

Google Groups

- Targeting Deploying Hadoop - Make this a priority
 - How are we going to get the data from storage into OpenStack
 - Where does the data live permanently
 - Where do we need to put that data to get it into Hadoop
 - Need to map out how data is going to flow through the system, do we need some hdfs nodes that will store the data for Hadoop.?
 - Want to meet with Robert & Dr. Denton about how they want this storage to work. May want to include Dr. Ludwig and find out exactly what she does to use Hadoop and how it flows from start to finish.
 - Do they pull in entire data set or just chunks?
 - Confusion on how data gets mapped across nodes. How do they load their processing jobs?
 - How is the job (map/reduce) written and how is it entered into Hadoop?
 - How to define a data set?
 - Create some sort of collateral document on how the AWS scripts work. What does it do?
 - Lets make OpenStack really work for Hadoop, make it optimized for that kind of use.
 - What are the different roles that servers play in Hadoop processing?
 - Job tracker and name node?
 - What kind of hardware should different roles use for Hadoop?
 - Job tracker has more memory?
 - name nodes have more cpu and storage?
 - Is this built into the AWS script? This why we need to know what the script does in the background.
 - Check with Dr. Ludwig to see if she has an architecture diagram of how Hadoop works when she's using it.

Quantum

- Can be really nice
- Nova network is probably a better choice
- Other ways than VLANs
- Lets get a dedicated switch
 - Need to setup a dual-homed box

Draw up a diagram of both the production and development instances

- Include storage (local, SAN, iscsi)
- Networking configurations

Grizzly

- LDAP should be a lower priority
- Make sure machines have virtualization extensions on hardware
- By default, most services are blocked by the firewall
 - netstat -tnap
- Ubuntu packages are probably out of date

For next week

- Get development instance running using Devstack. It doesn't matter if it's a single node or multi-node, we just need something that OpenStack will run on.
- Describe the workflow for Hadoop
 - How are the APIs actually going to work with Hadoop?
- What would a researcher see as an interface to Hadoop on OpenStack
- Document that describes how AWS script works
- Documentation on how Dr. Ludwig works with Hadoop
- Documentation on how Dr. Denton would like to work with Hadoop