vcluster:

A Framework for Auto Scalable Virtual Cluster System in Heterogeneous Clouds

March 2, 2012

Supercomputing Center
KISTI

Seo-Young Noh
Contents

1. Introduction

2. Related Work

3. vcluster: Automatic Scalable Virtual Cluster System

4. Development Status

5. Roadmap

6. Conclusions
1. Introduction

- Why we started this project...
  - GSDC at KISTI has been recently setup and is growing fast.
  - There is a strong need to tight collaboration with labs having strong experiences on grid, data, and cluster systems.
  - This project has been started as a part of collaboration with Fermilab.

- Why cloud was chosen...
  - Among 7 possible collaboration items, FermiCloud related project was attractive to both parties.
  - Cloud computing is one of hottest topics in IT world and no exception in HEP because of its flexibility.
  - Developed while KISTI staff stayed at Fermilab for 3 months (from May to August 2011)
1. Introduction

- Grid vs. Cloud
  - Grid and Cloud are both types of HTC (High Throughput Computing).
  - So, they are both focusing on system utilization.
  - But, there is a big difference between them.
  - Grid focuses on orchestration of scattered physical computing resources while cloud focuses on utilization of idle physical computing cycles through virtualization.

- Benefits from cloud...
  - Flexible for computing resource allocation.
  - Able to create multiple virtual machines on a single physical machine.
  - Able to make homogeneous and heterogeneous machines in OS and system level.
  - Easy to deploy huge computing cycles in a very short time.
  - Fast disaster recovery for service.
  - ...

Seo-Young Noh
2. Related Work

- Very similar articles...
  2. I. Gable et. al, “A batch system for hep applications on a distributed iaas cloud” in proceedings of computing in High Energy Physics 2010 (Taiwan)

- Concept of cloud scheduler

vcluster is very similar, but there are differences.

Figures from [1]
3. vcluster: Automatic Scalable Virtual Cluster System

- Design Concept
  - Cloud System Agnostic
  - Batch System Agnostic
  - Simple Cluster System
  - Auto Scalable on Demand
  - Dynamic Load Balancing
  - Power Management by Auto Migration
  - Security Monitoring
3. vcluster: Automatic Scalable Virtual Cluster System

- **Controller**
  - **Host Manager**
    - Host system management
    - VM migration & power on/off host systems
  - **Cloud Manager**
    - Single view for underlying cloud systems
  - **VM Manager**
    - Launching and managing vms.
  - **Plug-In Manager**
    - Handling plug-ins for batch and cloud systems
  - **Batch System Manager**
    - Single view for underlying batch systems
  - **Policy Generator**
    - Policy generation for jobs, running or idle vms.
    - Policy is passed to Load Balancer

- **Load Balancer**
  - **Policy**
    - Containing information how to distribute loads
  - **Plan Generator**
    - Creating an action plan for the policy
  - **Plan Executor**
    - Executing the plan generated by the plan generator

- **Monitoring**
  - **Cloud System Status Checker**
    - Monitoring the heartbeat of cloud systems
  - **Cluster System Status Checker**
    - Monitoring the heartbeat of virtual cluster system
    - Queue status & Pool status
  - **Security Checker**
    - Monitoring any security holes
4. Development Status

Some components have been partially implemented.
4. Development Status

```plaintext
vcluster > debug on
starting virtual machine manager
```
4. Development Status

vcluster > oneml list rsyoung
[1m   ID   USER  NAME STAT CPU   MEM    HOSTNAME TIME
1801 rsyoung clean-la runn 0 1024M fcl014 19 21:37:09
1825 rsyoung vm-wn runn 0 1024M fcl016 11 20:23:09

vcluster >
vcluster > cloudman set public 0
vcluster >
vcluster > dins
----------------------------------------
Inst ID  Status
----------------------------------------
i-1801  running
i-1825  running

vcluster >
vcluster > condor_status
Name OpSys Arch State Activity LoadAv Mem
ActvtyTime
Fermicloud007.fnal LINUX X86_64 Unclaimed Idle 0.010  1001
Fermicloud043.fnal LINUX X86_64 Unclaimed Idle 0.000  1001
Total Owner Claimed Unclaimed Matched Preempting Backfill
X86_64/LINUX 2 0 0 2 0 0 0 0

vcluster >
vcluster > cloudman set public 1
vcluster >
vcluster > dins
----------------------------------------
Inst ID  Status
----------------------------------------
i-b5ca1bdb stopped
i-3fed3d5a pending

vcluster >
vcluster > dins
Inst ID  Status
----------------------------------------
i-b5ca1bdb stopped
i-3fed3d5a running

vcluster >
vcluster > condor_status
Name OpSys Arch State Activity LoadAv Mem
ActvtyTime
Fermicloud007.fnal LINUX X86_64 Unclaimed Idle 0.010  1001
Fermicloud043.fnal LINUX X86_64 Unclaimed Idle 0.000  1001
ip-10-110-54-223.e LINUX X86_64 Unclaimed Idle 0.320  590
Total Owner Claimed Unclaimed Matched Preempting Backfill
X86_64/LINUX 3 0 0 3 0 0 0 0

vcluster >

vcluster > start i-3fed3d5a
vcluster > [MSG] : starting vm in a separated thread
vcluster > [MSG] : start a vm of i-3fed3d5a using REST API.
vcluster > [MSG] : starting is done in a separated thread

vcluster >
vcluster > dins
----------------------------------------
Inst ID  Status
----------------------------------------
i-b5ca1bdb stopped
i-3fed3d5a pending

vcluster >
vcluster > dins
----------------------------------------
Inst ID  Status
----------------------------------------
i-b5ca1bdb stopped
i-3fed3d5a running

vcluster >

vcluster > condor_status
Name OpSys Arch State Activity LoadAv Mem
ActvtyTime
Fermicloud007.fnal LINUX X86_64 Unclaimed Idle 0.010  1001
Fermicloud043.fnal LINUX X86_64 Unclaimed Idle 0.000  1001
ip-10-110-54-223.e LINUX X86_64 Unclaimed Idle 0.320  590
Total Owner Claimed Unclaimed Matched Preempting Backfill
X86_64/LINUX 3 0 0 3 0 0 0 0

vcluster >
4. Development Status

- **Improvements**
  - Centralized database to keep virtual machine status
  - Smart policy to estimate proper virtual machines from queue status
  - Smart recovery system from database corruption

```
vcluster > dins
Inst ID ------------------------------- Status
---------------------------------------------------
i-b5ca1bdb stopped
i-3fed3d5a running
---------------------------------------------------
vcluster >
vcluster > stop i-3fed3d5a
vcluster > [MSG]: stopping vm in a separated thread
vcluster > [MSG]: stop a vm of i-3fed3d5a using REST API.
vcluster > [MSG]: done in a separated thread
vcluster > condor_status
Name OpSys Arch State Activity LoadAv Mem
ActvtyTime
Fermicloud007.fnal LINUX X86_64 Unclaimed Idle 0.010 1001
Fermicloud043.fnal LINUX X86_64 Unclaimed Idle 0.000 1001
Total Owner Claimed Unclaimed Matched Preempting Backfill
X86_64/LINUX 2 0 0 2 0 0 0
Total 2 0 0 2 0 0 0
vcluster >
```
5. Roadmap of Y. 2012

We will develop continuously vcluster’s basic functionalities and apply it to small-size real experiments

- Cluster Monitoring Functionalities
- Policy-based VM Management

- Enabling vcluster on FermiCloud + EC2 + Gcloud

- Basic functionalities for launching and terminating vms
  - Condor-based
  - FermiCloud + EC2

- Tablet-based vcluster

- SC2012 Demonstration
6. Conclusions

- With successful collaboration with Fermilab – GCC (Grid and Cloud Computing Department), conceptual idea for vcluster has been developed.

- Design concept of vcluster focuses on developing cloud and batch system agnostic, and very simple virtual cluster system.

- vcluster can make a virtual cluster system over heterogeneous cloud systems like FermiCloud and Amazon EC2.

- Final development target of vcluster this year is to complete the feature of automatic scale up and down with FermiCloud + Gcloud + Amazon EC2.

- Final result will be demonstrated in SC2012
Thank You