

**CASE STUDY FOR RUNNING HPC
APPLICATIONS IN PUBLIC CLOUDS**

INTRODUÇÃO

- *Cloud computing* está a surgir como uma alternativa à computação tradicional. Nos últimos anos temos assistido a um aumento da utilização da *cloud computing* nas áreas da ciência, tais como:
 - Física Nuclear
 - Energia de Alta Tensão
 - Bioinformática
 - Astronomia
 - Investigação Climatérica
- É uma excelente ferramenta de colaboração que permite aos cientistas partilhar informação a nível global, replicar trabalhos e conjuntos de dados.



OBJECTIVO

- Alargar uma pesquisa realizada anteriormente a três plataformas públicas de Cloud Computing;
- Concentrar-se na avaliação das capacidades técnicas das actuais plataformas públicas de *Cloud Computing*;
- Execução de aplicações científicas, especialmente HPC (*High Performance Computing*);
- Comparar os resultados obtidos com os resultados de sistemas dedicados para aplicações HPC.
- Discutir técnicas que poderão ajudar cientistas para uma melhor utilização das plataformas públicas apesar das actuais anomalias.



BENCHMARKS E APLICAÇÕES

- NPB (NAS Paralell Benchmark) – desenvolvido pela NASA/NAS para avaliar o desempenho de supercomputadores paralelos.
- HPL (High Performance LINPACK), é um benchmark pertencente ao TOP500, utilizado em algumas aplicações científicas.
- CSFV (Cubed Sphere Finite-Volume) – aplicação NASA para previsão do tempo



AMAZON EC2 CLOUD

- Plataforma mais popular de Cloud Computing e é alvo para inúmeras aplicações acadêmicas e comerciais.
- Utiliza um processador quad-core Intel Xeon E5245@2.33GHz.

Server instance	RAM(GB)	Cores	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 CLOUD

- o Plataforma *GoGrid* utiliza processadores Intel 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



IBM CLOUD

- A IBM disponibiliza uma plataforma de *cloud computing*, actualmente em versão beta e gratuita.
- Utiliza um processador Intel 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



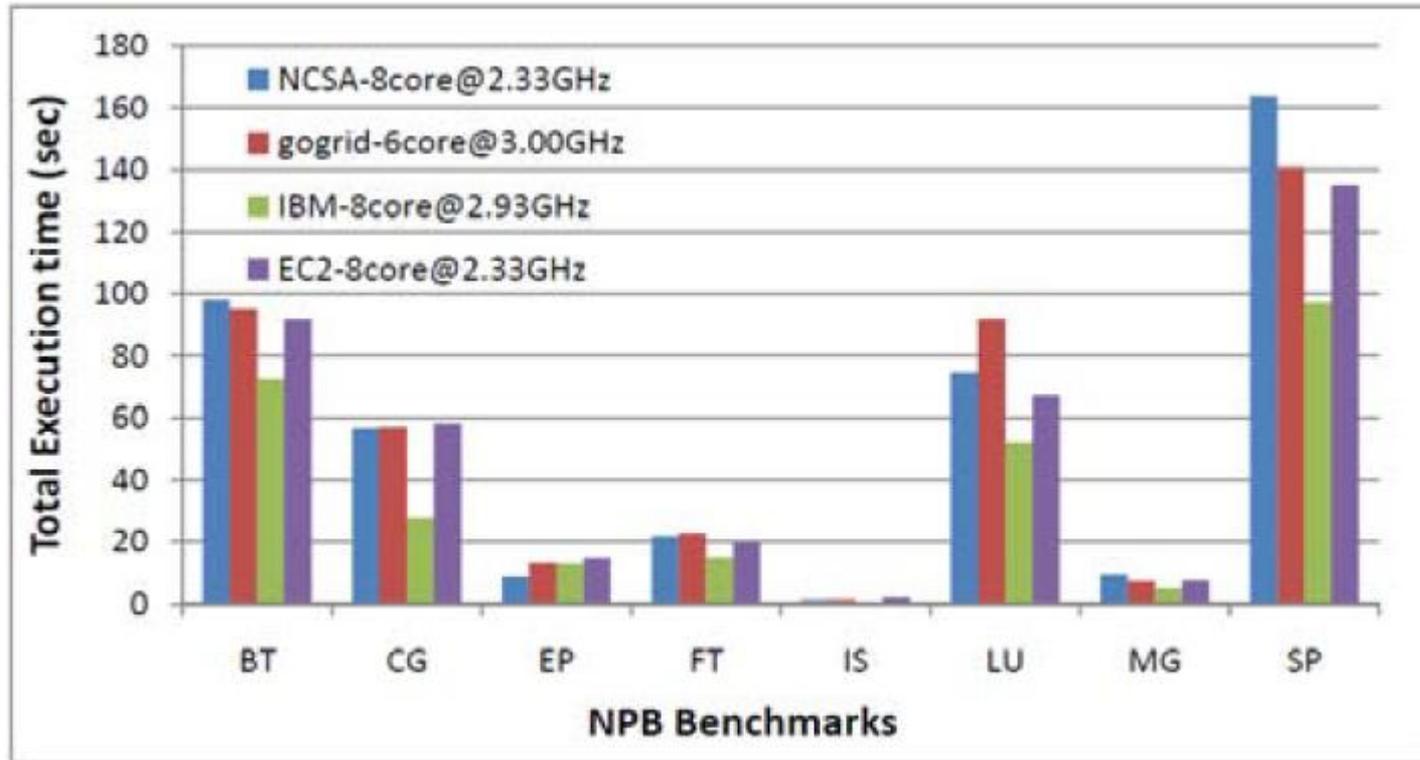
RPEAK

Quadro informativo do desempenho das várias plataformas Clouds

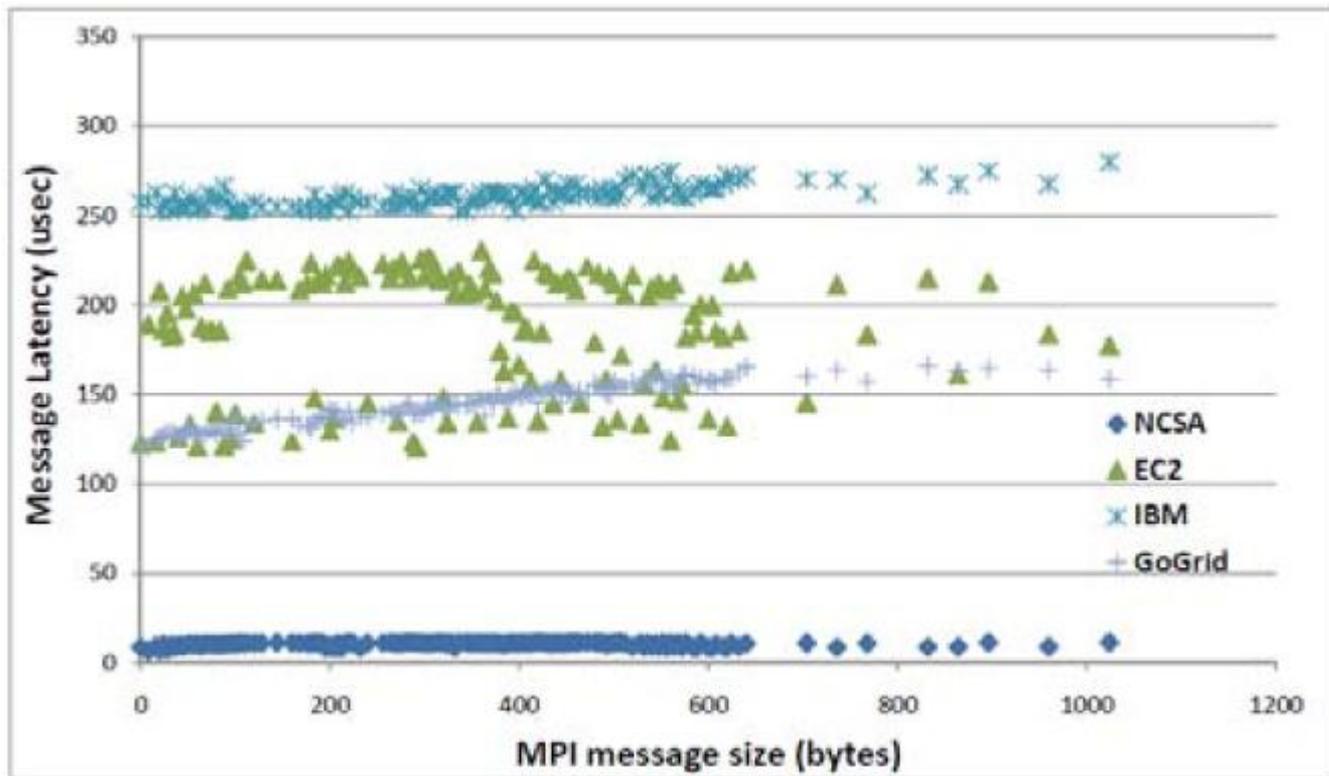
Server	CPU (GHz)	Hyper-visor	Processor/node	Core/Processor	RPeak (GFLOPS)
EC2	2.33	Xen	2	4	74.56
GoGrid	3.00	Xen	1	6	72.00
IBM	2.93	?	2	4	93.76



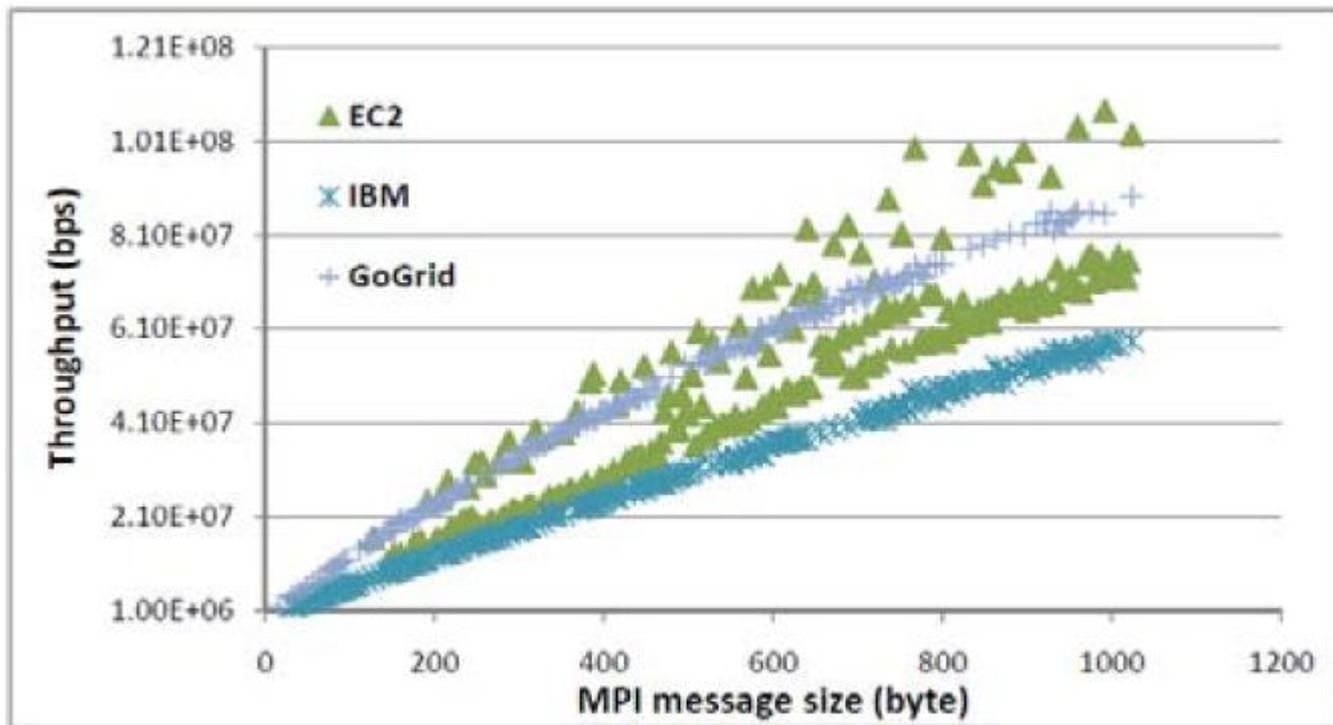
NPB OMP EM CLOUDS PUBLICAS



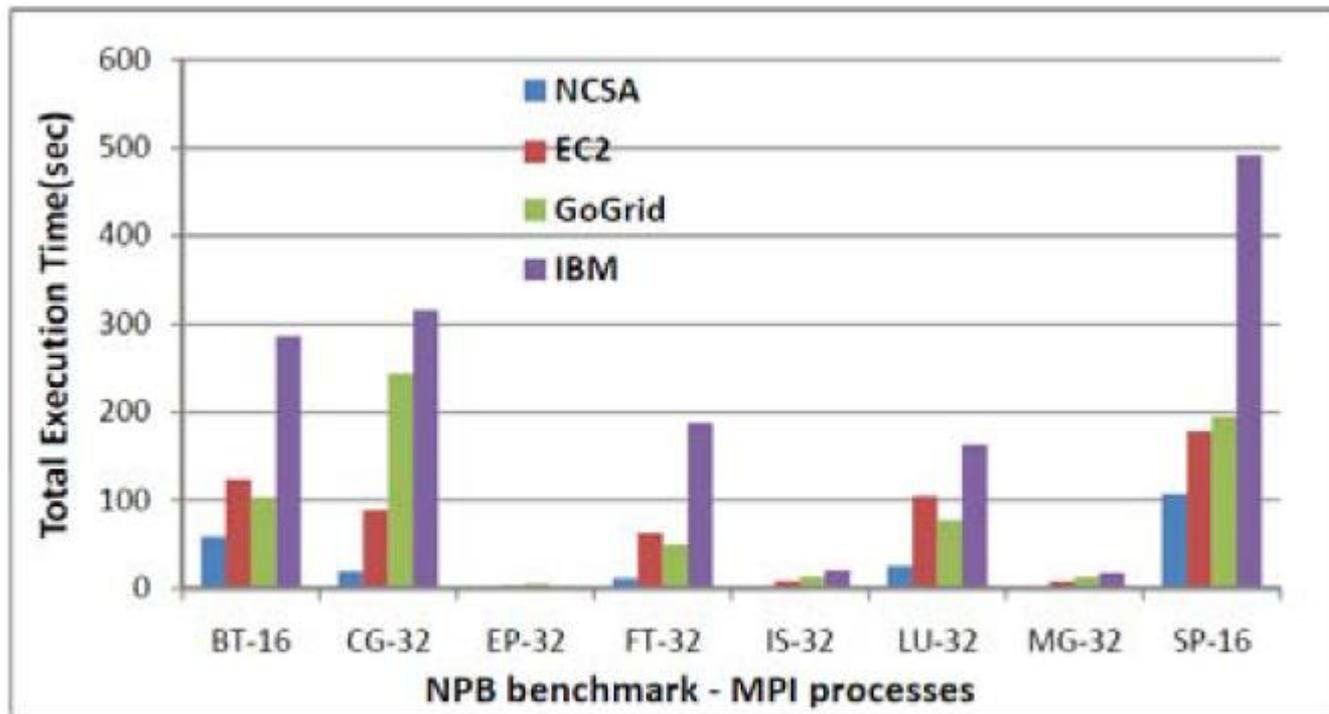
MPI LATÊNCIA DA MENSAGEM



MPI TRANSFERÊNCIA DE MENSAGENS

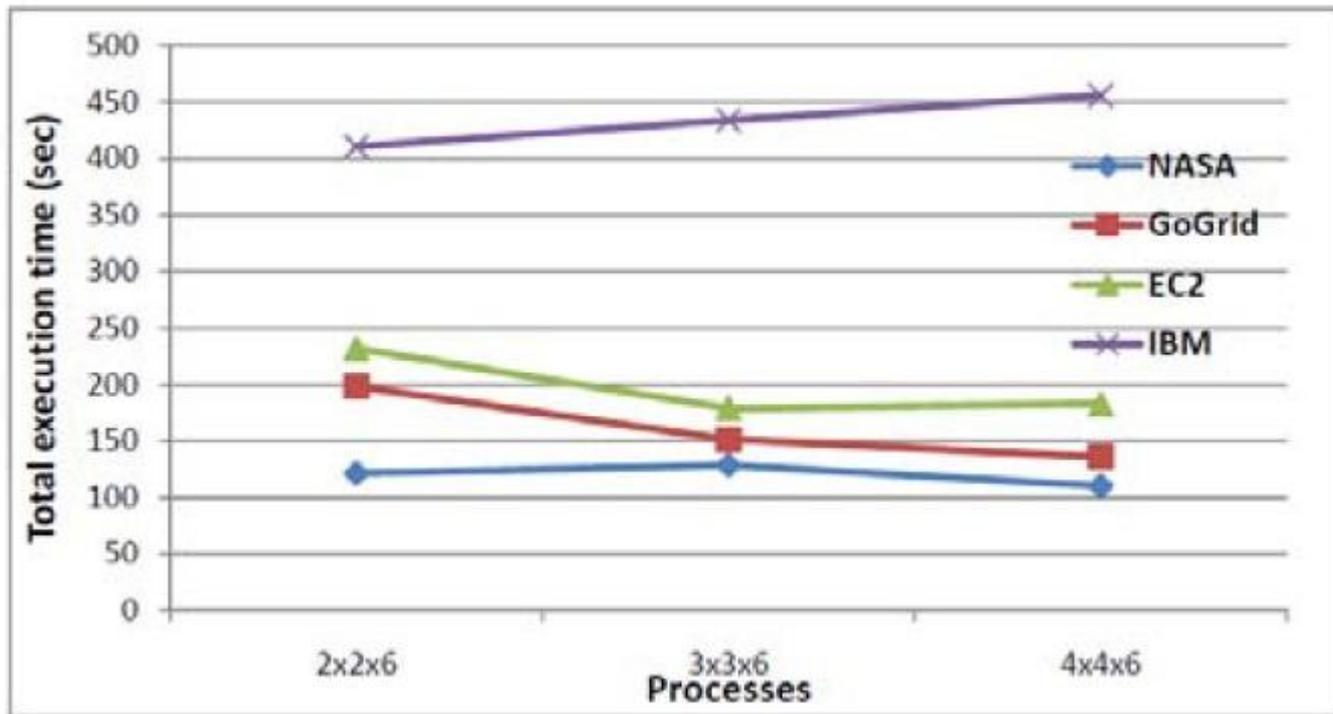


NPB-MPI RESULTADOS



CSFV

CLOUDS PÚBLICAS VS SISTEMA NASA



ANÁLISES E RESULTADOS

- IBM tem maior latência devido as suas redes lentas.
- EC2 tem a maior variação
- A rede GoGrid supera as outras duas plataformas em termos de latência média e variação.



OUTRAS OBSERVAÇÕES

- *Open Source* vs Software comercial
 - Não são observadas quaisquer diferenças de desempenho
- Embarrassingly Parallel (EP) benchmark é a aplicação com melhores resultados na *cloud*.
- As plataformas de Cloud Computing disponibilizam diversos serviços sendo que em alguns casos os utilizadores optam por soluções mais económicas.



TRABALHO FUTURO

- Analisar mais plataformas públicas de Cloud Computing
 - Penguin Computing HPC-as-a-Service.
- Analisar plataformas privadas, tais como:
 - NASA Nebula
 - DOE Magellan



CONCLUSÃO

- A maioria destas plataformas possui comunicações lentas, o que dificulta a execução deste tipo de aplicações.
- Apesar de não haver uma plataforma ideal para o desenvolvimento de aplicações HPC, a plataforma com o melhor desempenho foi a da IBM.



QUESTÕES ?

