



Cross-Institute Virtual Cluster Management in PRAGMA

Yuan Luo¹, Shava Smallen², Nadya Williams², Beth Plale¹, Philip Papadopoulos²

¹School of Informatics and Computing, Indiana University Bloomington

²San Diego Supercomputer Center, University of California San Diego



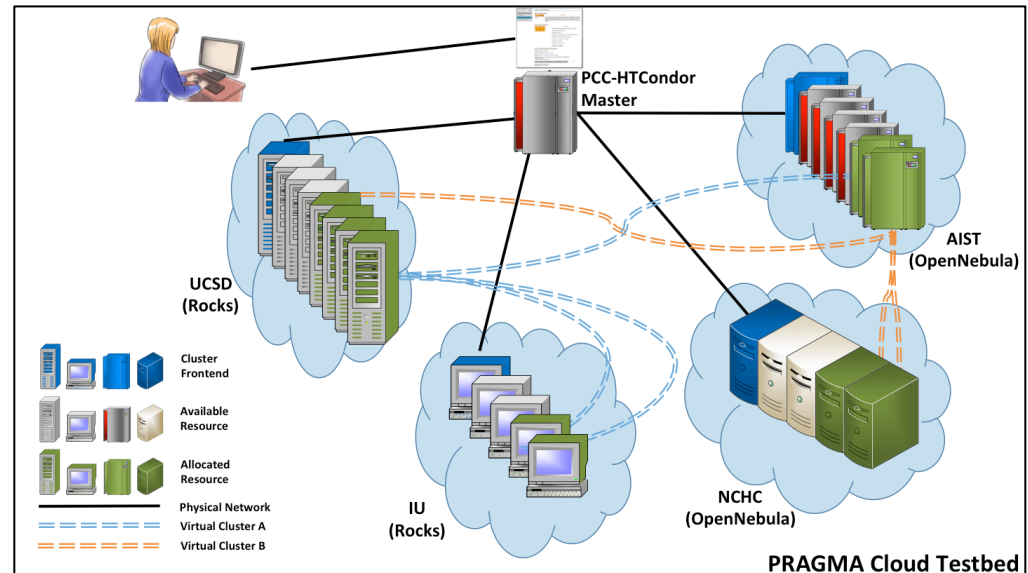
DATA TO INSIGHT CENTER

INDIANA UNIVERSITY
Pervasive Technology Institute

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

Goals of Personal Cloud Controller (PCC)

- Enable lab/group to easily manage **application virtual clusters** on available resources
- Leverage PRAGMA Cloud tools: PRAGMA Bootstrap, IPOP, ViNE, Rocks.
- Lightweight, extends HTCondor from U Wisc.
- Provide command-line and Web interfaces

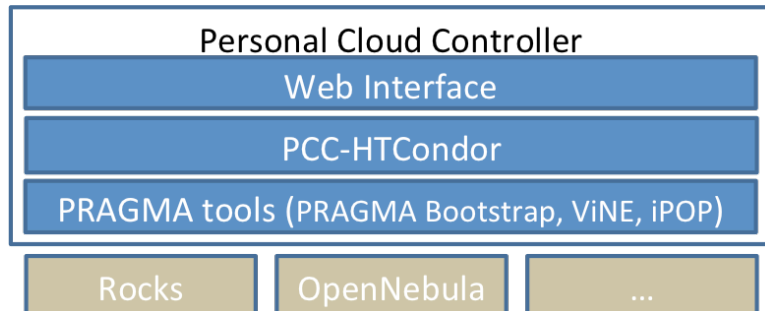


pragma_boot

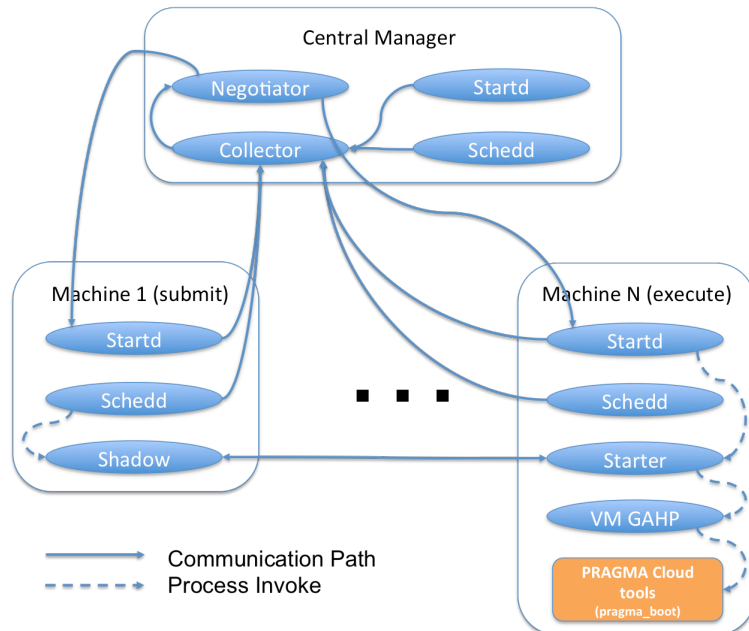
Working Group: Resources

Previous work

Demoed single site virtual cluster prototype at PRAGMA 26



High level architecture diagram of PCC



Architecture diagram of PCC-HTCondor

Web interface to launch and view status of virtual cluster. The interface is titled 'PRAGMA Personal Cloud Controller' and includes a navigation menu with 'Introduction', 'Launch a Virtual Cluster', and 'View Virtual Clusters'. The main content area shows the 'Launch a Virtual Cluster' process in three steps:

- Step 1: Select an Image**: The 'lifemapper' image is selected from a list that also includes 'dock6'. A description for 'lifemapper' is provided: 'The Lifemapper Project (www.lifemapper.org) is a computational and data resource for biogeographic research and education on ecological models of species distribution. Lifemapper's architecture is composed of back end computational modeling linked through web services to front end research clients.'
- Step 2: Select a Resource**: The 'nbc-224' resource is selected. Details for this resource are shown: Name: PRAGMA Virtual Cluster Manager Test Cluster, URL: http://www.sdsc.edu/, Organization: SDSC, Location: San Diego, California, US (N32.87 W117.22), Capacity: 4 Virtual Clusters, 12 core(s), Load: 0 Virtual Clusters, 0 core(s), Available: 12 core(s). A 'Select # of cores' dropdown is set to 8, and an 'Add to virtual cluster' button is present.
- Step 3: Submit Virtual Cluster Job Request**: The 'Image selected' is 'lifemapper' and the 'Resource selected' is 'nbc-224, 8 cores'. The 'Submit time' is 'Tue, 01 Apr 2014 19:01:19 -0700'. A message states: 'Created submit directory /var/log/pcc/submit/job/20140401.1396404079/ Submitting job(s). 1 job(s) submitted to cluster 71.' A progress bar shows 72% completion. The status at the bottom is 'Progress: Booting 'compute-1'...' and 'Elapsed time: 44.12 minutes'.

Web interface to launch and view status of virtual cluster

Progress for PRAGMA 27

Cross-institute virtual cluster using IPOP

Accomplished

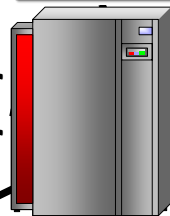
- ✓ Setup Rocks with KVM roll on 3-node cluster at IU
- ✓ Experimented with IPOP and measured initialization and bandwidth performance between IU and SDSC
- ✓ Drafted a paper “A Personal Cloud Controller Framework” for submission
- ✓ Developed new IPOP Rocks roll for easy installation of IPOP to any Rocks virtual cluster
- ✓ Added automated IPOP server/client initialization to PRAGMA Bootstrap
- ✓ Enabled multi-site virtual cluster creation via enhanced PCC-HTCondor (VM GAHP) and Condor DAG capabilities
- ✓ Part-way thru setup of OpenNebula/PRAGMA Bootstrap on 4-node cluster at AIST

TODO

- ☐ Automated reconfiguration of Rocks DB
- ☐ Debug OneNebula/PRAGMA Bootstrap issues
- ☐ Integrate changes into Web interface
- ☐ Rocks rolls for PCC-HTCondor and enhanced PRAGMA Bootstrap
- ☐ Live application demo with Lifemapper



PCC-HTCondor Master



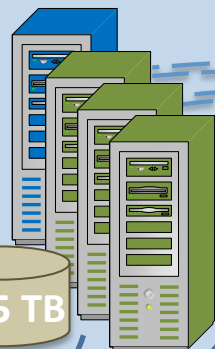
nbc-224.ucsd.edu

- (4) Dell PowerEdge SC1435
- (2) Dual-Core 2.4 GHz AMD Opteron
 - 8 GB Memory



v6.1

1.25 TB

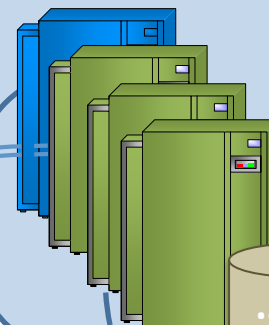


163.220.56.77 (AIST)

- (4) Dell PowerEdge M610
- (2) Quad-Core 2.4 GHz Intel Xeon
 - 24 GB Memory



.5 TB



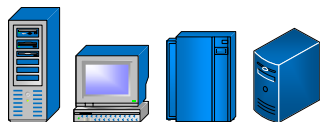
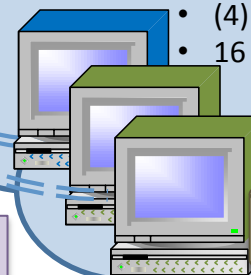
pragma8.cs.indiana.edu

- (3) Dell PowerEdge 6950
- (4) Dual-Core 1 GHz AMD Opteron
 - 16 GB Memory

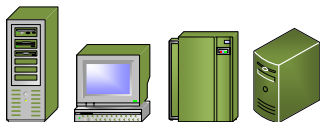


v6.1.1

1 TB



Cluster
Frontend



Allocated
Resource

Physical Network

Virtual Cluster
nbc-227.ucsd.edu

PCC Software



v8.0.6

PRAGMA Bootstrap

IPOP

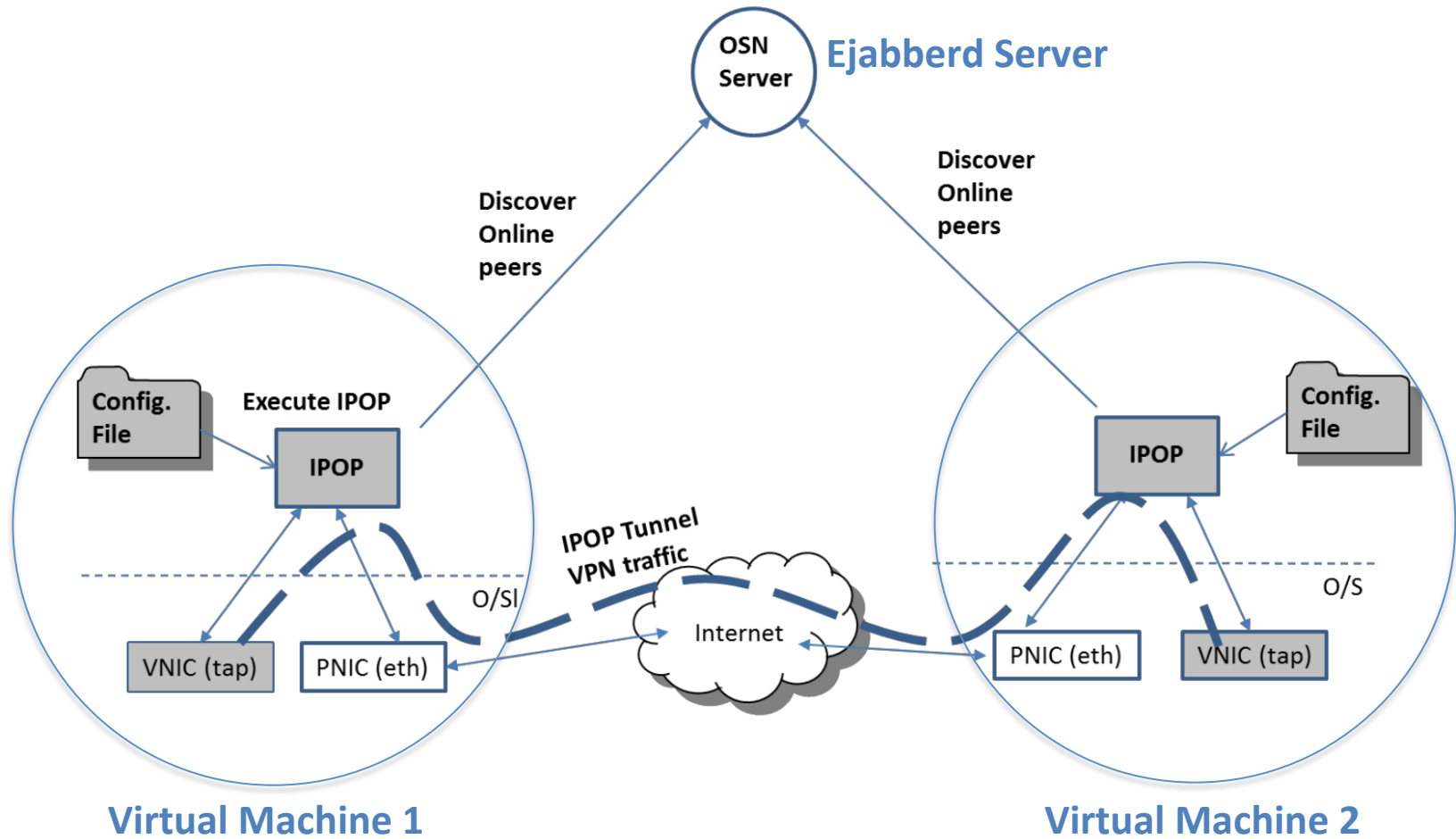


Image from IPOP White Paper, <http://ipop-project.org/wp-content/uploads/2014/07/IPPOP-WhitePaper-1407.pdf>

PCC Evaluation

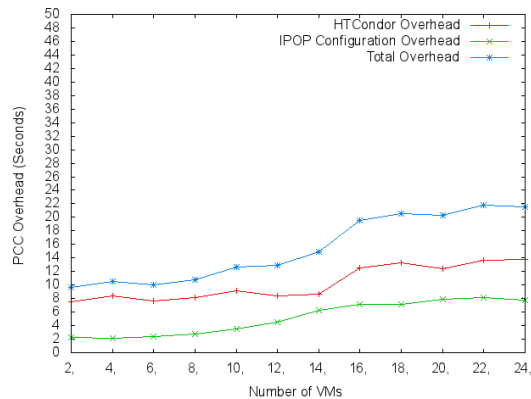
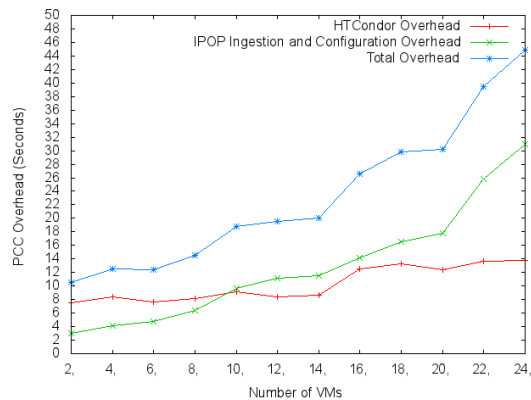
- ❑ We measure overhead of PCC as captured by overhead during the resource provisioning phase and overhead of application running over VPN.
- ❑ Testbed. Two clusters were selected: one at Indiana University(IU) and the other at the San Diego Supercomputer Center (SDSC).

Table 1. Testbed Specifications

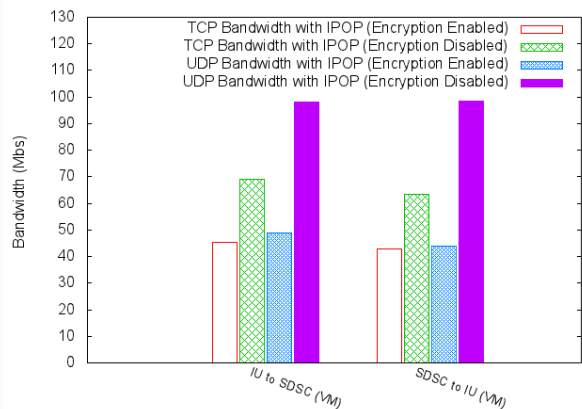
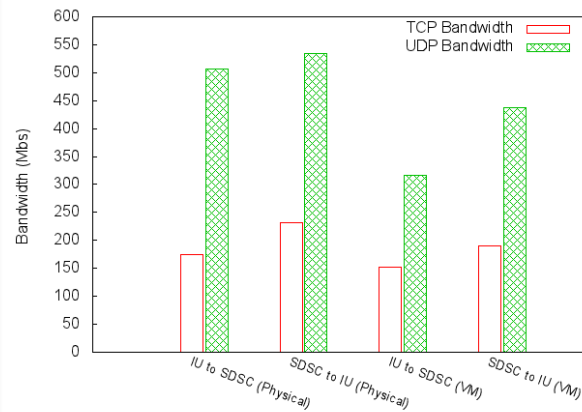
Cluster	Nodes	CPU	Cores	Mem	Ethernet	OS	VMM	Cloud Platform
SDSC	4	2.4GHZ	4	8GB	1000Base-T	CentOS 6	KVM	Rocks 6.1
IU	3	2.4GHZ	8	16GB	1000Base-T	CentOS 6	KVM	Rocks 6.1

PCC Evaluation – *cont'd*

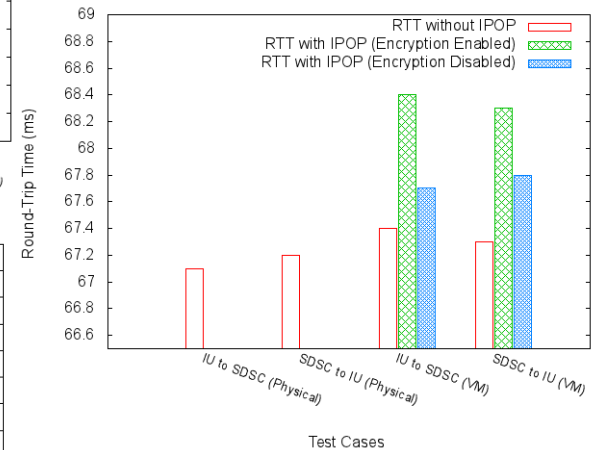
PCC Overhead Evaluation



Network Overhead Evaluation



Test Cases



Test Cases

IPOP Rocks Roll

 <https://github.com/pragmagrid/ipop>



v17.0



v14.07



v14.07

File	IPOP Server	IPOP Client	Purpose
/opt/ipop/ejabberd/bin/initEjabberd	X		(Re-)initializes Ejabberd for IPOP usage and produces ipopserver.info file.
/etc/init.d/ejabberd	X		Optionally initializes (via initEjabberd) and then starts Ejabberd
/var/www/html/ipop/ip.php	X		Distributes unique IP addresses to IPOP clients
/opt/ipop/bin/updateConfJson	X	X	Populates IPOP config.json file with ipopserver.info contents
/etc/init.d/ipop	X	X	Optionally initializes then starts IPOP

Many thanks to Nadya Williams!!

IPOP-enabled PRAGMA Bootstrap

PRAGMA Bootstrap

- Instantiates dynip-enabled virtual clusters within a single cluster
 - Utilizes “drivers” to support multiple cloud platforms (current support for Rocks and OpenNebula)
 - Allocates IP addresses, installs vc-out.xml (for dynip), and boots VMs

IPOP Enhancements

- **--enable-ipop-server=<URL>**
Starts up IPOP-enabled virtual cluster with the frontend serving as the IPOP server; fetches IPOP server info once initialization is complete
- **--enable-ipop-client=<URL>**
Start up the IPOP-enabled virtual cluster as an IPOP client (to another virtual cluster).

Enhanced PCC-HTCondor

(Leveraged HTCondor DAG capabilities to create multi-site virtual cluster)

```
# File name: vc-ipop.dag
#
JOB A vc1.sub
JOB B vc2.sub
JOB C vc3.sub
```

HTCondor DAG

```
universe           = vm
executable          = rocks_vc_1
requirements        = Machine == "nbcrc-224.ucsd.edu"
log                 = vc1.log.txt
vm_type             = rocks
vm_memory           = 64
rocks_job_dir       = /tmp/dag_vm
RequestMemory       = 64
rocks_should_transfer_files = Yes
RunAsOwner=True
queue
```

ROCKS

```
universe           = vm
executable          = rocks_vc_2
requirements        = Machine == "pragma8.cs.indiana.edu"
log                 = vc2.log.txt
vm_type             = rocks
vm_memory           = 64
rocks_job_dir       = /tmp/dag_vm
RequestMemory       = 64
rocks_should_transfer_files = Yes
RunAsOwner=True
queue
```

ROCKS

```
universe           = vm
executable          = rocks_vc_3
requirements        = Machine == "163.220.56.77"
log                 = vc3.log.txt
vm_type             = rocks
vm_memory           = 64
rocks_job_dir       = /tmp/dag_vm
RequestMemory       = 64
rocks_should_transfer_files = Yes
RunAsOwner=True
queue
```

OpenNebula.org

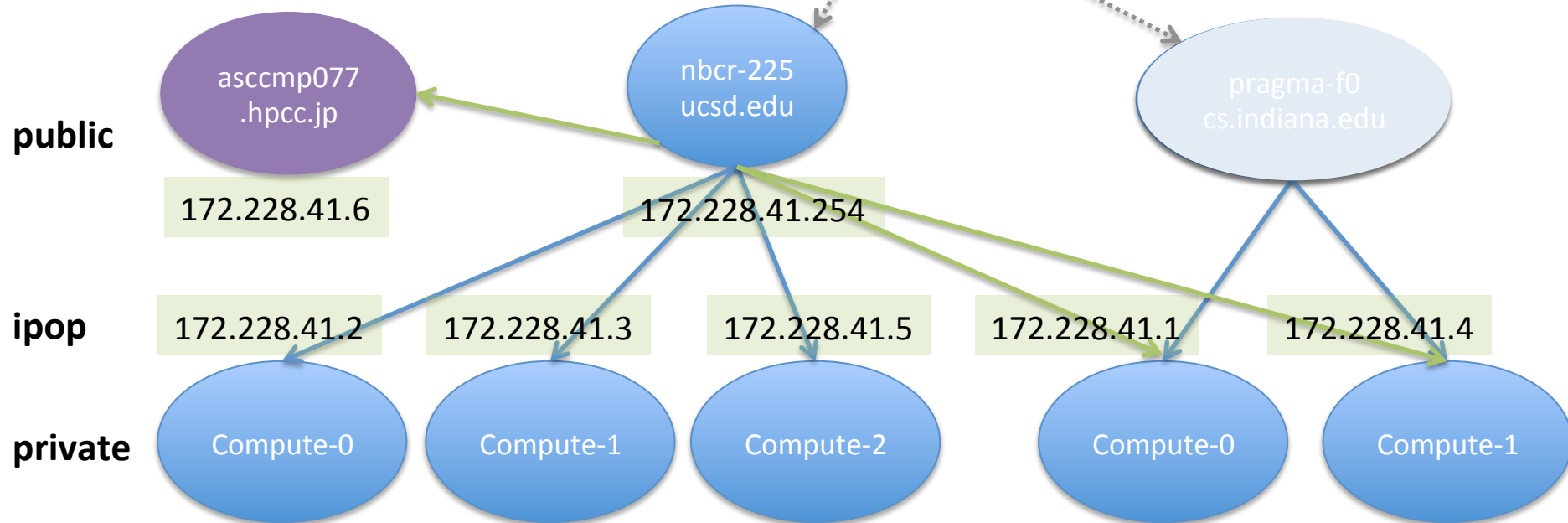
```
--executable      = pragma_boot
--basepath         = /opt/pragma_boot/vm-images
--key              = ~/.ssh/id_rsa.pub
--num_cores        = 2
--vcname           = lifemapper
--logfile          = shava_pragma_boot.log
--ipop-enable-server=${COLLECTOR_HOST_STRING}/ipop/register.php?jobid=${DAGManJobId}
--ipop-enable-client=${COLLECTOR_HOST_STRING}/ipop/register.php?jobid=${DAGManJobId}
```

VC configuration file

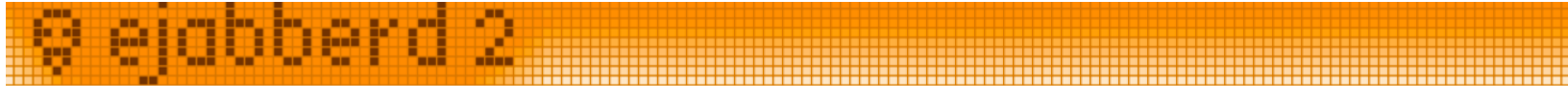
Parallel pragma_boot jobs

Instantiated virtual cluster

```
# File name: vc-ipop.dag
#
JOB  A  vc1.sub
JOB  B  vc2.sub
JOB  C  vc3.sub
```



Demo: View Ejabberd interface



ejabberd
Access Control Lists
Access Rules
Virtual Hosts
ejabberd
Access Control Lists
Access Rules
Users
Online Users
Last Activity
Nodes
Statistics
Shared Roster Groups
Nodes
Statistics

User **ipopuser@ejabberd**

Connected Resources:

- tincan1de34723728e58571ce3501c3d21a13f46a7e927 (tls://169.228.41.225:50745#ejabberd@nbcr-225.ucsd.edu)
- tincan9169df46acec4238922902480f0ecd393d023076 (tls://129.79.240.60:60407#ejabberd@nbcr-225.ucsd.edu)
- tincanfd5fbdde01244d74e7f54a1d72ff0441325d9fdb (tls://10.1.1.254:39350#ejabberd@nbcr-225.ucsd.edu)

Password:

Change Password

Last Activity

Online

Offline Messages:

0 Remove All Offline Messages

Roster

Remove User

```
[root@nbc-225 ~]# sh show-ipop

===== 172.228.41.1 =====
ping -c 1 172.228.41.1
PING 172.228.41.1 (172.228.41.1) 56(84) bytes of data.
64 bytes from 172.228.41.1: icmp_seq=1 ttl=64 time=69.9 ms

--- 172.228.41.1 ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 70ms
rtt min/avg/max/mdev = 69.905/69.905/69.905/0.000 ms
ssh 172.228.41.1 hostname
compute-0

===== 172.228.41.2 =====
ping -c 1 172.228.41.2
PING 172.228.41.2 (172.228.41.2) 56(84) bytes of data.
64 bytes from 172.228.41.2: icmp_seq=1 ttl=64 time=0.977 ms

--- 172.228.41.2 ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 1ms
rtt min/avg/max/mdev = 0.977/0.977/0.977/0.000 ms
ssh 172.228.41.2 hostname
compute-0

===== 172.228.41.3 =====
ping -c 1 172.228.41.3
PING 172.228.41.3 (172.228.41.3) 56(84) bytes of data.
64 bytes from 172.228.41.3: icmp_seq=1 ttl=64 time=0.872 ms

--- 172.228.41.3 ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 1ms
rtt min/avg/max/mdev = 0.872/0.872/0.872/0.000 ms
ssh 172.228.41.3 hostname
compute-1

===== 172.228.41.4 =====
ping -c 1 172.228.41.4
PING 172.228.41.4 (172.228.41.4) 56(84) bytes of data.
64 bytes from 172.228.41.4: icmp_seq=1 ttl=64 time=68.8 ms

--- 172.228.41.4 ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 69ms
rtt min/avg/max/mdev = 68.816/68.816/68.816/0.000 ms
ssh 172.228.41.4 hostname
compute-1

===== 172.228.41.5 =====
ping -c 1 172.228.41.5
PING 172.228.41.5 (172.228.41.5) 56(84) bytes of data.
64 bytes from 172.228.41.5: icmp_seq=1 ttl=64 time=0.892 ms

--- 172.228.41.5 ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 1ms
rtt min/avg/max/mdev = 0.892/0.892/0.892/0.000 ms
ssh 172.228.41.5 hostname
compute-2

===== 172.228.41.6 =====
ping -c 1 172.228.41.6
PING 172.228.41.6 (172.228.41.6) 56(84) bytes of data.
64 bytes from 172.228.41.6: icmp_seq=1 ttl=64 time=117 ms

--- 172.228.41.6 ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 117ms
rtt min/avg/max/mdev = 117.393/117.393/117.393/0.000 ms
ssh ssmallen@172.228.41.6 hostname
asccmp077.hpcc.jp
```

Future Work

Near-term goals

- **Oct-Dec**
 - Automated reconfiguration of Rocks DB
 - Integrate changes into Web interface
 - Rocks rolls for PCC-HTCondor and enhanced PRAGMA Bootstrap
 - Live application demo with Lifemapper
- **Jan – April**
 - Work with Aimee to develop load model for LM
 - Develop PCC auto-sizing capabilities



Longer-term goals

- Improve resource allocation algorithms.
- Enable resource to application information sharing.
- Extend the Hierarchical MapReduce model to support distributed sensitive data processing.
- Schedule application jobs based on VC topologies, and VM provenance.