Case Study for Running HPC Applications in Public Clouds

OPEN RESEARCH, INC.
QIMING HE*

NASA-GODDARD SPACE FLIGHT CENTER (GSFC)
SHUJIA ZHOU, BEN KOBLER, DAN DUFFY, TOM MCGLYNN

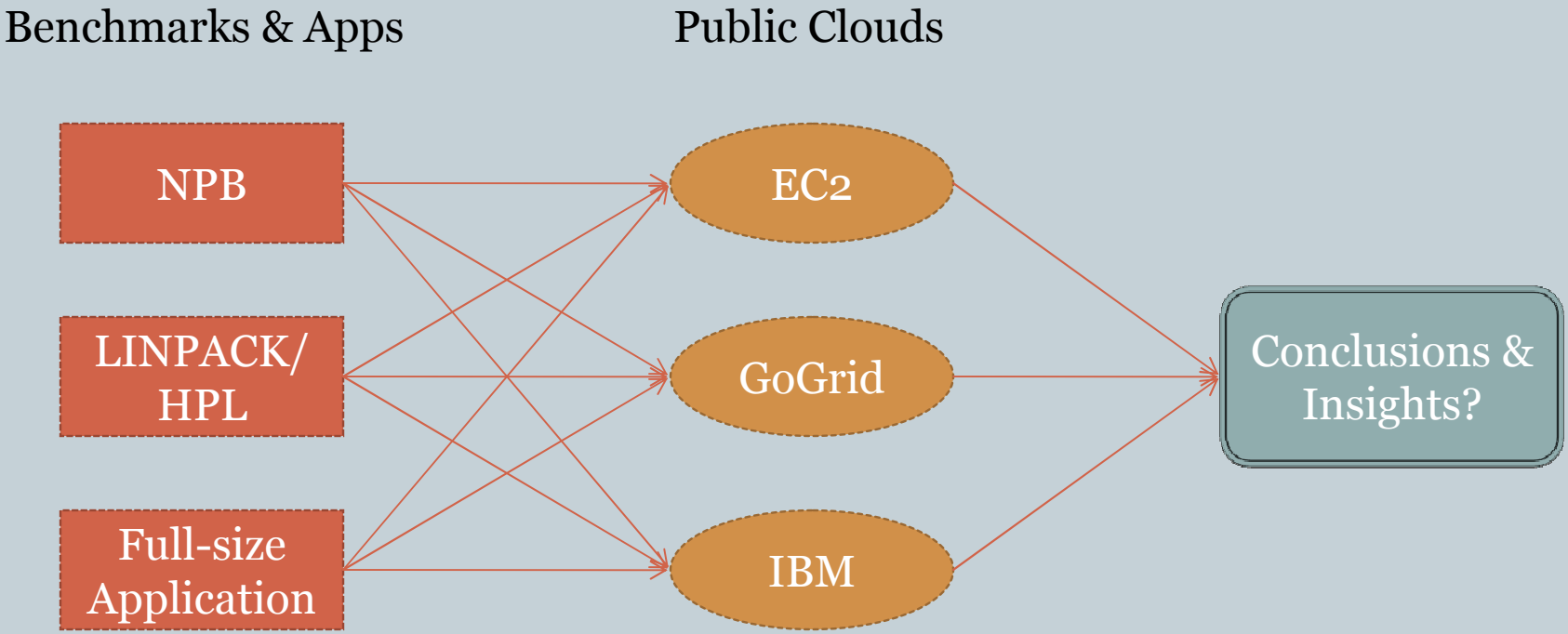
Motivation: HPC-in-the-cloud

- HPC: MPI-based parallel computing programs on distributed-memory systems.
- Public clouds:
  - Virtualized computing resources
  - Pay-as-you-go pricing model.
- Technically feasible?
  - Deploy and execute any-size app in any cloud?
  - Satisfactory performance vs. in-house HPC systems?
  - “HPC-friendly” clouds?
  - “HPC-friendly” application type?
- Economically feasible?
  - Cloud servers costs vs. local cost of person, facilities, energy,…

Previous Work

- Ed. Walker, "*Benchmarking Amazon EC2 for high-performance scientific computing*" , 2008.
  - NPB Benchmark on EC2 vs. NCSA
    - 5-20% performance gap for NPB-OMP.
    - 10%-1000% performance gap for NPB-MPI.
- Jeff. Napper, et. al. , "*Can cloud computing reach the top500?*", 2009.
  - LINPACK/HPL Benchmark on EC2
    - 1) not scalable as problem and cluster size increase.
    - 2) not economical in terms of FLOPS/$.
- Why EC2 only?

Our Approach

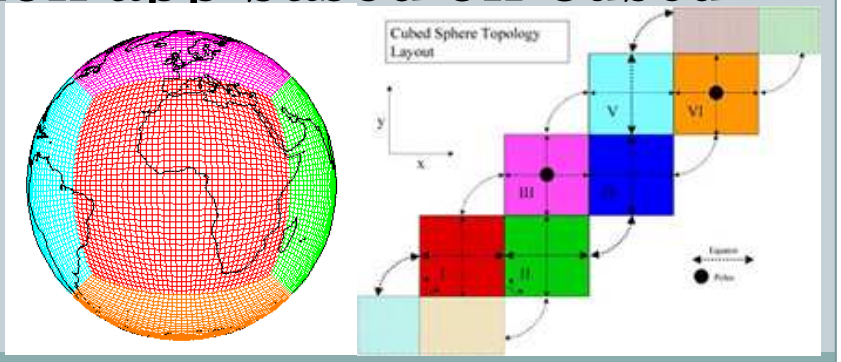


Benchmarks and Applications

- NPB (from NASA)

Benchmark	Name derived from	Description
BT	Block Tridiagonal	Solve a synthetic system of nonlinear PDEs using algorithms block tridiagonal.
CG	Conjugate Gradient	Estimate the largest eigenvalue of a sparse symmetric positive-definite matrix using the inverse iteration with the conjugate gradient method as a subroutine for solving systems of linear equations
EP	Embarrassingly Parallel	Generate independent Gaussian random variates using the Marsaglia polar method
FT	Fast Fourier Transform	Solve a three-dimensional partial differential equation (PDE) using the fast Fourier transform (FFT)
IS	Integer Sort	Sort small integers using the bucket sort
LU	Lower-Upper symmetric Gauss-Seidel	Solve a synthetic system of nonlinear PDEs using LU-SSOR algorithm
MG	MultiGrid	Approximate the solution to a three-dimensional discrete Poisson equation using the V-cycle multigrid method
SP	Scalar Pentadiagonal	Solve a synthetic system of nonlinear PDEs using scalar pentadiagonal algorithm

- LINAPCK/HPL (used by TOP500)
- CSFV (NASA weather prediction app based on Cubed-Sphere Finite-Volume Core)



Public Clouds Hardware Specifications

- EC2 (Xeon E5345 @2.33GHz)

Server instance	RAM(GB)	Core	Disk(GB)	Price/hr
Small	1.7	1	160	$0.085
Large	7.5	2	850	$0.34
Extra Large	15	4	1690	$0.68
Double Extra Large	34.2	4	850	$1.20
Quad Extra Large	68.4	8	1690	$2.40
High-CPU medium	1.7	2	350	$0.17
High-CPU Extra Large*	7	8	1690	$0.68

- GoGrid(Xeon E5459@3GHz)

Server Instance	RAM(GB)	Core	Disk(GB)	Price/hr
0.5G	0.5	1	30	$0.095
1G	1	1	60	$0.19
2G	2	1	120	$0.38
4G	4	3	240	$0.76
8G*	8	6	480	$1.52

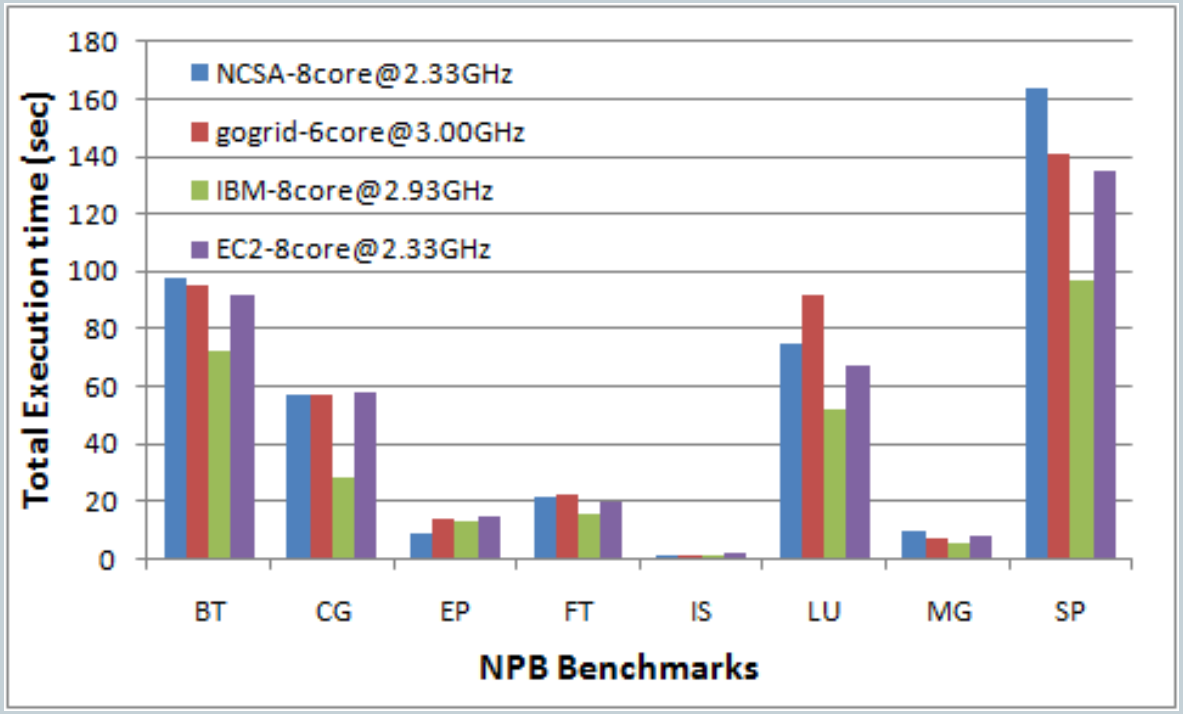
Now 8 cores

- IBM(Nehalem X5570@2.93GHz)

Server Instance	RAM(GB)	Core	Disk(GB)	Price/hr
Small	1	2	8	N/A
Medium	1.8	4	18	N/A
Large*	3.6	8	38	N/A

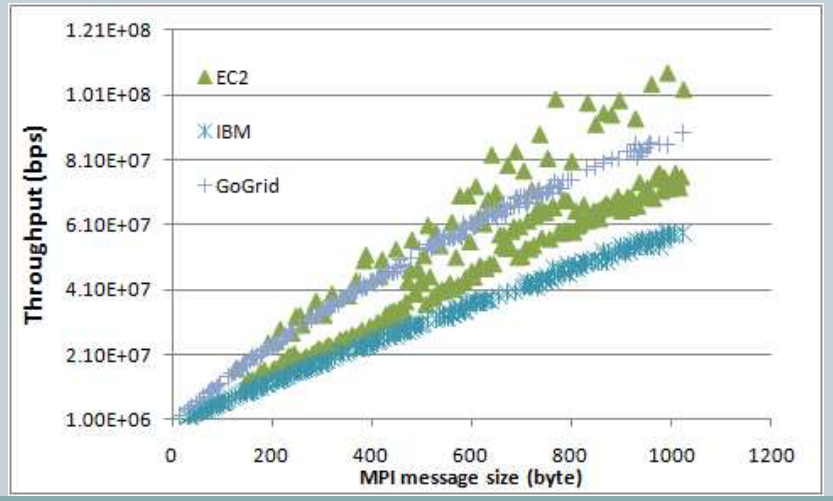
NPB-OMP

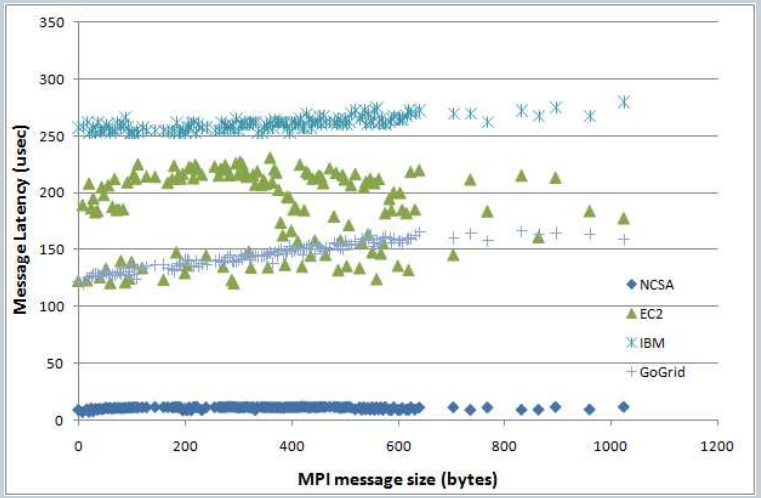
- NPB OpenMP Benchmark (No MPI over network)
- To evaluate virtualization overhead per node



Cloud Networking

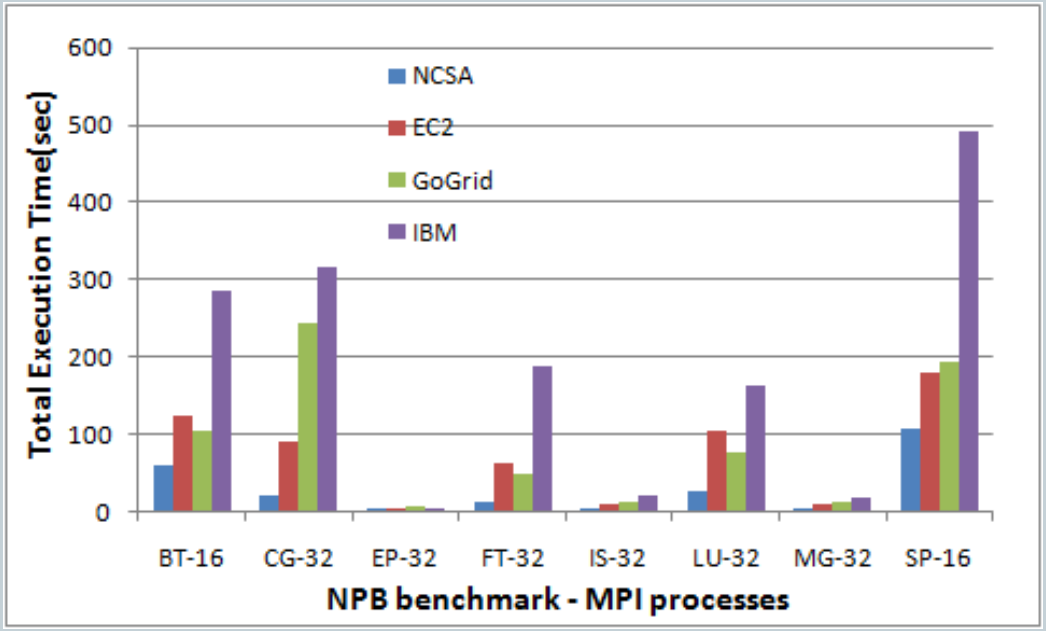
- Latency and Throughput (MPI Messages)
  - IBM (100Mbps): consistently slow
  - EC2 (1000Mbps?): inconsistently fast, multi-hop IP packets
  - GoGrid (1000Mbps): consistently fast
  - But all three are orders of magnitude slower than NCSA (with InfiniBand)





NPB-MPI

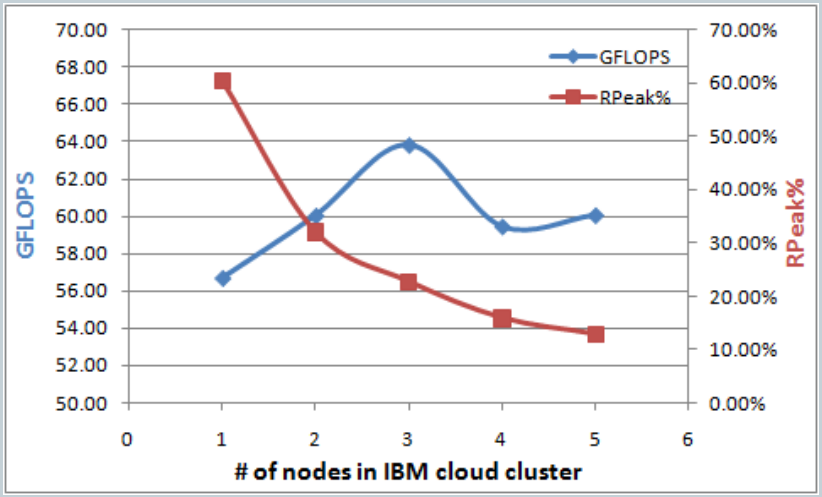
- Re-run Ed Walker's TestCase in three clouds

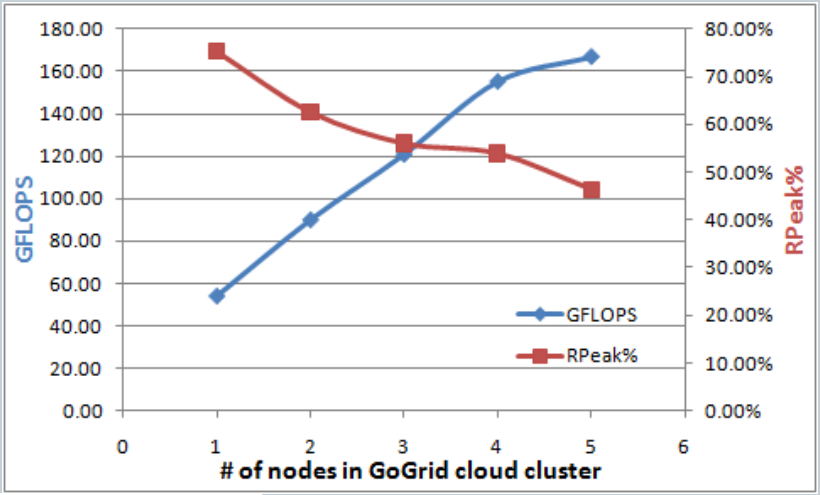


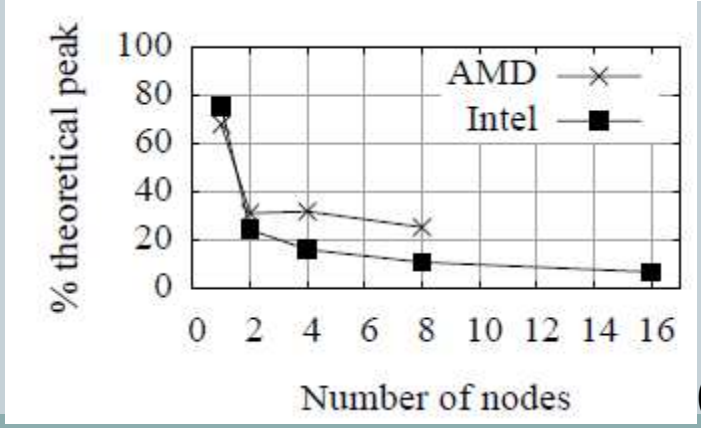
- Better network, better performance (in line with earlier researchers)
- GoGrid could be better if it has 8-core servers.

HPL

- Re-run Jeff Naper's TesCases in three clouds



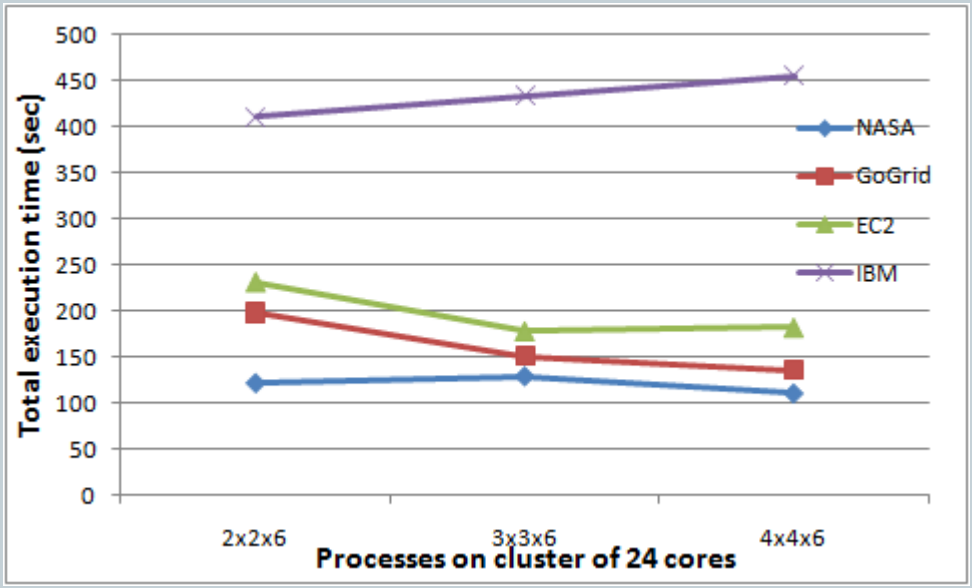




(from Naper's paper)

CSFV

- Full-size app (with 110K line of Fortran/MPI code).
- Compile- and run-time dependencies, e.g., NETCDF.



- Over-subscription helps (a little).
- Best-case scenario: only underperforms NASA system by 20%

Other Observations

- Programming paradigm matters
  - MPI, OpenMP, MPI+OpenMP
- Embarrassingly Parallel (EP) app works the best.
  - EP represents a wide-spectrum of apps.
- Commercial vs. Open source software in the cloud
  - No performance difference observed.
- Cloud economics
  - FLOPS/$, FLOPS/watt,…

Future Work

- Benchmarks with larger size
- Re-do GoGrid with 8-core server
- More public clouds
  - Terremark's vCloud Express (vmware-based)
  - Penguin Computing's HPC-as-a-Service™
  - IBM upgrades to (at least) 1GbE network soon?
- Private Clouds
  - NASA Nebula
  - DOE Magellan
- Q&A