CloudKon: DTS

Distributed Task Scheduling with Amazon STACK

Contents

[0. Table of Figures 1](#_Toc372489870)

[1. Abstract 2](#_Toc372489871)

[2. Introduction 2](#_Toc372489872)

[3. Background 3](#_Toc372489873)

[4. Motivation 3](#_Toc372489874)

[5. Proposed Solution 3](#_Toc372489875)

[6. Evaluation 3](#_Toc372489876)

[7. Related Work 3](#_Toc372489877)

[8. Conclusion and Future Work 3](#_Toc372489878)

[9. References 3](#_Toc372489879)

# Table of Figures

[Figure 1 : Architecture Diagram 4](#_Toc370105379)

# Abstract

Predictions are that by the end of this decade, we will have exascale system with millions of nodes and billions of threads of execution[1]. Task Scheduling and execution of tasks over these large scale, distributed systems plays an important role in achieving good performance and high system utilization. Many Task computing paradigm[2] aims to bridge the gap between High Performane Computing and High Throughput Computing . Tasks may be small or large, uniprocessor or multiprocessor, compute-intensive or [data-intensive](http://en.wikipedia.org/wiki/Data-intensive_computing) but MTC tasks include loosely coupled communication intensive tasks. Today’s jobschedulers have centralized Master/Slaves architecture (e.g. Slurm, Condor, PBS,SGE), where a centralized server is in charge of the resource provisioning and job execution. This architecture has worked well in modest scales and coarse granular workloads, but it has poor scalability at the extreme scales of petascale systems with fine granular MTC workloads. The goal of this project is to provide a efficient light weight and scalable distributed execution framework built on built on open source stack[HazleCast,Cassandra] to address MTC workloads deployed over Amazon Ec2 instance in cloud environment

# Introduction

The goal of an execution fabric is to effectively utilize the execution system aiming towards high throughput and also provide efficient results for executed tasks. Today’s workload involves a lot of fine granular workloads with execution times in seconds. Centralized schedulers are optimized towards high computational massive tasks where the complex decision policy and architecture of the schedulers play a major role. But they tend to add considerable overhead while scheduling these lots of small tasks. Moreover the centralized architecture tends to be a bottleneck in scheduling and execution. The solution to this problem is to have a decentralized and simple architecture. A decentralized architecture avoids the single point of failure, while a simple architecture reduces the considerable overhead involved in decision making for scheduling.

An execution fabric requires lot of computing resources to address the ever growing workload of today’s world. Clouds seem to be a viable solution to this problem. Our solution is to build a loosely coupled compact and distributed execution fabric over public cloud (Amazon Ec2 instance) with distributed building blocks such as Cassandra [3] and HazleCast [4]. The motivation behind using open source stack in favour of extensive use of AWS [Amazon Web Services] is to decouple the fabric from AWS and provide easier transition to private cloud environment.

Recent studies suggest that clouds were not suitable candidates for scientific HPC computing [5]. The problems listed were largely because of following the same approach involved in traditional clusters and grids. Clouds differ a lot from HPC applications as they are based on virtualization and shared resources.Our work involves running applications optimized for cloud environment. Traditional workloads can also be run on our execution fabric but with suitable decomposition of the workload at the client side.

In this project we implement a scalable distributed task execution framework. We have made extensive use of HazleCast , a highly scalable data distribution platform which acts as the reliable storage for the executable tasks . Our next building block is Cassandra, a distributed NoSQL store offering scalability and high availability for monitoring of the entire system. We also leverage the Amazon Elastic Compute Cloud (EC2) to manage virtual resources.

Today’s data analytics are moving towards shorter jobs with higher throughput and shorter latency. More applications are moving towards running higher number of jobs in order to improve the application throughput and performance. The focus is shifting towards Many Task Computing paradigm. Many task computing includes loosely coupled applications that are generally communication intensive.

We propose CloudKon as a job management system that achieves good load balancing and high system utilization. The heart of the CloudKon is the distributed queuing service. We have used HazleCast to facilitate this purpose. HazelCast performs the role of a highly available and reliable distributed pool of tasks to perfection. Worker Nodes are not adminstered by a centralized dispatcher or scheduler. Our work proposes an efficient pull architecture i.e worker nodes pull the tasks from the pool if they are idle. The system is loosely coupled and each component can be scaled based on the needs.

# Background

# Motivation

The main factors leading to our implementation of CloudKon are as follows

* Design and architect a light-weight task execution framework for MTC workloads
* Design a simple execution framework with a no frills user interface
* Design a robust framework which can easily switch between public and private cloud environments
* Design an extremely scalable execution framework
* Design a loosely coupled framework to support future enhancements
* Evaluate CloudKon with other state-of-the-art task execution systems
* Deliver excellent throughput with <5% codebase of the job management systems

# Proposed Solution

# Evaluation

# Related Work

# Conclusion and Future Work

# References

1. P. Kogge, et. al., “Exascale computing study: Technology challenges in achieving exascale systems,” 2008
2. IEEE Workshop on Many-Task Computing on Grids and Supercomputers (MTAGS08)
3. HazleCast <http://www.hazelcast.com/>
4. Cassandra <http://cassandra.apache.org/>
5. Q. He, S. Zhou, B. Kobler, D. Duffy, and T.McGlynn. “Case study for running HPC applications in public clouds,” In Proc. of ACM Symposium on High Performance Distributed

Computing, 2010.

1. CloudKon <http://www.cs.iit.edu/~iraicu/teaching/CS554-F13/proj/CloudKon-DTS.pdf>
2. D. Thain, T. Tannenbaum, M. Livny, “Distributed Computing in Practice: The Condor Experience” Concurrency and Computation: Practice and Experience 17 (2-4), pp. 323-356, 2005.
3. M. A. Jette et. al, Slurm: Simple linux utility for resource management. In Lecture Notes in Computer Sicence: Proceedings of Job Scheduling Strategies for Prarallel Procesing (JSSPP) 2003 (2002), Springer-Verlag, pp. 44-60
4. B. Bode et. al. “The Portable Batch Scheduler and the Maui Scheduler on Linux Clusters,” Usenix, 4th Annual Linux Showcase & Conference, 2000.
5. I. Raicu, et. al. “Falkon: A Fast and Light-weight tasK executiON Framework,” IEEE/ACM SC 2007
6. ZHT: A Light-weight Reliable Persistent Dynamic Scalable Zero-hop Distributed Hash Table, IPDPS 2013