Multivariate Analysis of Variance for High Energy Physics Software in Virtualized Environments

Víctor Fernández Albor, Marcos Seco, Víctor Méndez Muñoz, Tomás Fernández Pena, Juan Saborido Silva and Ricardo Graciani Diaz
Introduction

- Cloud benefits
  ● Flexible
  ● Reliable
  ● Scalable
- Computing center migration
- EGI FedCloud

Communities of different fields are served using cloud resources nowadays:
- Physics - BioMed - Life sciences etc...
Introduction

-One work based in HEP software performances such A. Chierici, R. Veraldi


Is the virtualized performance of typical clouds sufficient for HEP software and which are the suitable set-ups in that case?

-> Multivariate Analysis of Variance MANOVA
<table>
<thead>
<tr>
<th>HEP classification</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>1  AIRES used by Auger experiment</td>
</tr>
<tr>
<td>Event reconstruction</td>
<td>2  FastJet used by ALICE, ATLAS experiment</td>
</tr>
<tr>
<td>Physics analysis</td>
<td>1  GAUSS used by LHCb experiment</td>
</tr>
</tbody>
</table>
Test-bed and experimental set-up

Statement of the problem
- Study the performance of HEP scientific software on VM and BM

Subject matter
- Virtualization
- Amount of memory
- Number of CPU cores
- Length of the job
Test-bed and experimental set-up

Design aspects

- Treatment design: using a mixed factorial design $2^3 \times 3^2$

2 Level factors 3 Level factors
VM  No-VM  Core number (1-2-3)
2 GB  4 GB  Benchmarks  (HEP-Soft.)

  - Short job  Long job

- Error control: four times with different random seeds

- Sampling and data taking: maximization of CPU resources

  Response choice = job walltime
Test-bed and experimental set-up

VM
Scientific Linux 6.3 with gcc v.4.4.6.
CVMFS client v. 2.0.19-1.el6.

Host operating system
KVM 0.12.3+noroms-0ubuntu9.21
libvirt v. 0.7.5-5ubuntu27.23
Ubuntu 10.04
Selection of test-bed resources

Preliminary evaluation of platforms
- DELL Poweredge 1950 Intel Xeon 5160
- Supermicro 2022TG-HTRF AMD 6128 @2GHz

Architecture Performances Problems → GAUSS test was of 20-30% for AMD but of the order 5% for Intel.
Selection of test-bed resources

perf tracing:
loss performance traced to a specific function in GEANT4 software.

=> no-cpu flags in virtualization: 3dnow and 3dnowext

DELL Poweredge 1950 Intel Xeon 5160 CPU
Results (Mean and medians)

After 800 simulations with 28 different setups and hardware platforms
Results (Manova)

Separate analysis of variance

\[ P < 0.01 \]

\[ F > 1 \]

Hypothesis \((H_0)\) WallTime

<table>
<thead>
<tr>
<th></th>
<th>GAUSS BM</th>
<th>GAUSS KVM</th>
<th>AIRES BM</th>
<th>AIRES KVM</th>
<th>FastJet BM</th>
<th>FastJet KVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snedecor F</td>
<td>157.9286</td>
<td>171.4093</td>
<td>2082.0955</td>
<td>1559.9788</td>
<td>1233.9208</td>
<td>2925.2098</td>
</tr>
<tr>
<td>Probability P</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Exact Signif. | Wilk’s Lambda | Prob. P | Snedecor F | Pillai’s Trace |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CV 1</td>
<td>5.143E-005</td>
<td>0</td>
<td>2.3368</td>
<td>2.5391</td>
</tr>
<tr>
<td>CV 2</td>
<td>0.08621</td>
<td>0</td>
<td>1.4718</td>
<td>1.5397</td>
</tr>
<tr>
<td>CV 3</td>
<td>0.4338</td>
<td>0.9324</td>
<td>0.7788</td>
<td>0.7384</td>
</tr>
<tr>
<td>CV 4</td>
<td>0.6143</td>
<td>0.9913</td>
<td>0.6235</td>
<td>0.4446</td>
</tr>
<tr>
<td>CV 5</td>
<td>0.7893</td>
<td>0.9985</td>
<td>0.4691</td>
<td>0.2229</td>
</tr>
<tr>
<td>CV 6</td>
<td>0.8979</td>
<td>0.9833</td>
<td>0.4443</td>
<td>0.1020</td>
</tr>
</tbody>
</table>

Wilk’s Lambda close to 0 (< 0.08)
Pillai’s Trace is higher than 1

=> 2.53 CV 1 and 1.53 CV 2
Results
Results (Correlation analysis)

Occurrences of single core with 2-4 GB similar areas

Centroids correlation in two cores setup => increase memory not improve performances

Four cores setup shows stability with large jobs increasing memory
Conclusion

● This work present a Multianalysis of Variance MANOVA about performance of Virtualization a Non-Virtualization on typical Cloud KVM platform
● Libvirt version should be adapted to the CPU host model for improve performance
● The memory is not a relevant factor for increasing the performance
● Large jobs have trending of to be more stables when you are increasing CPU number
● In multi-core machines, the use of all cores simultaneously with the same number of cores will increase the job performance
Thank you for your attention!