Advanced Simulations using Volunteer Computing:

JANE ARLETH DELA CRUZ

Supervisors:
Laurence Field, Ben Segal, Francois Grey
<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATLAS@home</td>
<td>Event Simulations</td>
</tr>
<tr>
<td>Beauty</td>
<td>Beauty Physics Simulations</td>
</tr>
<tr>
<td>CMS@Home</td>
<td>Event Simulations</td>
</tr>
<tr>
<td>SixTrack</td>
<td>Beam Simulations</td>
</tr>
<tr>
<td>Test4Theory</td>
<td>Monte Carlo Events</td>
</tr>
</tbody>
</table>
LHC@Home Project for “SDG”

- Help some non-CERN scientists use the same LHC@Home infrastructure to solve health-related problems such as epidemiology modelling.

- Run Machine Learning using Keras and Tensorflow Test on BOINC for the first time.
Machine Learning for SDG

- Combining satellite imagery and machine learning to predict poverty (Stanford University, USA, National Bureau of Economic Research, USA)

- Classifying Smoking Urges Via Machine Learning (University of Pittsburgh, USA)

- Machine Learning Based Big Data Processing Framework for Cancer Diagnosis Using Hidden Markov Model and GM Clustering
<table>
<thead>
<tr>
<th>Explored Solutions</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtualization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtualbox</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hosts are heterogeneous in terms of OS</td>
<td>Requires Hypervisor</td>
</tr>
<tr>
<td></td>
<td>Provide a unified software environment</td>
<td>GPU support untested</td>
</tr>
<tr>
<td>Containerization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singularity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Runs natively on SLC6/CentOS7 (Can run on GPU’s)</td>
<td>Only runs on Linux</td>
</tr>
</tbody>
</table>
Build Singularity Container
Python2.7, Keras 2.2, Tensorflow 1.70

Test Container with ML jobs
Training and Prediction

$ singularity run container.img learn.py
LD_LIBRARY_PATH: /usr/local/cuda/lib64:/singularity.d/lib
PATH: /usr/local/cuda/bin:/bin:/sbin:/usr/bin:/usr/sbin:/usr/local/bin:
Arguments received: learn.py

from .. import h5g, h5i, h5o, h5r, h5t, h5l, h5p
Using TensorFlow backend.

Preparing model, please wait...
('x_train shape:', (50000, 32, 32, 3))
(50000, 'train samples')
(10000, 'test samples')
Train on 50000 samples, validate on 10000 samples
Epoch 1/1
50000/50000 [==============================] - 210s 4ms/step - loss: 1.8072 - acc: 0.3364 - val_loss: 1.5128 - val_acc: 0.4546
model.h5
Make the BOINC Application

- Submit the application to BOINC-server via Condor Submission

![BOINC & High Throughput Computing](image)

- Run application on Volunteer Computer via BOINC-Manager
Conclusion

- LHC@Home has potential to run ML jobs.
- Potential Internal ML Computing Resource using Desktop Computers
- First time to run ML (Keras+Tensorflow) on VC.
- (For now,) ML jobs can be run on Linux only with Singularity.
Thank You!

Questions?

jane.delacruz@obf.ateneo.edu

Jane Arleth dela Cruz
References

LHC@Home: http://lhcathome.web.cern.ch/

BOINC: http://boinc.berkeley.edu

Singularity Python Methods, University of Arizona HPC: https://docs.hpc.arizona.edu/display/UAHPC/Singularity+++CentOS7%2C+Tensorflow1.4.1%2C+Keras%2C+Python3.5%2C+Cuda8.0%2C+cuDNN6


Backfilling the Grid with Containerized BOINC in the ATLAS computing: https://indico.cern.ch/event/587955/contributions/2937192/attachments/1683819/2706562/Backfilling_the_Grid_with_Containerized_BOINC_in_the_ATLAS_computing.pdf

Innovative Big Data Approaches for Capturing and Analyzing Data to Monitor and Achieve the SDGs: https://reliefweb.int/sites/reliefweb.int/files/resources/Innovative%20Big%20Data%20Approaches%20for%20Capturing%20and%20Analyzing%20Data%20to%20Monitor%20and%20Achieve%20the%20SDGs.pdf

Jane Arleth dela Cruz 9