

**CHENNAI – PONDICHERRY**

**COST-EFFECTIVE RESOURCE PROVISIONING FOR**

**MAPREDUCE IN A CLOUD**

**Abstract:**

This paper presents a new MapReduce cloud service model, Cura, for provisioning cost-effective MapReduce services in a cloud. In contrast to existing MapReduce cloud services such as a generic compute cloud or a dedicated MapReduce cloud, Cura has a number of unique benefits. First, Cura is designed to provide a cost-effective solution to efficiently handle MapReduce production workloads that have a significant amount of interactive jobs. Second, unlike existing services that require customers to decide the resources to be used for the jobs, Cura leverages MapReduce profiling to automatically create the best cluster configuration for the jobs. While the existing models allow only a per-job resource optimization for the jobs, Cura implements a globally efficient resource allocation scheme that significantly reduces the resource usage cost in the cloud. Third, Cura leverages unique optimization opportunities when dealing with workloads that can withstand some slack. By effectively multiplexing the available cloud resources among the jobs based on the job requirements, Cura achieves significantly lower resource usage costs for the jobs. Cura's core resource management schemes include cost-aware resource provisioning, VM-aware scheduling and online virtual machine reconfiguration. Our experimental results using Facebook-like workload traces show that our techniques lead to more than 80 percent reduction in the cloud compute infrastructure cost with upto 65 percent reduction in job response times.