|epcc|

Hawk-i HPC CLoud Benchmark Tool

Visakh Muraleedharan

August 1, 2012

MSc in High Performance Computing

The University of Edinburgh

Year of Presentation: 2012

Abstract

Scientific computing unravel the mysteries of science by constructing mathematical models and numerical algorithms. This required massive computational power and High-Performance Computing (HPC) solutions like cluster and grid have been the answer to these needs for a long time. These HPC solutions are hard and expensive to setup, maintain and use. *Cloud computing* is a model of delivering the existing compute infrastructure where computation and storage can be dynamically provisioned on a pay as you go model. Using a vendor cloud service like Amazon Web Service (AWS) can significantly reduce the effort to access these on-demand high performance resources.

In this work we create a platform to study how useful Amazon EC2 cloud computing can be for scientific applications. We classify the applications based on the concept of computational *motif* [1]

Contents

List of Tables

List of Figures

Acknowledgements

Chapter 1

Introduction

Chapter 2

Background

2.1 Cloud Computing

2.1.1 Why Cloud Computing?

Resource on-demand

easy, accesibility

Pay as you go

Maintenance and Upgradation

Eco-friendly

2.1.2 Cloud Computing service models

IaaS

PaaS

Saas

2.1.3 Cloud Computing Deployment models

Private Cloud

Public Cloud

Hybrid Cloud

Community Cloud

2.2 Amazon Web Service

2.2.1 Amazon Machine Instance

2.2.2 Types of Instances

2.2.3 Amazon EC2

different instances chart and limitations of microinstance.

EC2 cluster compute

2.3 HPC in the cloud

Introduction Paragraph, comparison with Clusters

2.3.1 Advantages

2.3.2 Limitation

Network Limitations, Availability, Stability of performance Benifits

2.3.3 Amazon Cluster compute

2.4 Parallel Applications and 13 Dwarfs

Explain the 13 dwarfs and under the two sections explain N-body and FFT

2.4.1 N-Body Methods

MD, About the dwarf and application types optimisation, n2 and nlogn Communication pattern!!!

2.4.2 Spectral Methods

In scientific computing, a class of techniques used to numerically solve differential equations involving the use of Fast Fouries Transform are called spectral methods

Fourier transform Jean Baptiste Joseph Fourier (1768-1830) first employed what we now call Fourier transforms whilst working on the theory of heat Linear transform which takes temporal or spatial information and converts into information which lies in the frequency domain Who would use Fourier Transforms? âĂć Physics âĂŞ Cosmology (P3M N-body solvers) âĂŞ Fluid mechanics âĂŞ Quantum physics âĂŞ Signal and image processing âĂŞ Antenna studies âĂŞ Optics

âĂć Numerical analysis âĂŞ Linear systems analysis âĂŞ Boundary value problems âĂŞ Large integer multiplication (Prime finding)

âĂć Statistics âĂŞ Random process modelling âĂŞ Probability theory

Discrete Fourier Transform The discrete Fourier transform of N complex points fk is defined as

Communication pattern!!!

Chapter 3

Live Benchmark Tool Setup

3.1 System Design

3.1.1 Architecture

3.1.2 Sequence diagram

3.1.3 Database Design

3.2 Instance types Used

Refer to graph in Background Why these types?

3.3 Sun Grid Engine Cluster

Why sungrid Engine

3.4 SGE Clustering in Amazon cloud

benchmark to show how cluster instances are faster than normal high cpu instances

3.5 Web interface

Describe in detail

3.5.1 Admin panel

3.5.2 User Dashboard

Chapter 4

Results and Analysis

4.1 Serial

for each instance type time to result, execution time, increasing problem size N-body and FFT

4.2 Parallel program

execution time, increasing problem size, number of cores N-body and FFT

4.3 Stability of results

History of execution, snapshot from dashboard

Chapter 5

Conclusions

Appendix A

Cluster Computing Setup

A.0.1 Using Starcluster

A.0.2 Cluster Management package

A.0.3 Using Sun Grid Engine

A.0.4 Creating Dashboard

Appendix B

Benchmarking programs

B.0.5 Spectral Methods

B.0.6 N-body

Bibliography

[1] K. Asanovic, R. Bodik, J. Demmel, T. Keaveny, K. Keutzer, J.D. Kubiatowicz, E.A. Lee, N. Morgan, G. Necula, D.A. Patterson, et al. The parallel computing laboratory at uc berkeley: A research agenda based on the berkeley view. *EECS Department, University of California, Berkeley, Tech. Rep*, 2008.