

**CHENNAI – PONDICHERRY**

**COLLABORATION- AND FAIRNESS-AWARE BIG DATA MANAGEMENT IN DISTRIBUTED CLOUDS**

**ABSTRACT**

With the advancement of information and communication technology, data are being generated at an exponential rate via various instruments and collected at an unprecedented scale. Such large volume of data generated is referred to as big data, which now are revolutionizing all aspects of our life ranging from enterprises to individuals, from science communities to governments, as they exhibit great potentials to improve efficiency of enterprises and the quality of life. To obtain nontrivial patterns and derive valuable information from big data, a fundamental problem is how to properly place the collected data by different users to distributed clouds and to efficiently analyze the collected data to save user costs in data storage and processing, particularly the cost savings of users who share data. By doing so, it needs the close collaborations among the users, by sharing and utilizing the big data in distributed clouds due to the complexity and volume of big data. Since computing, storage and bandwidth resources in a distributed cloud usually are limited, and such resource provisioning typically is expensive, the collaborative users require to make use of the resources fairly. In this paper, we study a novel collaboration- and fairness-aware big data management problem in distributed cloud environments that aims to maximize the system throughout, while minimizing the operational cost of service providers to achieve the system throughput, subject to resource capacity and user fairness constraints. We first propose a novel optimization framework for the problem. We then devise a fast yet scalable approximation algorithm based on the built optimization framework. We also analyze the time complexity and approximation ratio of the proposed algorithm. We finally conduct experiments by simulations to evaluate the performance of the proposed algorithm. Experimental results demonstrate that the proposed algorithm is promising, and outperforms other heuristics.