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.

INTRODUCTION

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 is designed to provide a cost-effective solution to efficiently handle MapReduce production workloads that have a significant amount of interactive 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. 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 core resource management schemes include cost-aware resource provisioning, VM-aware scheduling and online virtual machine reconfiguration.

EXISTING SYSTEM

The existing cloud solutions for MapReduce work primarily based on a per-job or per-customer optimization approach where the optimization and resource sharing opportunities are restricted within a single job or a single customer.

DISADVANTAGE OF EXISTING SYSTEM

◆ Existing cloud solutions are largely optimized based on per-job and per-customer optimization which leads to poor resource utilization and higher cost.

◆ Existing solutions are not cost effective to deal with interactive MapReduce workloads that consist of a significant fraction of short running jobs with lower latency requirements.

◆ Existing models allow only a per-job resource optimization for the jobs.

PROPOSED SYSTEM

The growing trend in cloud computing is the combination of Big Data and Big Data analytics that has been driven with rapid evolution of data center technologies towards more cost-effective solutions. The most popular approach is using MapReduce for big data analytics. This project presents a cost-effective resource management framework. Cura usage model and techniques achieve higher service differentiation than existing models as Cura incorporates an intelligent multiplexing of the shared cloud resources among the jobs based on job requirements.
ADVANTAGE OF PROPOSED SYSTEM

- To minimize the infrastructure cost
- To minimize the number of servers required to handle the data center workload
- Depending upon deadlines for the jobs, VM aware scheduler needs to make future reservations.

ARCHITECTURE:
HARDWARE REQUIREMENTS:

- System: Pentium IV 2.4 GHz.
- Hard Disk: 40 GB.
- Floppy Drive: 44 Mb.
- Monitor: 15 VGA Colour.

SOFTWARE REQUIREMENTS:

- Coding Language: Java 1.7, Hadoop 0.8.1
- Database: MySql 5
- IDE: Eclipse