Docker, containers & creating scalable, reproducible, workflows in the cloud

Ian Muñoz - Computational Scientist
Docker, containers & creating scalable, reproducible, workflows in the cloud

THIS PRESENTATION WILL

BLOW YOUR MIND...

Ian Muñoz - Computational Scientist
Who is this guy?
- Docker
- AWS
- Demos
- Thoughts & Questions
Docker: A container platform
What is a container?

A “container” is just a term people use to describe a combination of Linux namespaces and cgroups. Linux namespaces and cgroups ARE first class objects. NOT containers.

Infrastructure as code (IaC):

is the process of managing and provisioning computer data centers through machine-readable definition files, rather than physical hardware configuration or interactive configuration tools.

Infrastructure as Code
MANAGING SERVERS IN THE CLOUD
Kief Morris
Software can be chaotic, but we make it work

Expert

Trying Stuff Until it Works

ORLY?

The Practical Developer
@ThePracticalDev
Dockerfile

FROM
RUN
CMD
LABEL
MAINTAINER
EXPOSE
ENV
ADD
COPY
ENTRYPOINT
VOLUME
USER
WORKDIR
ARG
ONBUILD
STOPSIGNAL
HEALTHCHECK
SHELL
Docker Swarm (mode)
Benefits:

- Cluster management
- Decentralized design
- Declarative service model
- Scaling
- Desired state reconciliation
- Multi-host networking
- Service discovery
- Load balancing
- Secure by default
- Rolling updates
Definition: *Cloud Computing*

NIST: “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

● ~70% of IT professionals are using the cloud in some shape or form.
● 1.5 clouds on average
● Professionals are using between 1 and 6 clouds
● There are many clouds
● The cloud ultimately provides:
  ○ infrastructure (IaaS)
  ○ platforms (Paas)
  ○ software (Saas)

What is AWS (Amazon Web Services)?
What is AWS (Amazon Web Services)?

Source: The Next Platform
What is AWS (Amazon Web Services)?
What is AWS (Amazon Web Services)?
Storage & Content Delivery

**Amazon Simple Storage Service**
- Amazon S3
- bucket
- bucket with objects
- object

**AWS Import/Export**
- AWS Import/Export

**Amazon Elastic Block Store**
- Amazon EBS
- volume
- snapshot

**AWS Storage Gateway**
- AWS Storage Gateway
  - non-cached volume
  - cached volume
  - virtual tape library

**AWS Import/Export**
- AWS Import/Export
Amazon Elastic Compute Cloud

Amazon EC2 instance instances AMI DB on instance with CloudWatch Elastic IP optimized instance Amazon Lambda

Amazon Virtual Private Cloud

Amazon VPC router Internet gateway customer gateway virtual private gateway VPN connection VPC peering
Price Trend of Each Product in the Matched Model Index

- m1.small
- m1.large
- m1.xlarge
- m2.2xlarge
- m2.4xlarge
- c1.medium
- c1.xlarge

Time Period:
- 200912
- 201006
- 201012
- 201006
- 201112
- 201206
- 201212
- 201306

Price ($/hour):
- 0
- 0.5
- 1
- 1.5
- 2
- 2.5
- 3

http://repository.wellesley.edu/cgi/viewcontent.cgi?article=1472&context=thesiscollection
<table>
<thead>
<tr>
<th>Region</th>
<th>Instance Type</th>
<th>Platform</th>
<th>Price start date</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>sa-east-1</td>
<td>m1.large</td>
<td>Linux/Unix</td>
<td>Dec. 15 2011</td>
<td>$0.46</td>
</tr>
<tr>
<td>sa-east-1</td>
<td>m1.large</td>
<td>Linux/Unix</td>
<td>Mar. 8 2012</td>
<td>$0.46</td>
</tr>
<tr>
<td>sa-east-1</td>
<td>m1.large</td>
<td>Linux/Unix</td>
<td>Jan. 11 2013</td>
<td>$0.32</td>
</tr>
<tr>
<td>sa-east-1</td>
<td>m1.large</td>
<td>Linux/Unix</td>
<td>Jan. 4 2014</td>
<td>$0.23</td>
</tr>
</tbody>
</table>

https://www.cloudyn.com/blog/analyzing-aws-ec2-price-drops-over-the-past-5-years/
Batch Processing

Where does your data live?

How does it get To the cloud?

Never underestimate the bandwidth of a station wagon full of tapes hurtling down the highway.

—Andrew Tanenbaum, 1981
Simple Queue Service

Elastic Container Service
Potential Issues with the cloud...
What’s next?
What's next?

- GPU Cluster
- Amazon EFS
- Better State (Dynamo)
- More automated configuration
- Better Spot configuration
Things I am excited about

More dynamic GPUs

Abstracting batch processing
Thank You!

ian@cgrb.oregonstate.edu