Basic Primer on using Virtual Resource for Mastering HPC and Cloud Computing

Chris Bording
Senior IT Architect Research and User Support Environment
IBM Research

- 3,000+ (researchers and engineers)
- 5 Nobel Prizes
- 6 Turing Awards
- 10 National Medals of Technology
- 5 National Medals of Science
- Averaging 9.3 patents per day
- 22 years of patent leadership
- Collaboration with clients and universities
- Multi-disciplinary projects that lead to prototypes or long-term projects
TACC – Texas Advance Computing Center
HPC Resources

**FRONTERA** – 38.7 PETAFLOP
- 8008 nodes x 56 cores (2x28)

**STAMPEDE 2** – 18 PETAFLOP
- 4200 Intel KNL nodes x 68 cores
- 1736 nodes x 48 cores (2x24) Intel Skylake

**LONESTAR 5** – 1.3 PETAFLOP
- 1252 nodes x 24 cores (2x12) Haswell (CRAY XC40)
HPC Resources at University of Sheffield

The University of Sheffield HPC Facilities are freely available to all staff and research **Students!**

- **Bessemer cluster** - 26 nodes -> 1040 cores
- 1 node with 4 NVIDIA Tesla V100 GPUs

- **ShARC** – Sheffield Advanced Research Computer
  - 98 nodes -> 6272 cores Intel Haswell
  - 4 large memory nodes with 256 Gbytes ram
  - 2 nodes each with 8 x NVIDIA Tesla K80

- **SGE** – Son of Grid Engine is the scheduler.

**Sheffield HPC website**
Internship Training Programme

The initial training programme was:

- week-long "bootcamp".
- uses the Software Carpentry material
  https://software-carpentry.org/

1. The Unix Shell
2. Version Control with Git
3. Programming with Python
4. Internally develop course on using LSF/HPC.
The initial training programme was:
• week-long "bootcamp".
• uses the Software Carpentry material https://software-carpentry.org/
  1. The Unix Shell
  2. Version Control with Git
  3. Programming with Python
  4. Internally develop course on using (LSF, SLURM, SGE)
  5. Intro to HPC Architecture
The expectation is that should be sufficient training to move from your Laptop to…

or you can try lots of thing like...

THIS???
You can also try to do giant food challenges!

But should you?
What researchers need is support and resources that allow them to become experts with the computational applications in their scientific domains that they will use everyday during their research careers!

Secret of HPC ---- The resources typically only last 3-5 years then get binned once the warranty goes out. What do you think is the cause of this??
What service is TACC and other HPC centres now offering that will help researcher as they begin.

CLOUD

Okay, these are really Virtual Machines but they are self service and on demand!
At TACC the Cloud and Interactive Computing group provides a number of services for the national research science Community.

IaaS –
- Chameleon
- Jetstream
- Rodeo

PaaS –
- Agave Science-as-a-Service API
- TACC JupyterHub
- ABACO
HPC training is absolutely necessary!

This training approach has been well received by the interns and useful for their internship projects and they have better understanding of how to use compute resources in their PhD working.

1. Improving Collaboration
2. Automation of workflows
3. Reproducible work
4. Basics of HPC

The outcome of this training has in general been a success!

BUT

In the past two years we have only had one IBM intern that has had any prior experience with using the Cloud or VMs!!
In case you have not heard, the cloud computing market is worth over 250 billion dollars in 2019. Not 10 years from now but this year!!

So Virtual Machines are a thing.

As well as IBM and Redhat have all made Multi-Billion Dollar investments in the Cloud technologies and services.

P.S.

There is a skill shortage and they are all hiring!
IBM Research and the Hartree Centre are collaborating with UK industry through the Innovation Return on Research programme to create digital assets.

To achieve the goals of this collaboration a strong foundation of compute is needed. However, the interns need to become familiar with new technologies to allow them to become the innovators we want them to be.

What we identified during training is that we have had only one student that had any extensive experience with using Cloud/virtual resources. This is our real knowledge gap in training.

- Few if any of the interns have access to a research cloud at their home universities.
- Despite the promise of AWL and Azure, and a plethora of others providers including IBM cloud. Limited uptake by students.
Requirements for a platform which can be used for training on how to use Virtual Machines (VMs) or Infrastructure as a Service (IaaS).

1. Limited training time – half day
2. Focused training on automate provisioning VMs.
3. EASE OF USE!! (remember some of the students might have just learned BASH two days earlier)

Solution – Combination of Virtualbox and Vagrant
What is VirtualBox and What is Vagrant?

VirtualBox
- is a virtualization tool
- It can run multiple instances of VirtualBox simultaneously
- Will allow you to create a small network of VMs or a virtual cluster

Vagrant
- Is a tool for building and managing virtual machine environments in a single workflow.
- Provides an easy-to-use workflow.
- Is made with automation in-mind!
- Shortens the environment setup time.
- If done right allows for a reproducible environment that is portable and easy to configure. (ie using git!)
How I integrated this into training.

- Abstract the initial use of VirtualBox and Vagrant as much as possible.
- The interns installed VirtualBox and Vagrant on their laptops.
- Created a Vagrantfile that used an Ubuntu-18.04 image.
- Used wget to copy the software carpentry setup material "data-shell.zip" for the "The Unix Shell" course. This was extracted into the default Vagrant HOME directory on the VM.
- Put everything in git!
Vagrant Basics

Basic Vagrant commands

- **vagrant init** initializes a new Vagrant environment by creating a Vagrantfile.

- **vagrant up** starts and provisions the vagrant environment

- **vagrant ssh** connects to the machine via SSH

- **vagrant halt** stops the vagrant machine

- **vagrant destroy** stops and deletes all traces of the vagrant machine

The most important command is

- **vagrant help**
VirtualBox and Vagrant training

Prior to the training session on using Virtualbox and Vagrant the interns had to use:

• vagrant up
• vagrant ssh
• vagrant halt

The synopsis of the vagrant training was:

1. Quickly reviewed the vagrant basic commands.
2. Reviewed the Vagrantfile file used for the Software Carpentry training all 14 lines
# provision the instance with the swc data!
config.vm.provision "shell", inline: <<-SHELL
  apt-get update
  apt-get install -y zip
  wget http://swcarpentry.github.io/shell-novice/data/data-shell.zip
  unzip data-shell.zip
  rm data-shell.zip
  chown -R vagrant:vagrant data-shell
SHELL

# define the hardware resources in the vm
config.vm.provider "virtualbox" do |vb|
  vb.memory = "1024"
end

# sync folders on host and vm
config.vm.synced_folder "../scripts", "/home/vagrant/provision"

# -*- mode: ruby -*-
# vi: set ft=ruby :
Vagrant.configure("2") do |config|
  config.vm.box = "bento/ubuntu-18.04"
end

Vagrantfile- 14 lines
Select the bento/ubuntu-18.04 image

# sync folders on host and vm
config.vm.synced_folder "../scripts", "/home/vagrant/provision"

# define the hardware resources in the vm
config.vm.provider "virtualbox" do |vb|
  vb.memory = "1024"
end

# -*- mode: ruby -*-
# vi: set ft=ruby :
Vagrant.configure("2") do |config|
  config.vm.box = "bento/ubuntu-18.04"
If you need more than zip??
Then try RPMs – The ubuntu repository!
https://packages.ubuntu.com/bionic/

<table>
<thead>
<tr>
<th>Math</th>
<th>Database</th>
<th>Electronics</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Fenics</td>
<td>• Mariadb</td>
<td>• Alliance</td>
<td>• Abinit</td>
</tr>
<tr>
<td>• BLAS</td>
<td>• MySQL</td>
<td>• Electric</td>
<td>• Astromatic</td>
</tr>
<tr>
<td>• Scalapack</td>
<td>• PostgreSQL</td>
<td>• TclSpice</td>
<td>• Bamtools</td>
</tr>
<tr>
<td>• FFTW</td>
<td>• Redis</td>
<td>• qflow</td>
<td>• Bioperl</td>
</tr>
</tbody>
</table>
Visualization Demo with Paraview and ViSIT

1. Install Paraview and ViSIT on my laptop.

2. Using the initial Vagrantfile it was modified (added 2 lines) to create a new VirtualBox image and installing zlib, szip, cmake, HDF5 and Silo data format libraries from source.

3. From the vagrant instance I was able to created a set of silo data files for ViSIT and a h5 data file for Paraview.

4. Rendered the images on the host machine!
THINK - DEVOPS

The Researchers need to start to think in terms of DEVOPS!

- How do I provision my VM with the correct environment?
- What is the workflow?
- How do I automate my workflow?
- How do I present my results?

- Minimal software requirements (THINK containers)
- Performance! What are the hardware resource requirements ie cores, gpus, or memory.
- Who is running this and from where? Website, Ipad, Iphone or command line?
The initial programme was:
• week-long "bootcamp".
• uses the Software Carpentry material https://software-carpentry.org/
  1. The Unix Shell
  2. Version Control with Git
  3. Programming with Python
  4. An internally developed course on using LSF/HPC.

Training Goals
1. Want to insure the success of all the internship projects.
2. Fill the knowledge gaps in basic skills.
Current Intern Projects

Multi-phase CFD using Volume of Fluid methods with OpenFOAM.
Jobs range from ~ 512 cores to 2000 cores

MD Simulations with LAMMPS
Average job size is ~512 cores

DPD Simulations using \ DL-MESO
Average job size is 512 cores

Energy Aware Computing a Hollistic Approach
Collects Tbytes of data from sensors on Cluster using IOT and containers.

Multi Fidelity Bayesian Optimization
Measuring algorithm performance
From a single node with 1 GPU to N nodes with 4N GPUs.
IBM Research Internships at the Hartree Centre

- PhD students.
- Both 3 and 6 month internships opportunities.
- 3 cycles per calendar year – Spring Summer and Autumn.
- 4-6 Interns per cycle.
- Work in one of the four IBM Research groups: Chemistry, Life Sciences, Engineering or Enabling Technologies
IBM Research @ Hartree Centre

Projects & team

www.research.ibm.com/labs/uk

Location

IBM Research
The Hartree Centre STFC Laboratory
Sci-Tech Daresbury
Warrington WA4 4AD
UK

Contact

Director: martyn.spink@uk.ibm.com
Business enquiries: deborah_sahota@uk.ibm.com