bandwidth cost for data transfer1. Hence, there is a trade-off among computation, storage and bandwidth in clouds, where different storage and regeneration strategies lead to different total costs for storing the generated application data. In light of this, users need comprehensive understanding of cost in clouds in order to take advantage of the cost-effectiveness of cloud computing, especially for storing and regenerating data with multiple cloud service providers.

**Proposed System:**

We propose a novel GT-CSB algorithm that can find the best trade-off among computation, storage and bandwidth costs in clouds. This trade-off is represented by the theoretical minimum cost strategy for storing and regenerating application data among multiple cloud service providers. This minimum cost is a very important reference for cloud users in the following three aspects: 1) it can be used to design minimum cost benchmarking approaches for evaluating the cost effectiveness in clouds; 2) it can guide cloud users to develop cost effective storage strategies for their applications; and 3) it can demonstrate the constitution of different costs in clouds and help users to understand the impact of different workloads on the total cost.