HPC In The Cloud?

Michael Kleber

Department of Computer Sciences
University of Salzburg, Austria

July 2, 2012

Content

Motivation

- wide spread availability of cloud services
- easy access to compute resources
- easy to scale
- pay-as-you-go

Questions

- Performance?
- suitable for every HPC application?
- shortcomings and problems?
- cost-effective?

Cloud
Amazon

Amazon EC2

- well known
- available to everyone
- CCI targets HPC user

Cloud

- virtualisation
- processor type per instance is not specified
- performance (CU) and Memory are specified
- 1GbE or 10GbE (CCI)
- (real) network capacity, quality and topology are not known

HPC Cluster

- no virtualisation
- processor type and Memory ar known
- InfiniBand, HT-Link or ethernet (1Gb oder 10Gb)
- several network topologies, e.g. 2D-mesh
- (real) capacity and quality are known

Outline

Overview

- EC2 CCI versus local cluster [1]
- "synthetic benchmarks"
- network performance
- compute performance
- parallel I/O
- performance variation
- cost-effective

EC2 CCI Setup

- 2 quad core Xeon X5570, 23GiB RAM
- 10GbE
- per instance LBS 1690 GiB
- EBS + S3
- Amazon Linux 64 Bit
- Intel compiler

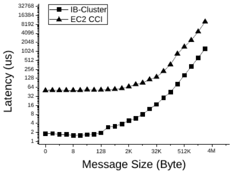
Cluster Setup

- six core Xeon 5670, 32 GiB RAM
- QDR InfiniBand
- NFS
- Intel compiler
- similar compute performance

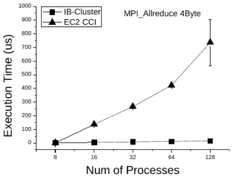
Benchmarks

- Intel MPI
- NBP

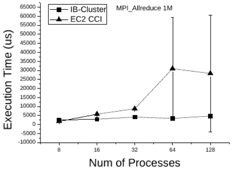
Intel OpenMPI



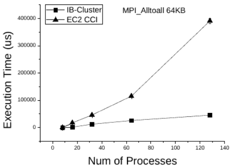
(a) Ping-Pong



(b) MPI_Allreduce (4B)

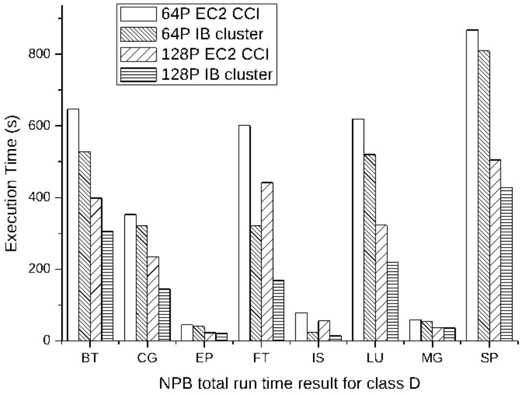


(c) MPI_Allreduce (1 MB)



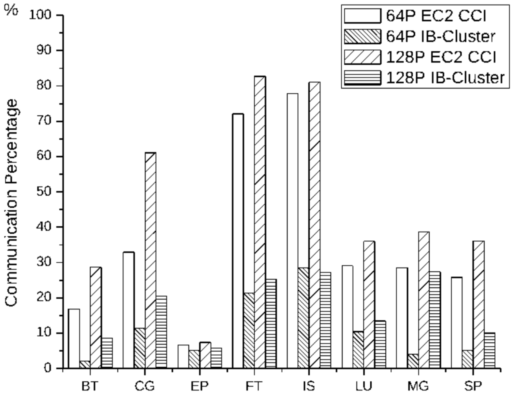
(d) MPI_Alltoall (64 KB)

NBP
Completion Time



NPB

Communication



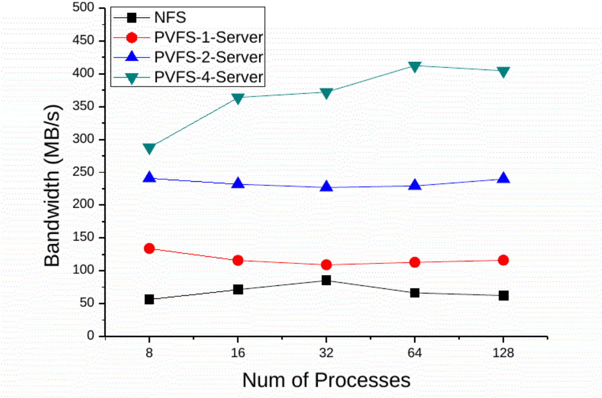
Parallel I/O

Introduction

- NFS common on small and some mid range clusters, parallel file systems are popular on large clusters
- some parallel file systems are PVFS, GPFS and Lustre
- changing the file system every session is only possible with cloud services
- tested file systems:
  - NFS
  - PVFS
- benchmarks:
  - IOR micro benchmarks
  - BT-IO

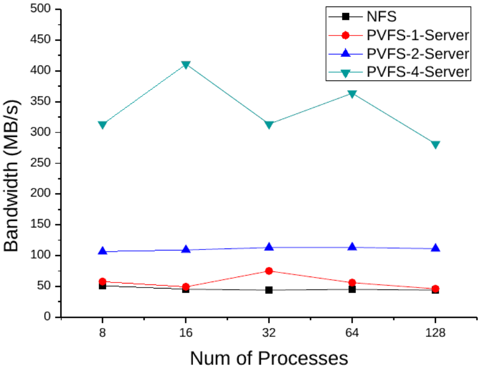
Parallel I/O

IOR Read Rate



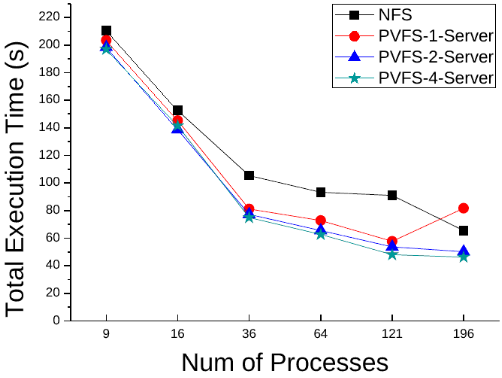
Parallel I/O

IOR Write Rate

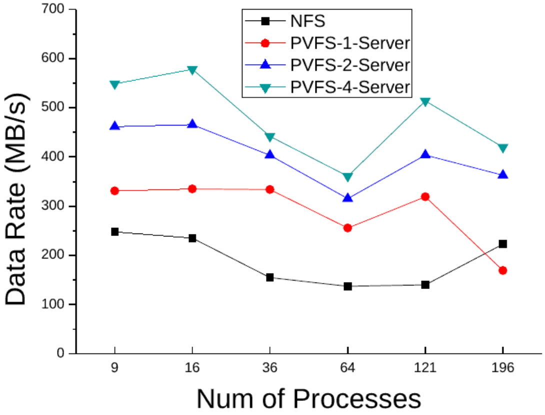


Parallel I/O

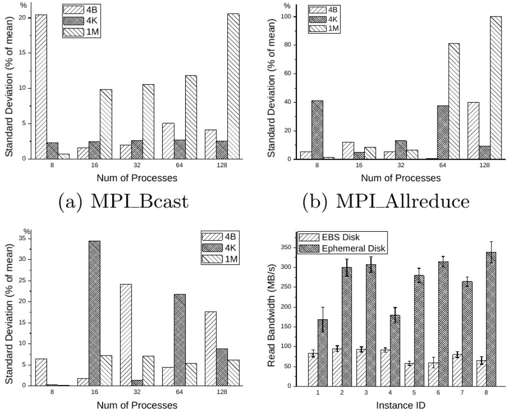
BT-IO Completion Time



Parallel I/O
BT-IO Data Rate



Performance Variation



(a) MPI_Bcast
(b) MPI_Allreduce
(c) MPI_Alltoall
(d) Disk Read Bandwidth

Outline

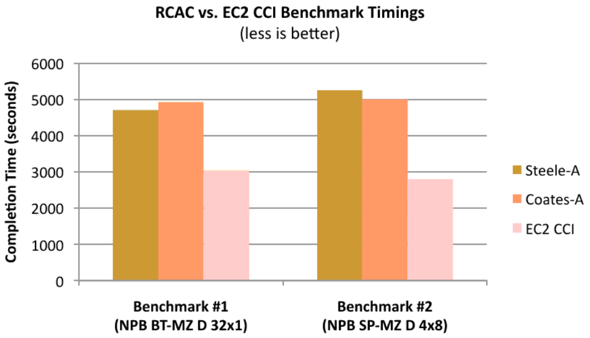
Background

- case study by Perdue [2]
- 2 community cluster
  - Steele: 902 nodes, Xeon E5410, utilisation 68.4%
  - Coates: 993 nodes, Opteron 2380, utilisation 80.1%
- Amazon EC2 CCI: 1.6$ per node hour + storage => ca 1.75$

Cluster Cost

Resource	Q2 2010 Core-Hours	Nodes Purchased	TCO per Node	EC2 Equivalent / Node-Hour
Steele	10,815,169	902	$7,643.10	$0.42
Coates	13,936,475	993	$8,315.82	$0.40

Performance



Outline

MUSCLE

- Multi-Scale Simulation MUSCLE [3]
- usage examples:
  - blood flow
  - solid tumor models
  - stellar system
  - virtual reactor

Setup

- in-stent restenosis application
- 15 and 150 iterations
- HP cluster platform with 2.26 GHz Xeon processors, InfiniBand
- EC2 high CPU extra large instances including 7 GiB of RAM, 20 CUs (8 virtual cores) and 1690 GiB of local storage
- EC2 cluster compute quadruple extra large including 23 GiB of RAM, 33.5 CUs and 1690 GiB of local storage

Performance

ISR 2D 15 iterations								
Infrastructure	**Setting up**		**Execution**		**Sending Output**		**Total**	
	min - max (sec)		avg (sec)	σ			min - max (sec)	
Local HPC Cluster	6 - 363		190	16	N/A		196 - 553	
	avg(sec)	σ	avg (sec)	σ	avg(sec)	σ	avg(sec)	σ
AWS Cloud m1.xlarge	81	6	250	20	100	10	430	20
m2.4xlarge	80	10	187	3	130	20	400	20
ISR 2D 150 iterations								
Infrastructure	**Setting up**		**Execution**		**Sending Output**		**Total**	
	min - max (sec)		Avr (sec)	σ			min - max (sec)	
Local HPC Cluster	6 - 363		1500	130	N/A		1506 - 1863	
	avg(sec)	σ	avg (sec)	σ	avg(sec)	σ	avg(sec)	σ
AWS Cloud m1.xlarge	72	4	2068	15	120	20	2260	30
m2.4xlarge	74	4	1526	4	110	60	1710	60

Outline

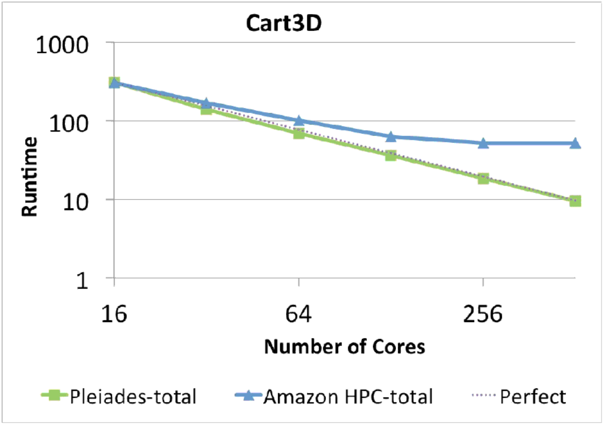
Setup

- NASA Benchmarks [4]
- Amazon EC2 quadruple extra large, NFS, SLES 11, Intel compiler 12.4, OpenMPI 1.4.4
- Pleiades: SGI ICE with Xeon X5570, 24 GiB of RAM, InfiniBand 4XQDR, MPT 2.04, Intel compiler 12.4, Lustre parallel FS

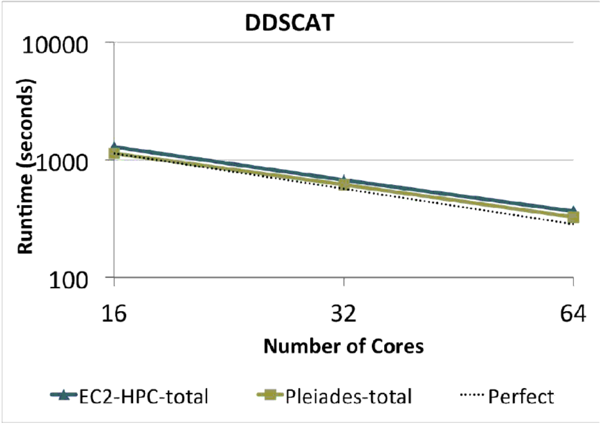
Applications

- 4 real world applications
- Cart3D - CFD fluid simulation
- DDSCAT - calculates absorption and scattering properties
- Enzo - to simulate cosmological structure formation
- MITgen - global ocean simulation model

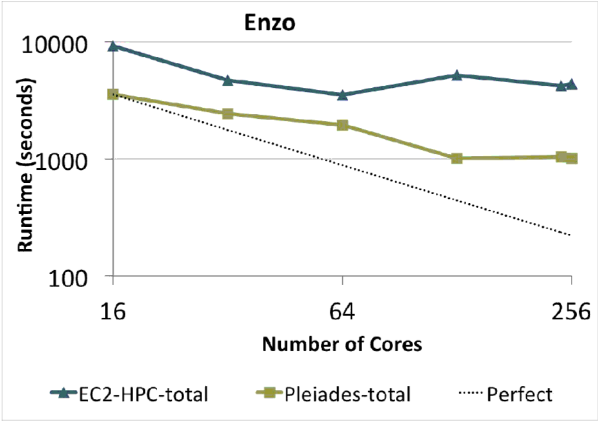
Cart3D


Cart3D
1000
100
10
1
Runtime
16
64
256
Number of Cores
Pleiades-total
Amazon HPC-total
Perfect

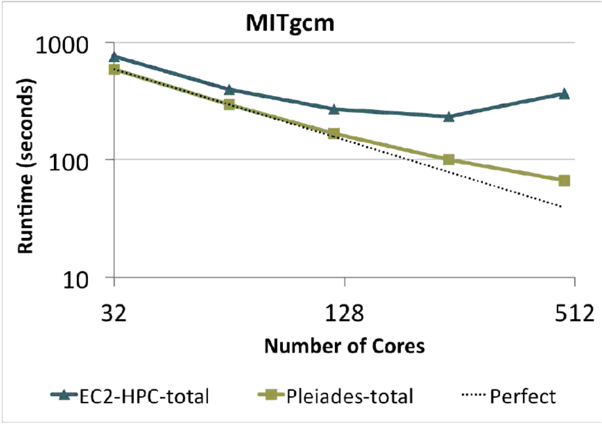
DDSCAT


DDSCAT
10000
1000
100
Runtime (seconds)
16
32
64
Number of Cores
EC2-HPC-total
Pleiades-total
Perfect

Enzo


Enzo
10000
1000
100
Runtime (seconds)
16
64
256
Number of Cores
EC2-HPC-total
Pleiades-total
Perfect

MITgen


MITgcm
1000
100
10
Runtime (seconds)
32
128
512
Number of Cores
EC2-HPC-total
Pleiades-total
Perfect

Conclusion

- good suited for tigthly coupled application
- high variance
- comparable costs
- easy to setup, low maintanance
- easy to switch filesystems
- no waiting time

Reference

1 Cloud Versus In-house Cluster: Evaluating Amazon Cluster Compute Instances for Running MPI Applications
2 Cost-effective HPC: The Community or the Cloud?
3 Comparison of Cloud and Local HPC approach for MUSCLE-based Multiscale Simulations
4 Performance Evaluation of Amazon EC2 for NASA HPC Applications