



# **CloudCRV and The ATLAS Cluster on the Cloud**

**What do we need? And what do we have?**

**Yushu Yao  
Lawrence Berkeley National Laboratory**

# Overview

- I'll only talk about **technical issues** in this talk. Will leave the policy issues to the more qualified personnels.
- No matter **who runs** our job on the cloud, and **which cloud** they run on (commercial or pledged resources), **who pays** for it, we need an easy and user-transparent way to scale our jobs to the cloud
- And we don't have a way to do that
- What are needed, and what can we do?

Let's see...

# Mode of Operation

**Who calls for Scale-Up?**

# Modes of Operation?

## Centralized:

- Like a tier2 in the sky, deployed by one, run jobs for many.
- E.g. the cluster we are building on Magellan (details later in the talk)

## De-centralized:

- Deployed (and paid) by one, run jobs for himself (e.g. a univ. prof with a credit card and a paper deadline)

Both modes are possible. Independent of which mode, we face the same problems...

# Cloud Problems

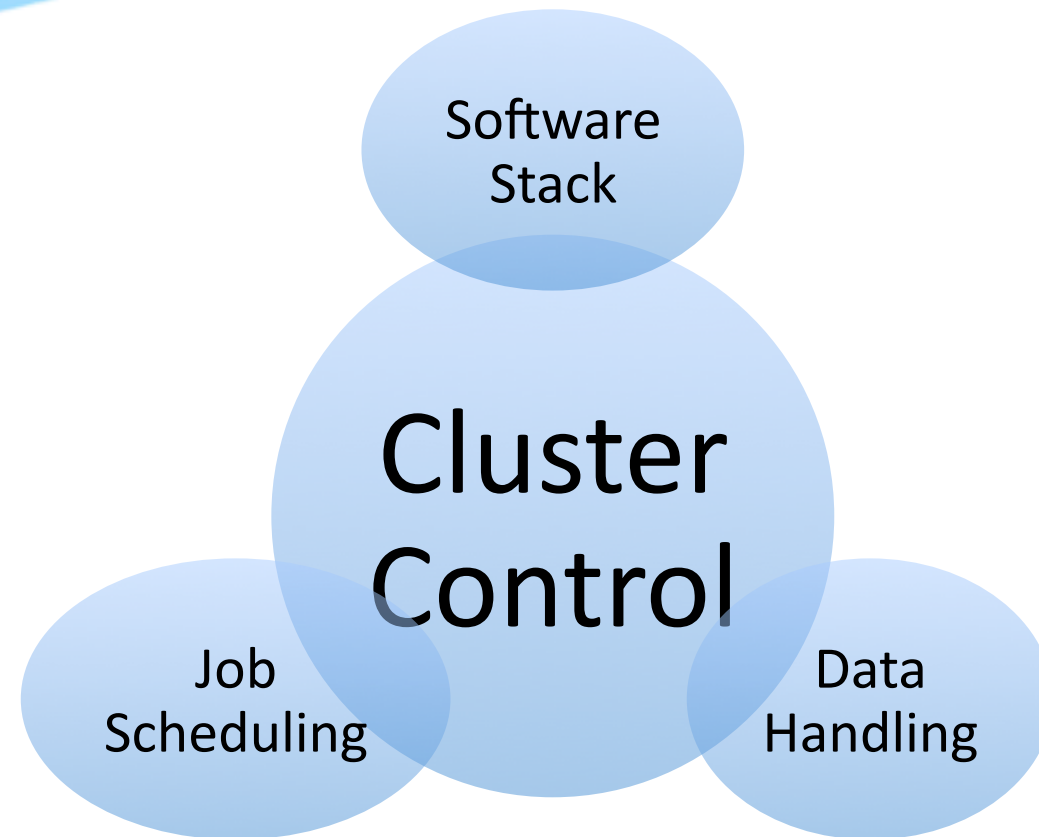
## What do we need to solve?

- **Agility** (Super Scalable! Isn't it supposed to be a benefit?)
  - Yes, but there's no easy way to use it so far (how to setup the resource to run ATLAS, how to distribute jobs, etc)
- **Data** (Two aspects):
  - Getting data from/to cloud is expensive and inefficient
  - Storing the data in the cloud is tricky.

Several Key Components are need to solve these problems...

# Key Components Needed for us to use the Cloud

- Software Stack
- Data Handling
- Job Scheduling
- Cluster Control/  
Management

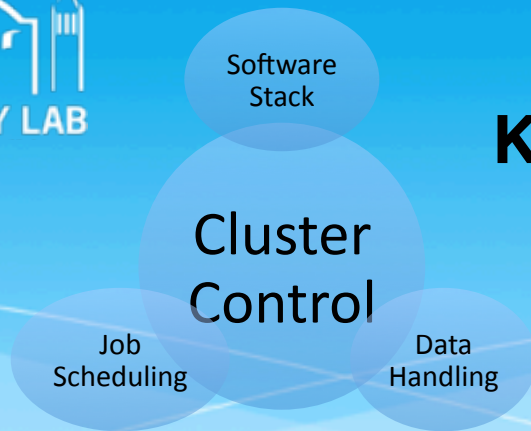


**We need them to be scalable, efficient, and user friendly**



# Software Stack

## Key Components for ATLAS on the Cloud

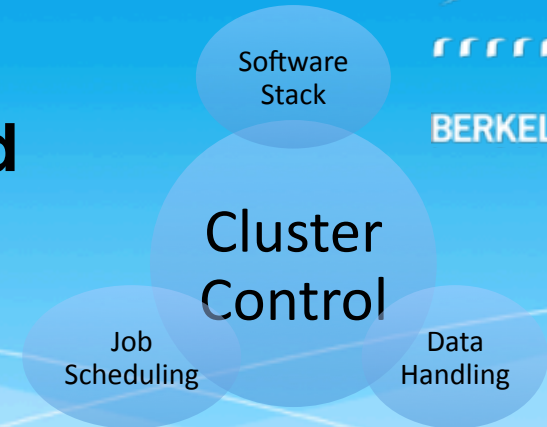


## Thanks to CernVM(-FS)

- Everything except data is provided by CernVM-FS
- Use CernVM we also get an OS for free
- With proxy servers, it can scale as big as we need
- **Cloud Ready, Great!**

# Data Handling

## Key Components for ATLAS on the Cloud



- **Storage on Worker:**

- Very Important: we can't dedicate too many storage nodes, that's waste of money (when no worker is running)

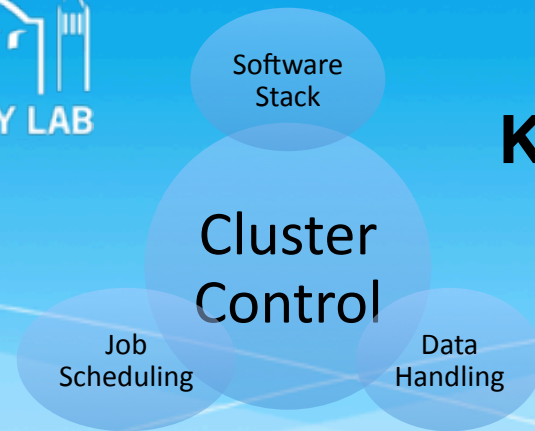
- **Smarter Data Transfer**

- Pre-staging dataset, reuse of data across jobs, etc

- **Possible Solutions:**

- Mount HDFS across the scalable cluster
  - transparent add/remove node (Agility required)
  - Simplify data staging (1-step staging, no need to move from storage to worker)
- Xrootd confederation (discover/transfer data better)
- Reserved links (when possible, reduces transfer time)
- ... ..



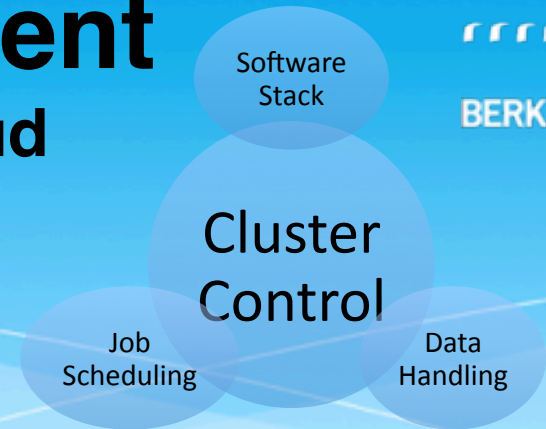


## Panda

- Well tested, works well for ATLAS jobs
- Low overhead (management, etc)
- “Data Smart”, sort of
- Schedule whole node jobs with AthenaMP (much easier to handle when trying to take a node offline)

# Cluster Control/Management

## Key Components for ATLAS on the Cloud



## We need a tool to:

- Allocate cloud resource when needed, release resource when done
- Configure the resources to do ATLAS work
  - tasks like: install CernVM-FS, configure HDFS, setup Panda, etc.
  - note that: each of the above task needs an expert to do
- Any one who need to setup such a cluster should be able to do this with one button click (especially for de-centralized modes)

We developed CloudCRV to do this job...

# An Analogy



Substances

A Cluster	↔	A Fridge
Various of Apps and Services	↔	Parts in a Fridge
Computing Power from Cloud	↔	Electricity

Actors

Cloud Provider	↔	Electricity Provider
Cluster Designer	↔	Fridge Manufacturer
Cluster Manager/Deployer	↔	Fridge User



The guy who needs to deploy a Cluster

Actions

Design a Cluster	↔	Manufacture a Fridge (Need expert)
Deploy a Cluster	↔	Plug the Fridge to the Wall (Simple)

# Pack and Ship ATLAS Cluster like a Fridge

## The Components (roles) in an ATLAS Cluster

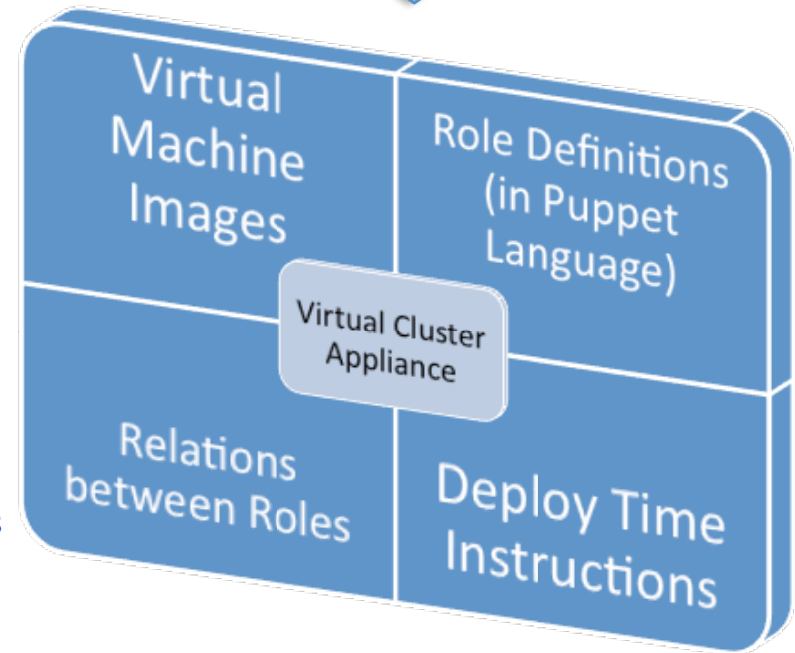
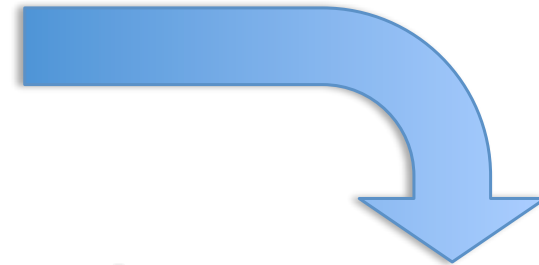


Zoom out version in Backup Slides



Expert

## Create a Virtual Cluster Appliance



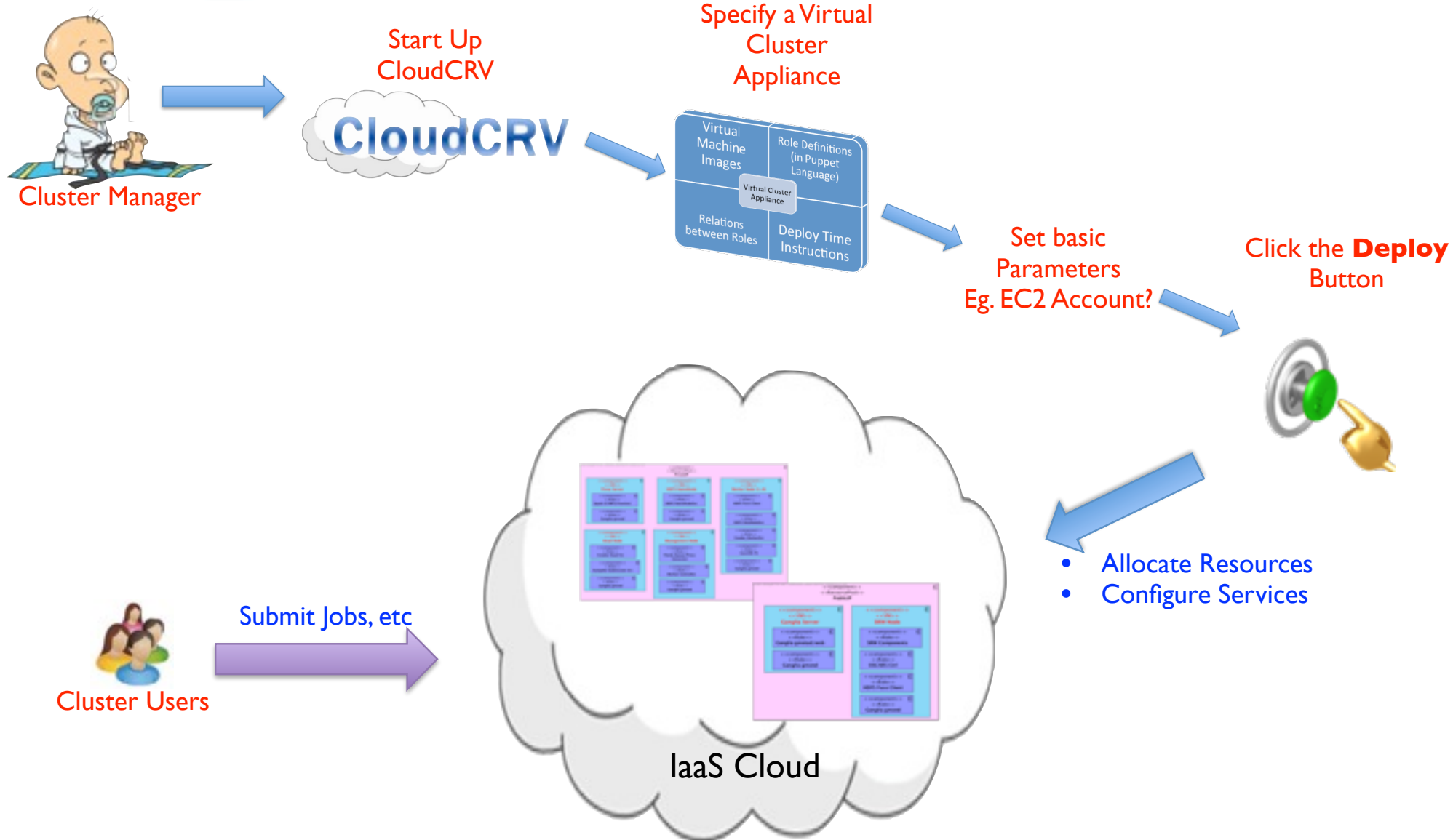
## Ship to cluster manager



Remember those univ. profs with credit cards? Or his postdoc?

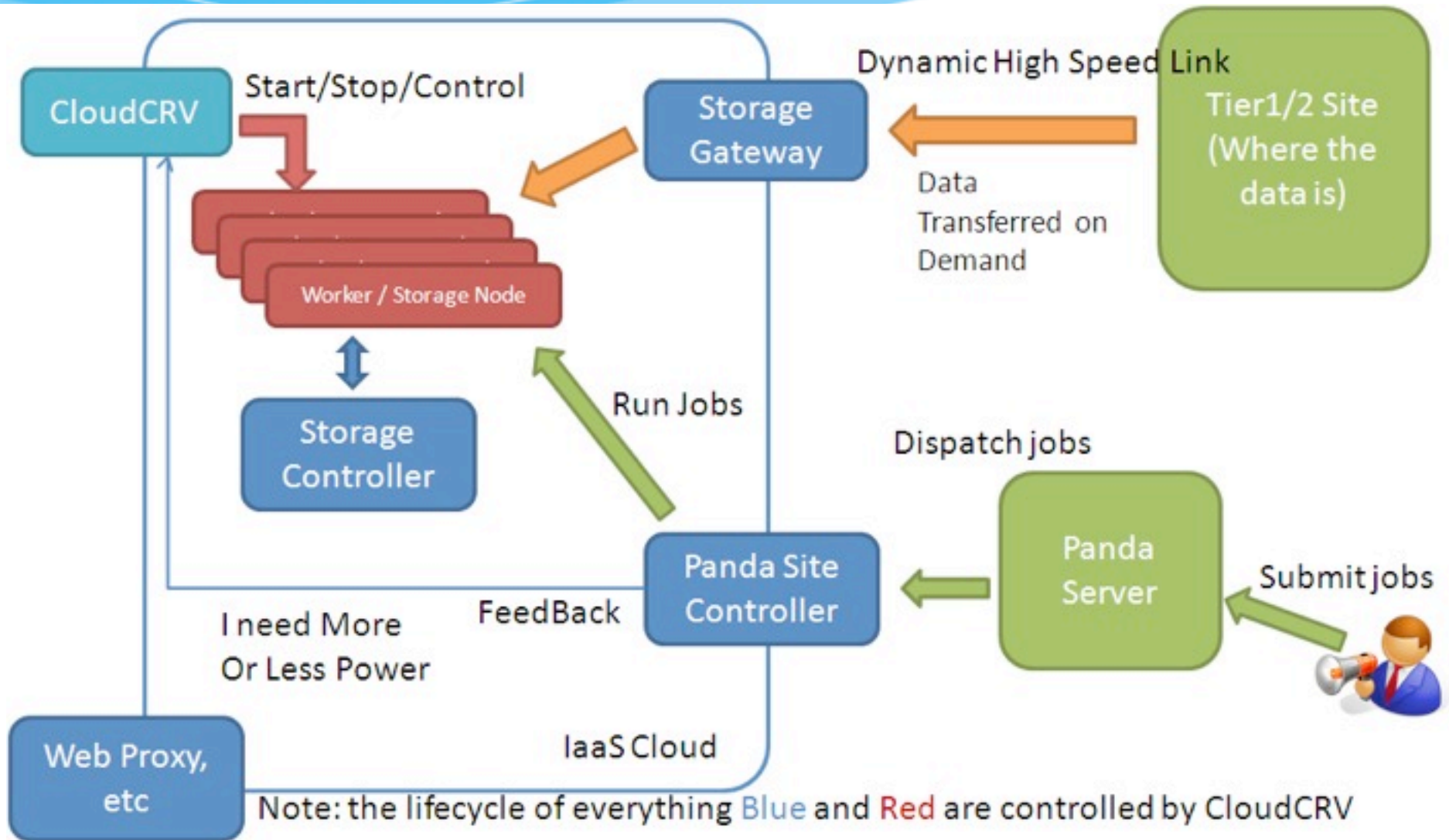


# Deploying an ATLAS Cluster like Plugging in a Fridge



# The ATLAS Virtual Cluster Appliance

# ATLAS Virtual Cluster Appliance



- Scale on-demand
- Storage on Worker (HDFS)
- Panda Based
- High Speed Data Link

# CloudCRV in Action



## CloudCRV

Drive on the Cloud

Clusters Roles Resource Pools/VMs

Login

Now Viewing: login



Only for logged in users

Login

Username:

Password:

Login



Please refer to <https://code.google.com/p/cloudcrv/>

Powered by TurboGears 2

# List of Clusters



The screenshot displays the CloudCRV web interface. At the top, the title "CloudCRV" is shown with the tagline "Drive on the Cloud". Below this is a navigation bar with tabs for "Clusters", "Roles", and "Resource Pools/VMs", along with "Admin" and "Logout" links. The main content area features a table with the following data:

Name	Current Status ↓	Target Status	Attributes	Roles	Action
tier3	INITIALIZED	INITIALIZED	<a href="#">Show Attr</a>	<a href="#">Show Roles</a>	<input type="button" value="Start"/>

# List of Roles

## CloudCRV

Drive on the Cloud

Clusters Roles Resource Pools/VMs
Admin Logout

**INITIALIZED**

**STARTINGVM**

**APPLYING**

**RUNNING**

**REMOVING**

**STOPPINGVM**

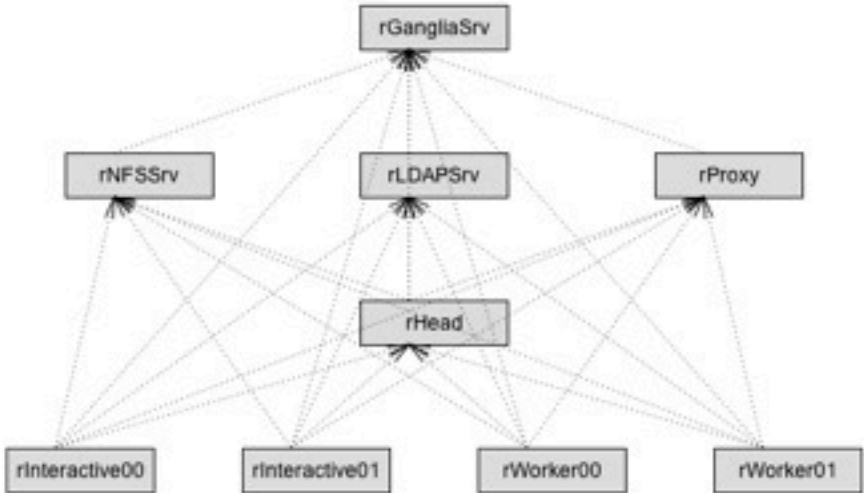
**FIXING**

**EXCEPTION**

Note:

STOPPED = INITIALIZED

STARTED = RUNNING



```

graph TD
    rGangliaSrv --> rNFSSrv
    rGangliaSrv --> rLDAPSrv
    rGangliaSrv --> rProxy
    rGangliaSrv --> rHead
    rNFSSrv --> rHead
    rLDAPSrv --> rHead
    rProxy --> rHead
    rHead --> rInteractive00
    rHead --> rInteractive01
    rHead --> rWorker00
    rHead --> rWorker01
    
```

VM	RoleID	Name	Current Status	Target Status	Attributes	dependOn	Local depOn	dependBy	Local depBy	roleDef
<a href="#">ymGangliaSrv</a>	1	rGangliaSrv	INITIALIZED	INITIALIZED	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	GangliaSrv
<a href="#">ymNFSSrv</a>	2	rNFSSrv	INITIALIZED	INITIALIZED	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	NFSSrv
<a href="#">ymNFSSrv</a>	3	autorole_rNFSSrv_GangliaClient	INITIALIZED	INITIALIZED	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	GangliaClient
<a href="#">ymLDAPSrv</a>	4	rLDAPSrv	INITIALIZED	INITIALIZED	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	LDAPSrv
<a href="#">ymLDAPSrv</a>	5	autorole_rLDAPSrv_GangliaClient	INITIALIZED	INITIALIZED	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	GangliaClient
<a href="#">ymProxy</a>	6	rProxy	INITIALIZED	INITIALIZED	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	Proxy
<a href="#">ymProxy</a>	7	autorole_rProxy_GangliaClient	INITIALIZED	INITIALIZED	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	<a href="#">Show</a>	GangliaClient

# List of ResourcePool/VM

## CloudCRV

Drive on the Cloud

Clusters Roles Resource Pools/VMs
Admin Logout

List of Resource Pools and VMs

### Resource Pool #1: publicIP

*Resource Pool with Publicly Addressed VMs, Puppet Profiles and RHEL5 Clients*

[Show Details](#)

#### List of VMs

VM_ID	Name	Current Status	Target Status	Identifier	PublicIP	PrivateIP	Attributes	Roles
1	vmGangliaSrv	INITIALIZED	INITIALIZED				<a href="#">Show</a>	<a href="#">Show</a>
6	vmInteractive00	INITIALIZED	INITIALIZED				<a href="#">Show</a>	<a href="#">Show</a>
7	vmInteractive01	INITIALIZED	INITIALIZED				<a href="#">Show</a>	<a href="#">Show</a>


### Resource Pool #2: privateIP

*Resource Pool with Privately Addressed VMs, Puppet Profiles and RHEL5 Clients*

[Show Details](#)

#### List of VMs

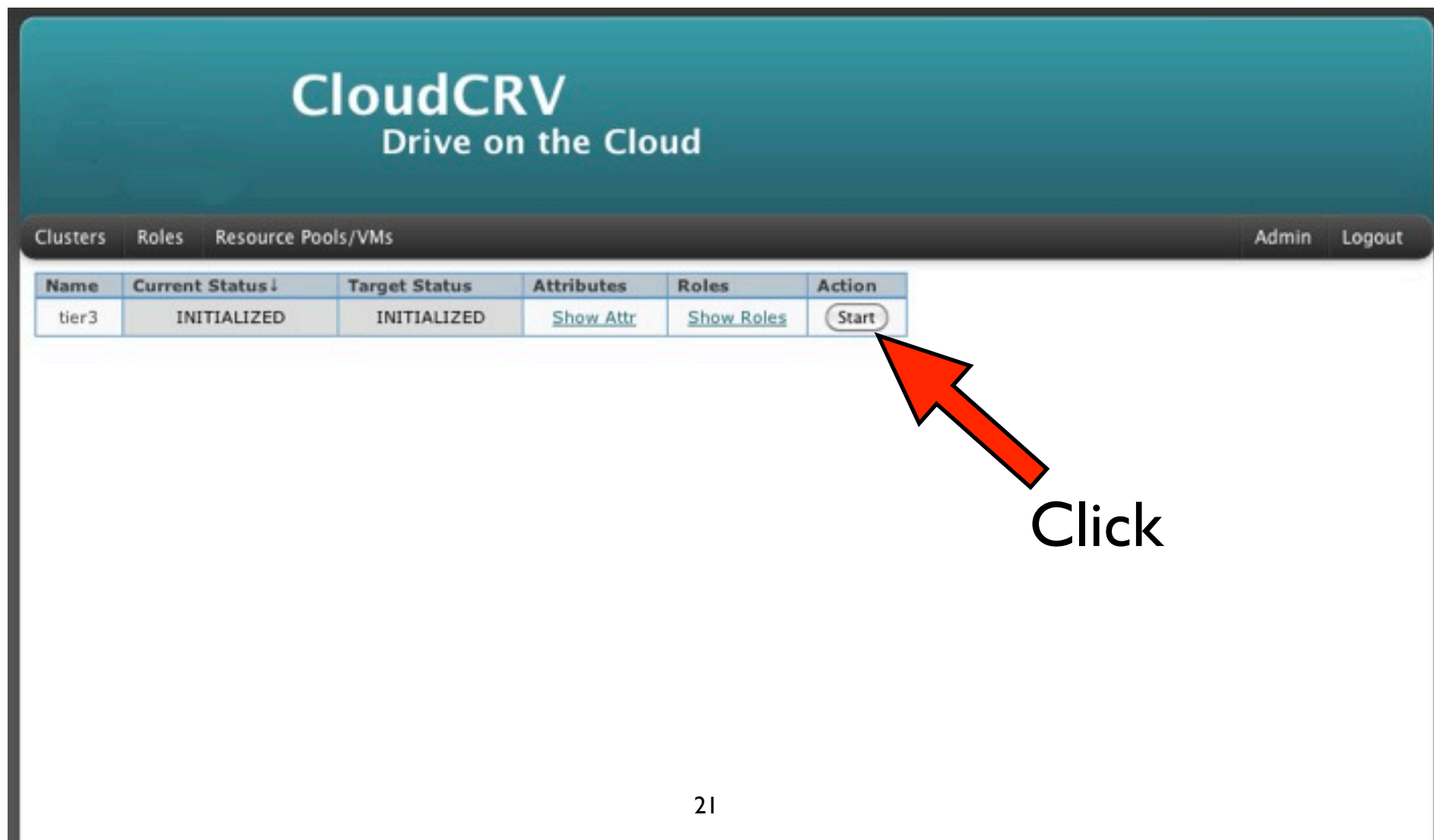
VM_ID	Name	Current Status	Target Status	Identifier	PublicIP	PrivateIP	Attributes	Roles
5	vmHead	INITIALIZED	INITIALIZED				<a href="#">Show</a>	<a href="#">Show</a>
3	vmLDAPSRV	INITIALIZED	INITIALIZED				<a href="#">Show</a>	<a href="#">Show</a>
2	vmNFSSrv	INITIALIZED	INITIALIZED				<a href="#">Show</a>	<a href="#">Show</a>
4	vmProxy	INITIALIZED	INITIALIZED				<a href="#">Show</a>	<a href="#">Show</a>
8	vmWorker00	INITIALIZED	INITIALIZED				<a href="#">Show</a>	<a href="#">Show</a>
9	vmWorker01	INITIALIZED	INITIALIZED				<a href="#">Show</a>	<a href="#">Show</a>



Please refer to <https://code.google.com/p/cloudcrv/>

Powered by TurboGears 2

# Click the Start Button



The screenshot shows the CloudCRV web interface. At the top, there is a teal header with the text "CloudCRV Drive on the Cloud". Below the header is a navigation bar with tabs for "Clusters", "Roles", and "Resource Pools/VMs", and links for "Admin" and "Logout". The main content area displays a table with the following data:

Name	Current Status ↓	Target Status	Attributes	Roles	Action
tier3	INITIALIZED	INITIALIZED	<a href="#">Show Attr</a>	<a href="#">Show Roles</a>	<input type="button" value="Start"/>

A red arrow points to the "Start" button in the "Action" column of the table. The word "Click" is written below the arrow.

# Resources Allocated

## CloudCRV

Drive on the Cloud

Clusters Roles Resource Pools/VMs
Admin Logout

List of Resource Pools and VMs

### Resource Pool #1: publicIP

Resource Pool with Publicly Addressed VMs, Puppet Profiles and RHEL5 Clients

[Show Details](#)

#### List of VMs

VM_ID	Name	Current Status	Target Status	Identifier†	PublicIP	PrivateIP	Attributes	Roles
1	vmGangliaSrv	RUNNING	RUNNING	i-510D08F3	131.243.2.18	192.168.2.4	<a href="#">Show</a>	<a href="#">Show</a>
6	vmInteractive00	RUNNING	RUNNING	i-4C6E09C4	131.243.2.26	192.168.2.10	<a href="#">Show</a>	<a href="#">Show</a>
7	vmInteractive01	RUNNING	RUNNING	i-49FF09D8	131.243.2.29	192.168.2.11	<a href="#">Show</a>	<a href="#">Show</a>


### Resource Pool #2: privateIP

Resource Pool with Privately Addressed VMs, Puppet Profiles and RHEL5 Clients

[Show Details](#)

#### List of VMs

VM_ID	Name	Current Status	Target Status	Identifier†	PublicIP	PrivateIP	Attributes	Roles
9	vmWorker01	RUNNING	RUNNING	i-579508C1	0.0.0.0	192.168.2.13	<a href="#">Show</a>	<a href="#">Show</a>
8	vmWorker00	RUNNING	RUNNING	i-57770A55	0.0.0.0	192.168.2.12	<a href="#">Show</a>	<a href="#">Show</a>
3	vmLDAPsrv	RUNNING	RUNNING	i-4FAD0946	0.0.0.0	192.168.2.6	<a href="#">Show</a>	<a href="#">Show</a>
2	vmNFSSrv	RUNNING	RUNNING	i-42060899	0.0.0.0	192.168.2.5	<a href="#">Show</a>	<a href="#">Show</a>
4	vmProxy	RUNNING	RUNNING	i-3AB60680	0.0.0.0	192.168.2.7	<a href="#">Show</a>	<a href="#">Show</a>
5	vmHead	RUNNING	RUNNING	i-32D205BE	0.0.0.0	192.168.2.9	<a href="#">Show</a>	<a href="#">Show</a>



Please refer to <https://code.google.com/p/cloudcrv/>

Powered by TurboGears 2

# Roles Defined

## CloudCRV

Drive on the Cloud

Clusters Roles Resource Pools/VMs
Admin Logout

**INITIALIZED**

**STARTINGVM**

**APPLYING**

**RUNNING**

**REMOVING**

**STOPPINGVM**

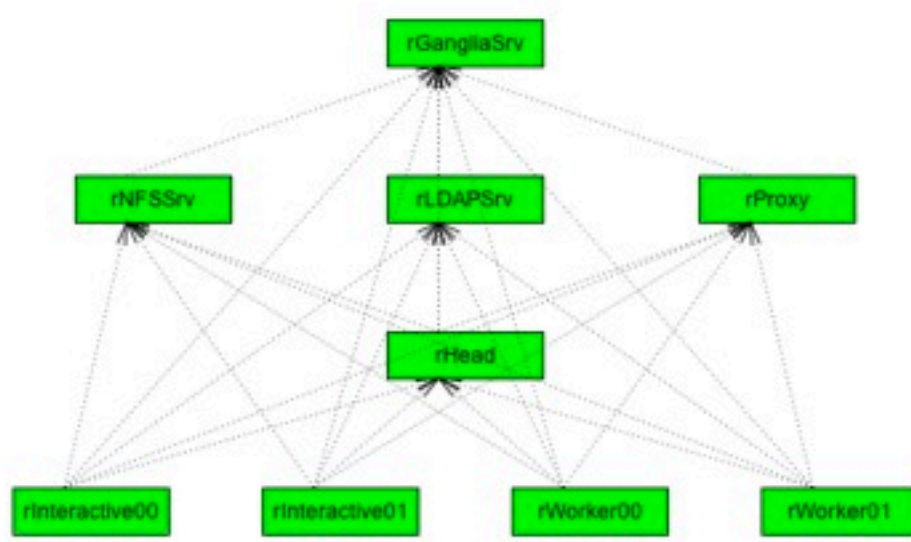
**FIXING**

**EXCEPTION**

Note:

STOPPED = INITIALIZED

STARTED = RUNNING



```

graph TD
    rGangliaSrv --- rNFSSrv
    rGangliaSrv --- rLDAPsrv
    rGangliaSrv --- rProxy
    rGangliaSrv --- rHead
    rNFSSrv --- rinteractive00
    rNFSSrv --- rinteractive01
    rLDAPsrv --- rinteractive00
    rLDAPsrv --- rinteractive01
    rLDAPsrv --- rWorker00
    rLDAPsrv --- rWorker01
    rProxy --- rinteractive00
    rProxy --- rinteractive01
    rProxy --- rWorker00
    rProxy --- rWorker01
    rHead --- rinteractive00
    rHead --- rinteractive01
    rHead --- rWorker00
    rHead --- rWorker01
    
```

Hide Autoroles

VMi	RoleID	Name	Current Status	Target Status	Attributes	dependOn	Local depOn	dependBy	Local depBy	roleDef	enable
vmGangliaSrv	1	rGangliaSrv	RUNNING	RUNNING	Show	Show	Show	Show	Show	GangliaSrv	True
vmHead	8	rHead	RUNNING	RUNNING	Show	Show	Show	Show	Show	Head	True
vmHead	9	autorole_rHead_LDAPClient	RUNNING	RUNNING	Show	Show	Show	Show	Show	LDAPClient	True
vmHead	10	autorole_rHead_NFSCClient	RUNNING	RUNNING	Show	Show	Show	Show	Show	NFSCClient	True
vmHead	11	autorole_rHead_GangliaClient	RUNNING	RUNNING	Show	Show	Show	Show	Show	GangliaClient	True
vmInteractive00	12	rInteractive00	RUNNING	RUNNING	Show	Show	Show	Show	Show	Interactive	True

## CloudCRV Drive on the Cloud

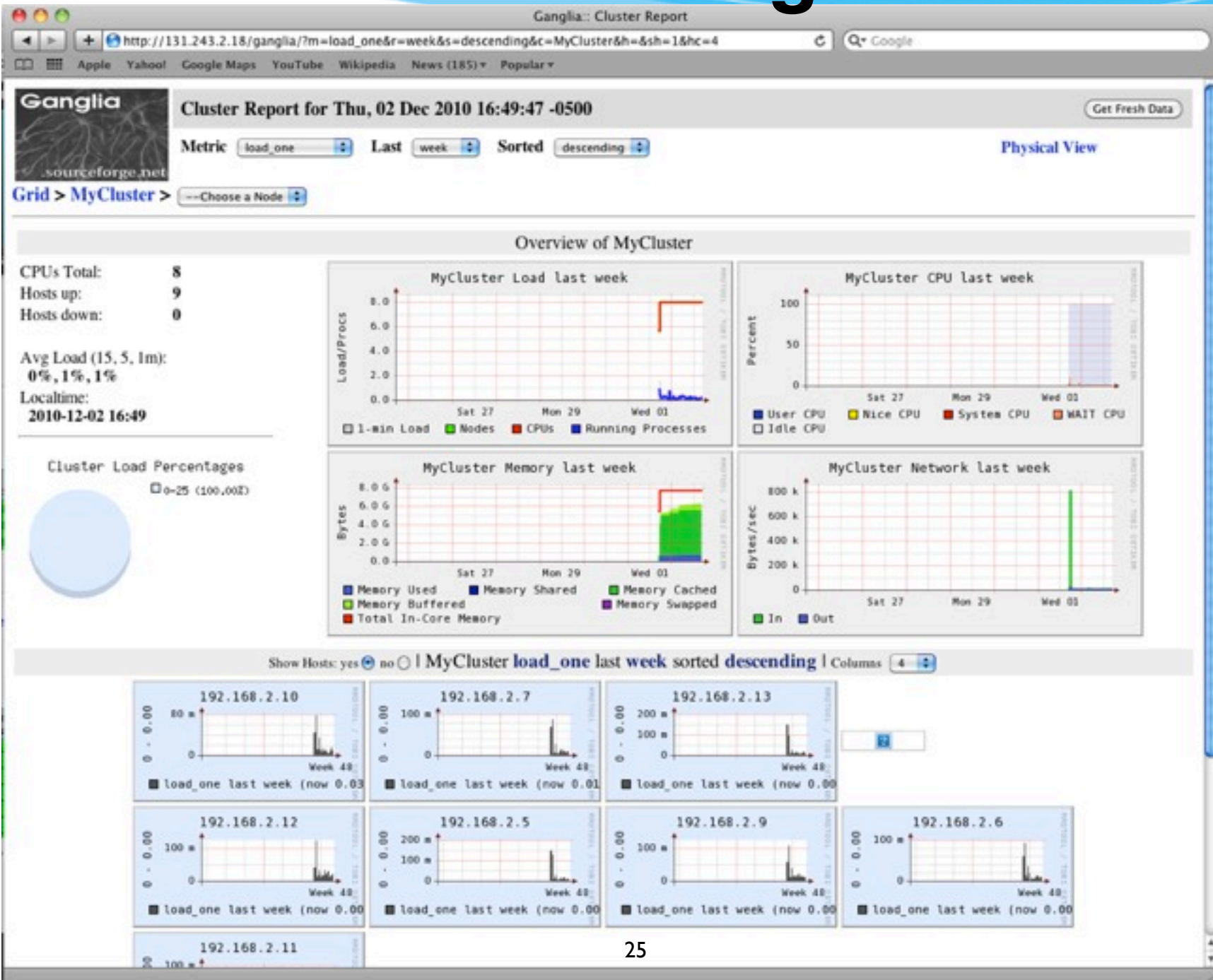
Clusters Roles Resource Pools/VMs

Admin Logout

Name	Current Status ↓	Target Status	Attributes	Roles	Action
tier3	<b>RUNNING</b>	<b>RUNNING</b>	<a href="#">Show Attr</a>	<a href="#">Show Roles</a>	<input type="button" value="Stop"/>



# Testing The Cluster



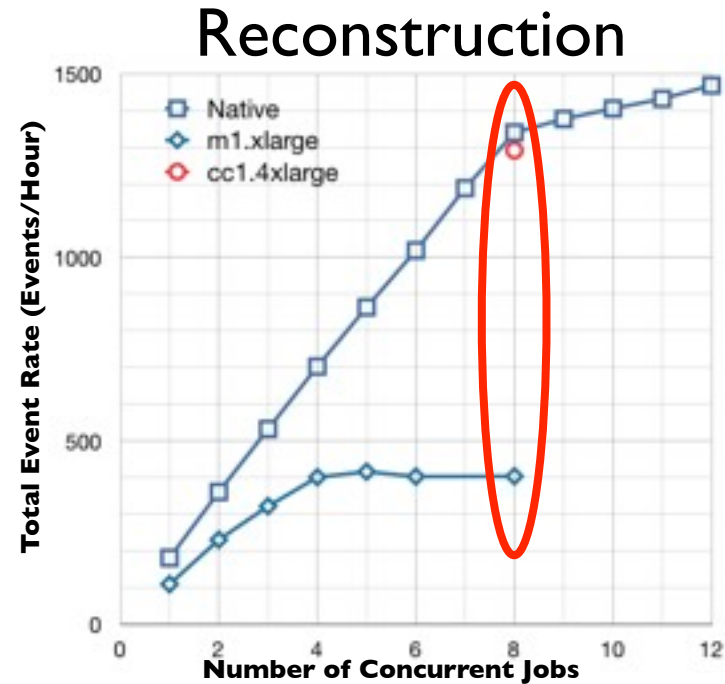
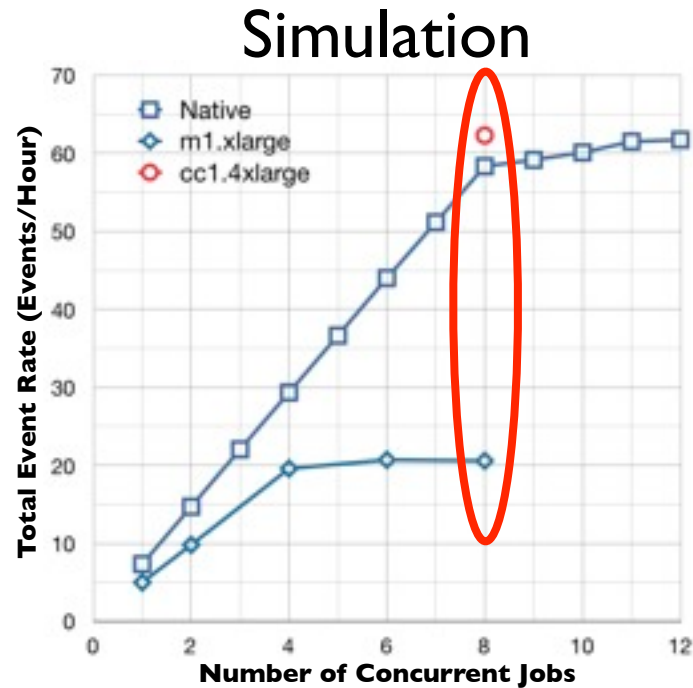
# Summary

- We need a way to use the resource on the cloud!
- We propose to use **CloudCRV and Virtual Cluster Appliance** to help deploy ATLAS Cluster to the Cloud
- Key components: Software Stack, Job Scheduling, Data Handling
- As a proof of concept, we are building an ATLAS cluster on **Magellan**
- Help needed! We need to work together!

# Cost?

**EC2 is costly so far, however...**

# Cost Estimate (Cost per 1K Evt)



Assume we run 8 concurrent jobs for all cases, the cost per 1k event is calculated.

Cost / 1K Event (USD)	EC2		Tier3 Size Center	Large Center (hundreds K cores)
	m1.xlarge	cc1.4xlarge		
Simulation	37	26	11	5
Reconstruction	1.88+Storage	1.24+Storage	0.48	0.24

Details please refer to my talk in the previous ATLAS SW Workshop

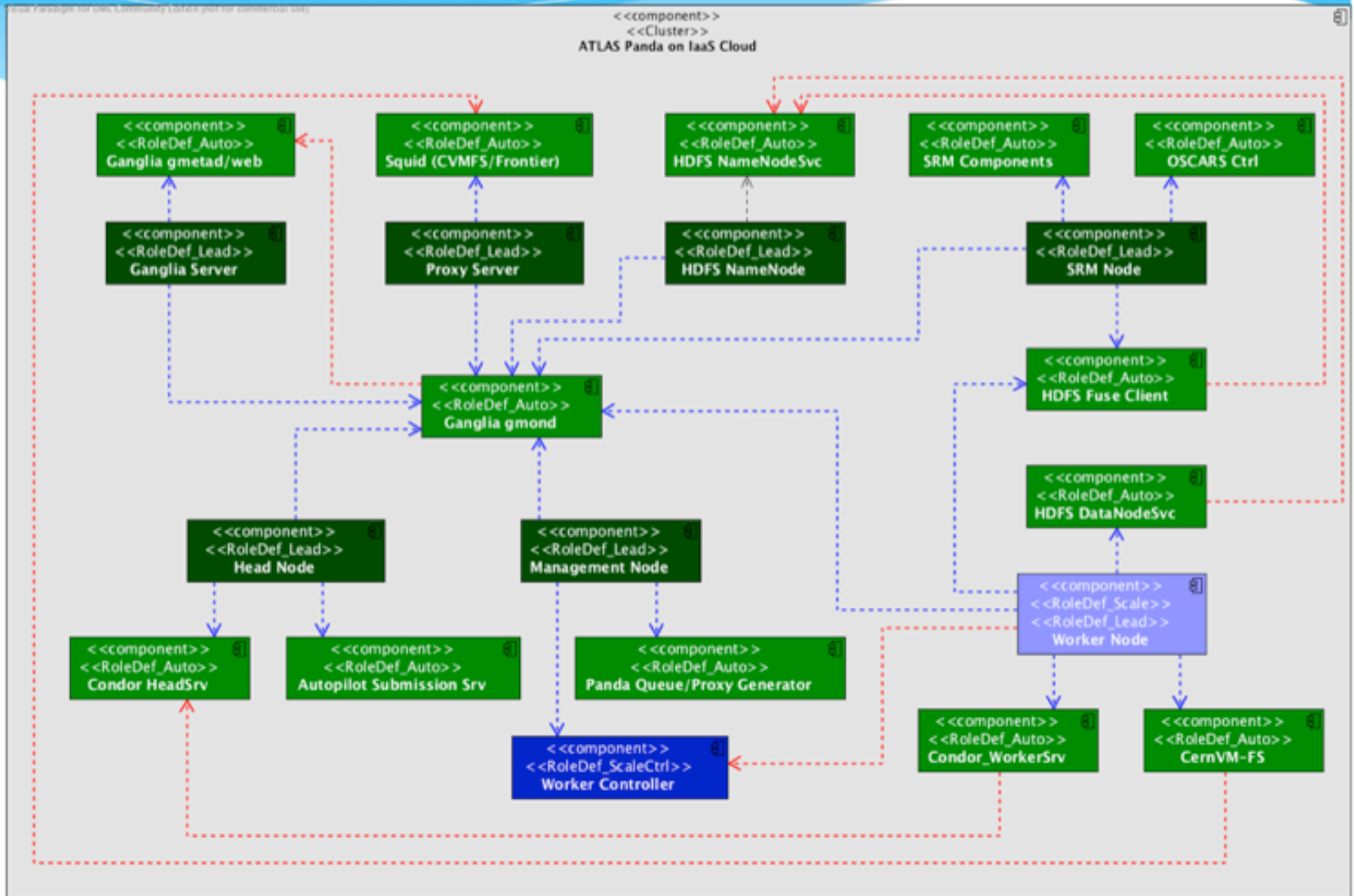
However with EC2 **spot instances**, the game might change!

**Stay Tuned for the next talk...**



# Backup

# Roles in An ATLAS Cluster



# Cost Calculation Assumptions



Not for accurate calculation, cost might be different for individual computer center.

- m l.xlarge: **\$0.19** per core-hour (Storage excluded)
- c l.xlarge: **\$0.20** per core-hour (Storage excluded)
- For US ATLAS Tier3 Center a rough estimate is around \$0.05-0.10 per core-hour (including initial hardware and support), **we use the number \$0.08** (Storage included)
- Large data centers (hundreds of K cores), \$0.02-0.06 per core-hour, **we use the number \$0.04** (Storage included)