

HPC IN THE CLOUD - MASTER OF HPC

PRACTICAL ASSIGNMENT: VIRTUAL CLUSTERS

Scalability evaluation of Computational Fluid Dynamics (CFD) kernels in virtual clusters

Objective

Design and perform an experiment, using the CFD kernels of the MPI implementation of the NAS Parallel Benchmarks (NPB) suite, to evaluate the scalability of computation-intensive and communication-intensive programs in AWS HPC and HTC virtual clusters.

Description

In this assignment you have to design and perform an experiment to evaluate the scalability of HPC programs in AWS virtual clusters. **Start by reading the article and the blog referenced at the end of this document**, to learn the basics of scalability assessments. **Then watch the video(s) you can find in Aula Cesga** about how to deploy a virtual cluster on AWS and run an MPI job.

The experiment you design has to fulfil the following conditions:

1. You have to use two different NPB kernels of the MPI implementation, one intensive in communications and the other intensive in computation.
2. For the communication-intensive kernel, you have to perform the evaluation using two virtual clusters, an HPC and an HTC, and compare the results. Use the same type of instances in both virtual clusters.
3. For the computation-intensive kernel, you have to perform the evaluation using two virtual clusters of the same type, both HPC or both HTC, and compare the results. Use general-purpose instances in one of them and computing-intensive instances in the other. To highlight performance differences, select instances with as different configurations as possible (e.g. from different families and generations).
4. The result for each configuration evaluated in the experiment will be the average of a total of 5 runs.

The rest of the parameters of the experiment are left to your choice, but they have to be adequately justified (e.g. in terms of execution time, scalability, costs, etc).

Important notices

You'll be charged for the resources you use in the experiment. Keep this in mind when choosing the NPB problem size and kernels (i.e. take into account their execution times), instance types, etc. **Estimate the cost of your experiment in advance!**
<https://calculator.s3.amazonaws.com/index.html>

Deliverables

Your submission has to include at least the following:

1. The NPB configuration files of your experiment: `make.def` and `suite.def`
2. Any script that you use (e.g. job submission scripts) in the experiment.
3. A spreadsheet containing the experiment raw data (i.e. the execution times of all the runs of each tested configuration).
4. An experiment report, **in PDF format**, with at least the following sections:
 - *Experimental configuration*: explain the configuration of your experiment in detail with a justification of the parameters you chose. Include screenshots of the AWS CloudFormation and EC2 consoles, showing detailed information about the stacks and nodes for each virtual cluster used in your experiment.
 - *Experiment results*: using tables and charts, summarize the results (i.e. execution time, speedup and parallel efficiency) for each NPB kernel and explain your conclusions.

For the language of the report, Galician, Spanish, Portuguese or English are welcome.

Submission instructions and due date

Use the assignment in Aula Cesga to upload a **compressed file** containing your submission. **The assignment is due by 23:59 PM, December 17th.**

References

- NAS Parallel Benchmarks (NPBs)
<https://www.nas.nasa.gov/publications/npb.html>
- **Article**: Performance analysis of HPC applications in the cloud
Some contents of this article are a bit outdated (e.g. instance types, network virtualization support, network bandwidth) but the experimental setup and conclusions remain valid. Sections of interest for this assignment are: 1, 3, 4, 5.3, 5.4 and 6.
<http://www.des.udc.es/~juan/papers/fgcs2013.pdf>
- **Blog**: Real World AWS Scalability
<https://aws.amazon.com/blogs/compute/real-world-aws-scalability/>

Additional materials

- **Repository**: CloudFormation template to deploy virtual clusters on AWS
<https://github.com/micap-hpcn/virtual-clusters>
- **Video**: how to deploy a virtual cluster and run an MPI job
Available in Aula Cesga.