CloudMan

Working with CloudMan

Enis Afgan

@ GCC 2014, Baltimore

Welcome to Galaxy on the Cloud
managed by CloudMan

Your history is empty. Click 'Get Data' on the left pane to start

Researchers

CloudMan

100+

100sGB

Cloud resources

How to use the cloud?

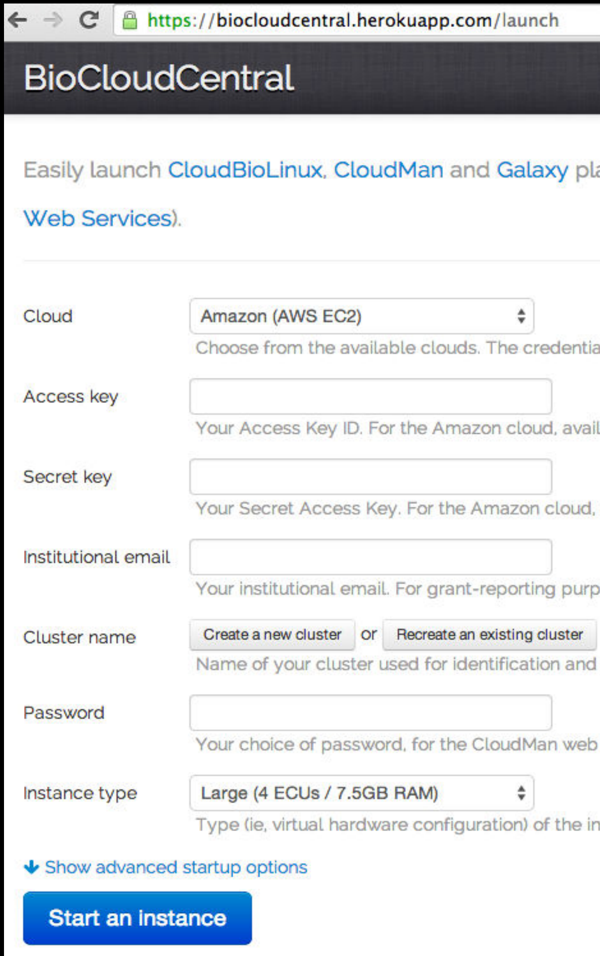
1. Get an account on the supported cloud
2. Start a master instance via the cloud web console or CloudLaunch
3. Use CloudMan's web interface on the master instance to manage the platform
4. Use or customize Galaxy

YOUR TURN

Launch an instance

1. Visit biocloudcentral.org
2. Enter the access key and secret key provided by Dave Clements on 6/25
3. Provide your email address
4. Use your initials as the cluster name
5. Set any password (and remember it)
6. Keep Large instance type
7. Start your instance

Wait for the instance to start (~2-3 minutes)



https://biocloudcentral.herokuapp.com/launch

BioCloudCentral

Easily launch CloudBioLinux, CloudMan and Galaxy pla

Web Services).

Cloud: Amazon (AWS EC2)
Choose from the available clouds. The credentia

Access key:
Your Access Key ID. For the Amazon cloud, avail

Secret key:
Your Secret Access Key. For the Amazon cloud,

Institutional email:
Your institutional email. For grant-reporting purp

Cluster name: Create a new cluster or Recreate an existing cluster
Name of your cluster used for identification and

Password:
Your choice of password, for the CloudMan web

Instance type: Large (4 ECUs / 7.5GB RAM)
Type (ie, virtual hardware configuration) of the in

For more details, see wiki.galaxyproject.org/CloudMan

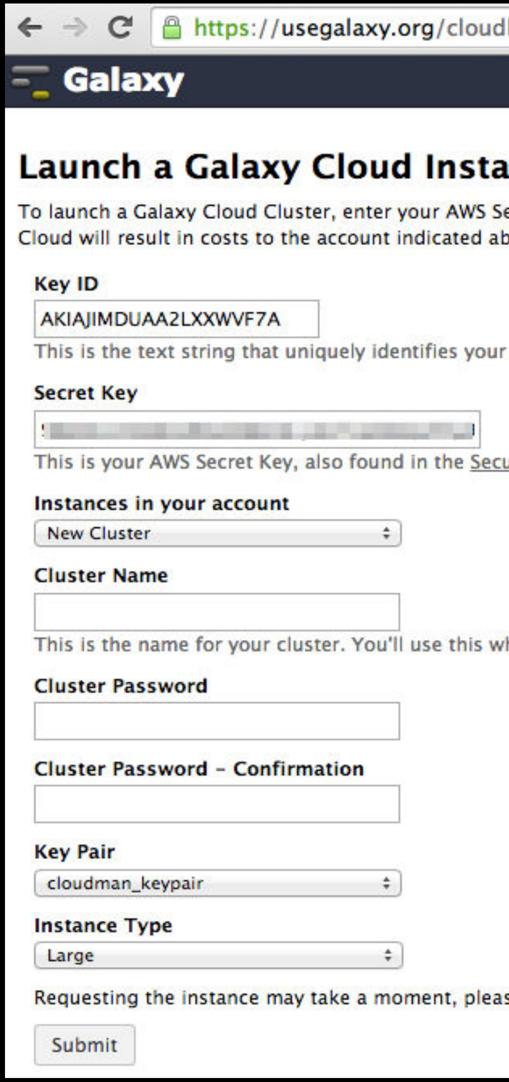
amazon
web services

Launch an instance

1. Visit usegalaxy.org/cloudlaunch
2. Enter your access key and secret key

   (credits provided by Dave Clements on 6/25)
3. Choose New cluster
4. Set any name as the cluster name
5. Set any password
6. Keep cloudman_keypair
7. Keep Large instance type
8. Launch your instance

   Wait for the instance to start (~2-3 minutes)



Launch a Galaxy Cloud Insta

To launch a Galaxy Cloud Cluster, enter your AWS Se
Cloud will result in costs to the account indicated ab

Key ID

AKIAJIMDUAA2LXXWVF7A

This is the text string that uniquely identifies your

Secret Key

[redacted]

This is your AWS Secret Key, also found in the Secu

Instances in your account

New Cluster

Cluster Name

This is the name for your cluster. You'll use this wh

Cluster Password

Cluster Password - Confirmation

Key Pair

cloudman_keypair

Instance Type

Large

Requesting the instance may take a moment, pleas

For more details, see
wiki.galaxyproject.org/CloudMan

Customize your instance - install a new tool

```
cd /mnt/galaxy/export
wget http://heanet.dl.sourceforge.net/project/dnaclust/
arallel_release_3/dnaclust_linux_release3.zip
unzip dnaclust_linux_release3.zip
cd dnaclust_linux_release3
chmod +x *
```

Use the new tool in the cluster mode

1. Create a new sample shell file to run the tool; call it *job_script.sh* with the following content:

   ```
   #$ -cwd
   ./dnaclust -l -s 0.9 /mnt/workshop-data/mtDNA.fasta
   ```

2. Submit single job to SGE queue

   ```
   qsub job_script.sh
   ```

3. Check the queue: `qstat -f`
4. Job output will be in the local directory in file *job_script.sh.o#*
5. Start a number of instances of the script:

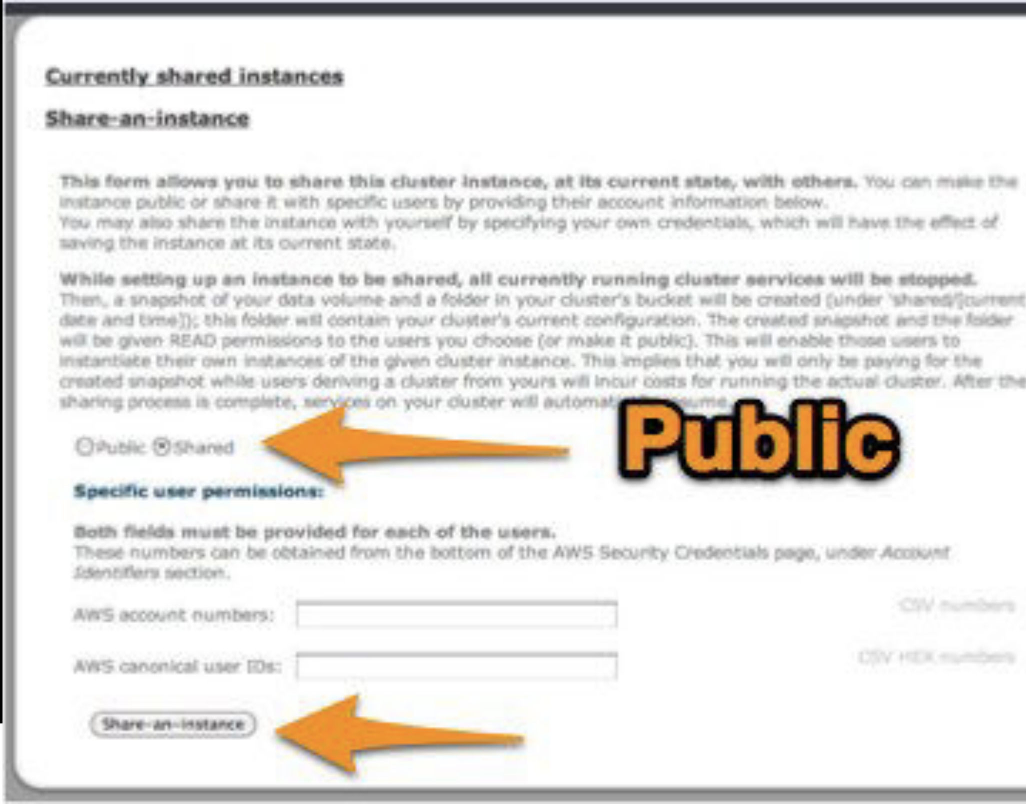
   `qsub job_script.sh` (*10)

   `watch qstat –f`

    1. See all jobs lined up
6. See auto-scaling in action (using /cloud) [1.5-2 mins]
7. See jobs being distributed

Sharing-an-Instance

- Share the entire CloudMan platform
  - Includes all of user data and even the customizations
- Publish a self-contained analysis
- Make a note of the *share-string* and send it to your neighbor



Currently shared instances

Share-an-instance

This form allows you to share this cluster instance, at its current state, with others. You can make the instance public or share it with specific users by providing their account information below.
You may also share the instance with yourself by specifying your own credentials, which will have the effect of saving the instance at its current state.

While setting up an instance to be shared, all currently running cluster services will be stopped. Then, a snapshot of your data volume and a folder in your cluster's bucket will be created (under 'shared/[current date and time]); this folder will contain your cluster's current configuration. The created snapshot and the folder will be given READ permissions to the users you choose (or make it public). This will enable those users to instantiate their own instances of the given cluster instance. This implies that you will only be paying for the created snapshot while users deriving a cluster from yours will incur costs for running the actual cluster. After the sharing process is complete, services on your cluster will automat[illegible]ume.

○ Public ● Shared

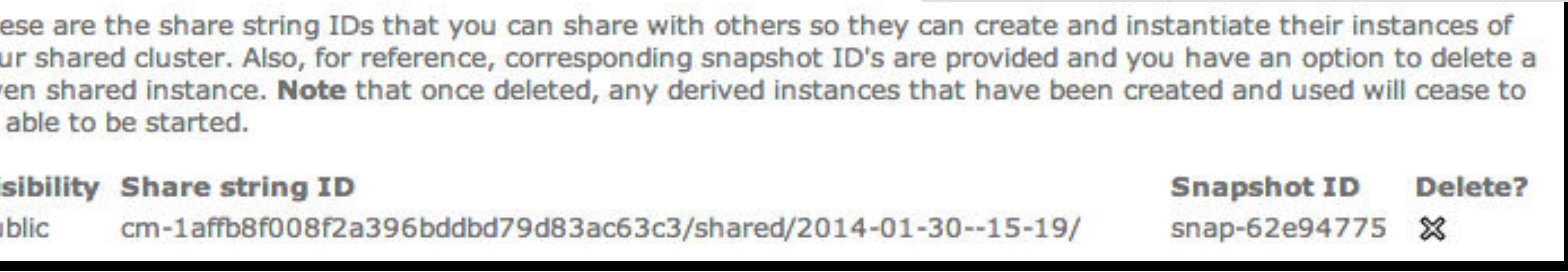
Specific user permissions:

Both fields must be provided for each of the users.
These numbers can be obtained from the bottom of the AWS Security Credentials page, under *Account Identifiers* section.

AWS account numbers: CSV numbers

AWS canonical user IDs: CSV HEX numbers

urrently shared instances

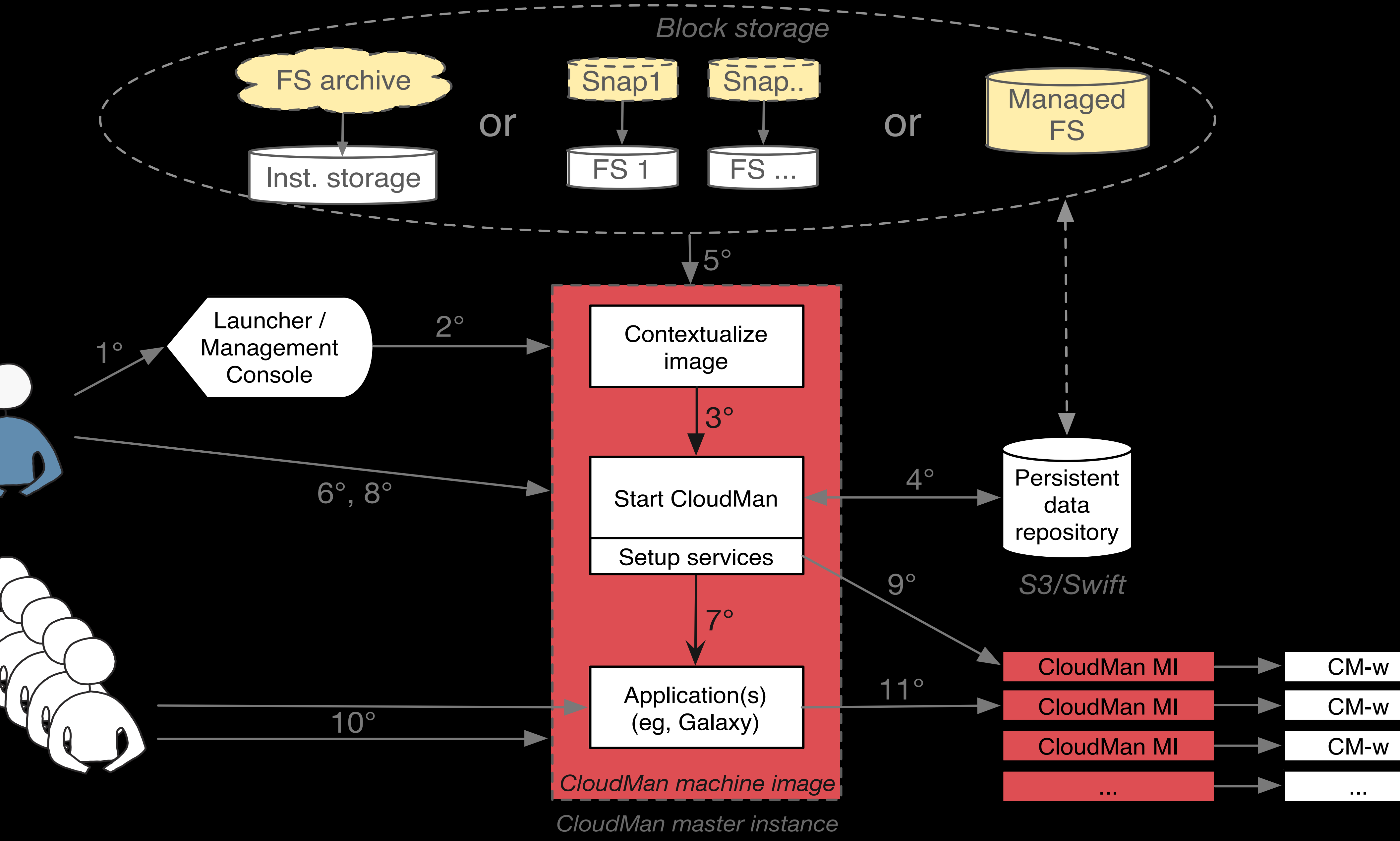


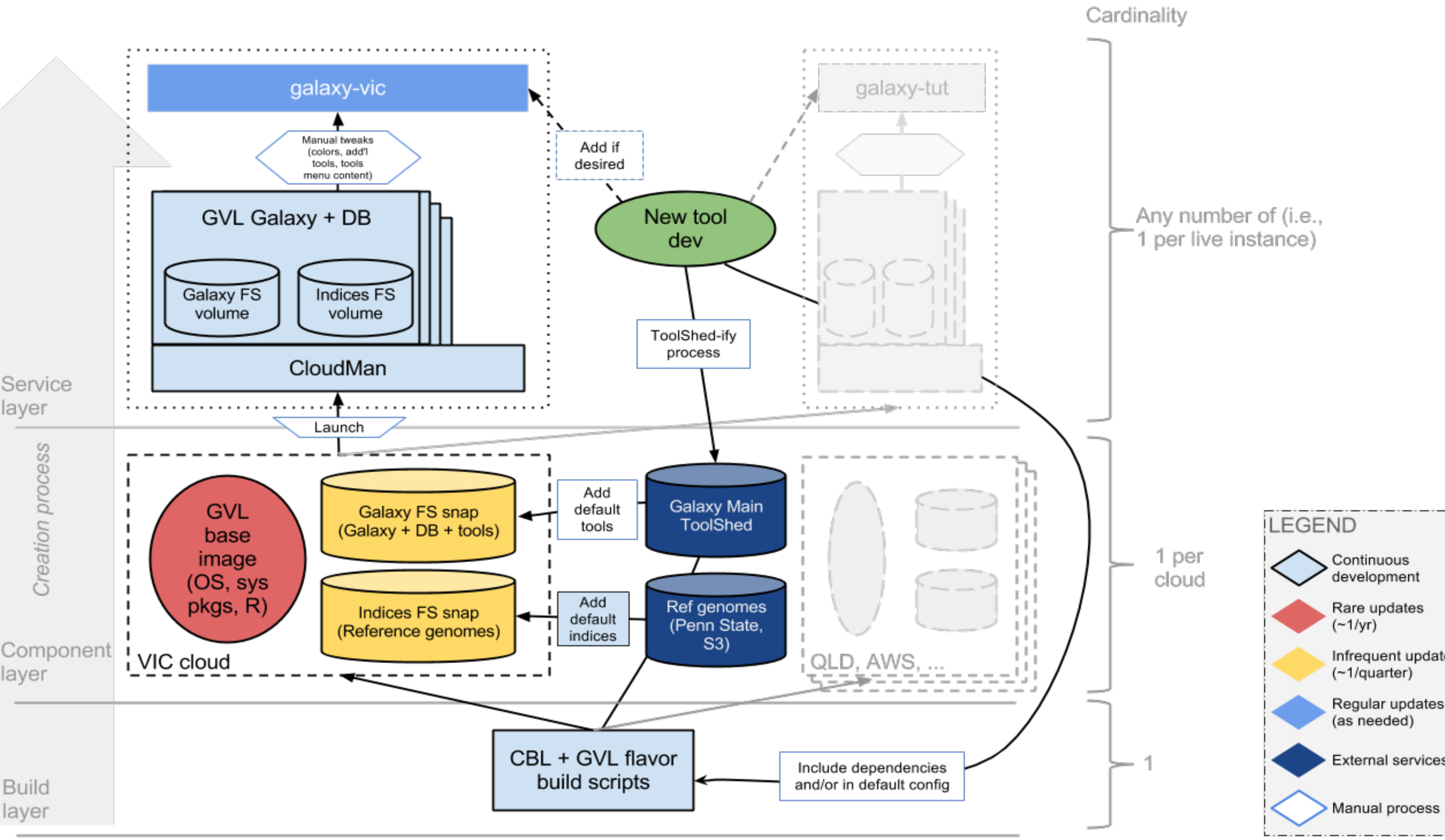
ese are the share string IDs that you can share with others so they can create and instantiate their instances of ur shared cluster. Also, for reference, corresponding snapshot ID's are provided and you have an option to delete a ven shared instance. **Note** that once deleted, any derived instances that have been created and used will cease to able to be started.

sibility	Share string ID	Snapshot ID	Delete?
blic	cm-1affb8f008f2a396bddbd79d83ac63c3/shared/2014-01-30--15-19/	snap-62e94775	☒

Technical details

UNDERPINNING CONCEPTS

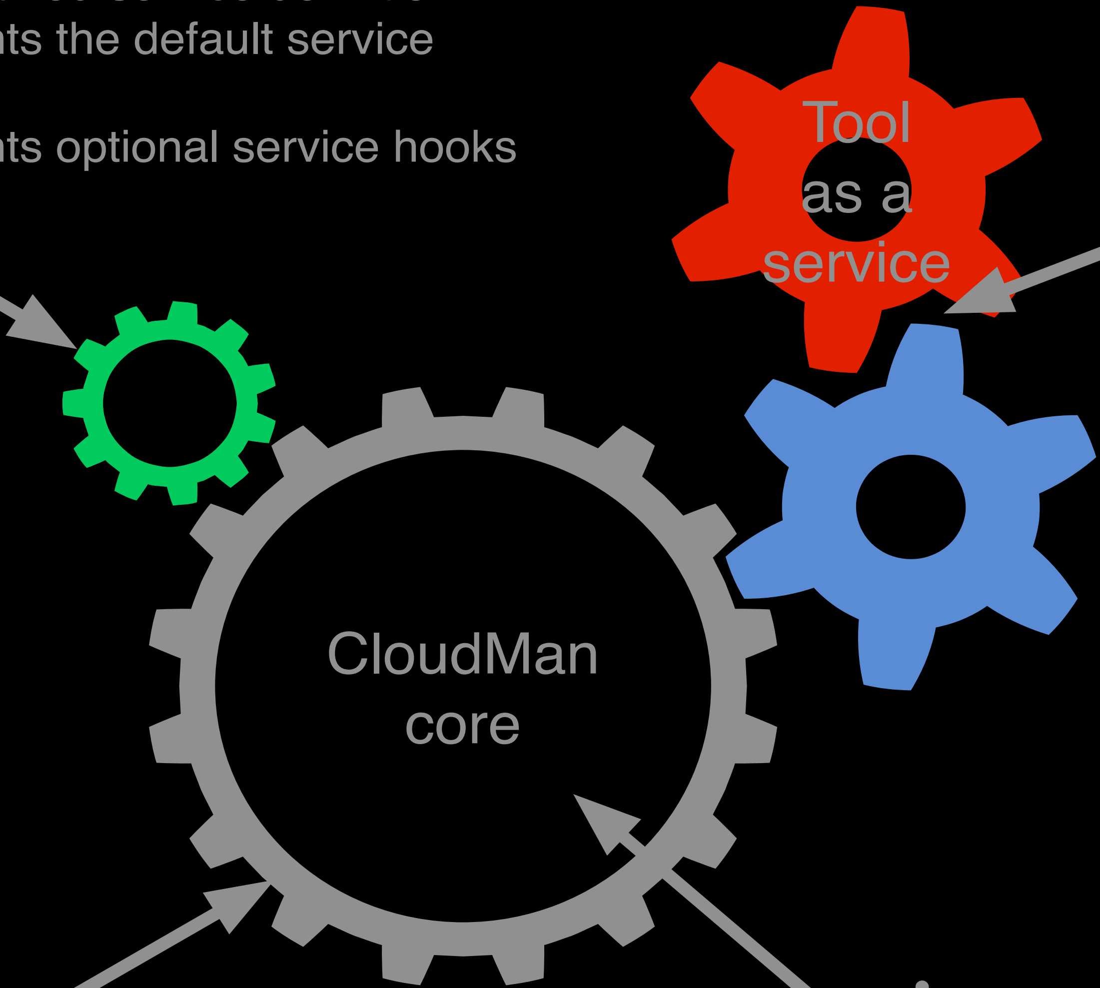

Block storage
FS archive
Inst. storage
or
Snap1
FS 1
Snap..
FS ...
or
Managed FS
5°
1°
Launcher / Management Console
2°
Contextualize image
3°
Start CloudMan
Setup services
6°, 8°
4°
Persistent data repository
S3/Swift
9°
7°
Application(s) (eg, Galaxy)
10°
11°
CloudMan MI
CloudMan MI
CloudMan MI
...
CM-w
CM-w
CM-w
...
CloudMan machine image
CloudMan master instance


Cardinality
galaxy-vic
Manual tweaks (colors, add'l tools, tools menu content)
GVL Galaxy + DB
Galaxy FS volume
Indices FS volume
CloudMan
Add if desired
New tool dev
ToolShed-ify process
galaxy-tut
Any number of (i.e., 1 per live instance)
Service layer
Launch
Creation process
GVL base image (OS, sys pkgs, R)
Galaxy FS snap (Galaxy + DB + tools)
Indices FS snap (Reference genomes)
VIC cloud
Add default tools
Add default indices
Galaxy Main ToolShed
Ref genomes (Penn State, S3)
QLD, AWS, ...
1 per cloud
Component layer
CBL + GVL flavor build scripts
Include dependencies and/or in default config
1
Build layer
LEGEND
Continuous development
Rare updates (~1/yr)
Infrequent updat (~1/quarter)
Regular updates (as needed)
External services
Manual process

What do you get?

- Cluster-in-the-cloud: SGE
- Galaxy on the Cloud + control
- Customizable: tools, configs, data
- Sharable

- Extensible



- self-contained service definition
- implements the default service interface
- implements optional service hooks

- service dependencies
- automatic dependency manageme
- dynamic loading

- well-defined interface / API
- web interface

- cloud resource management
- multi-cloud interface
- service management and monitoring
- user interaction
- system state

Supported cloud middleware

1. Amazon Web Services
2. OpenStack
3. Eucalyptus
4. OpenNebula

Building

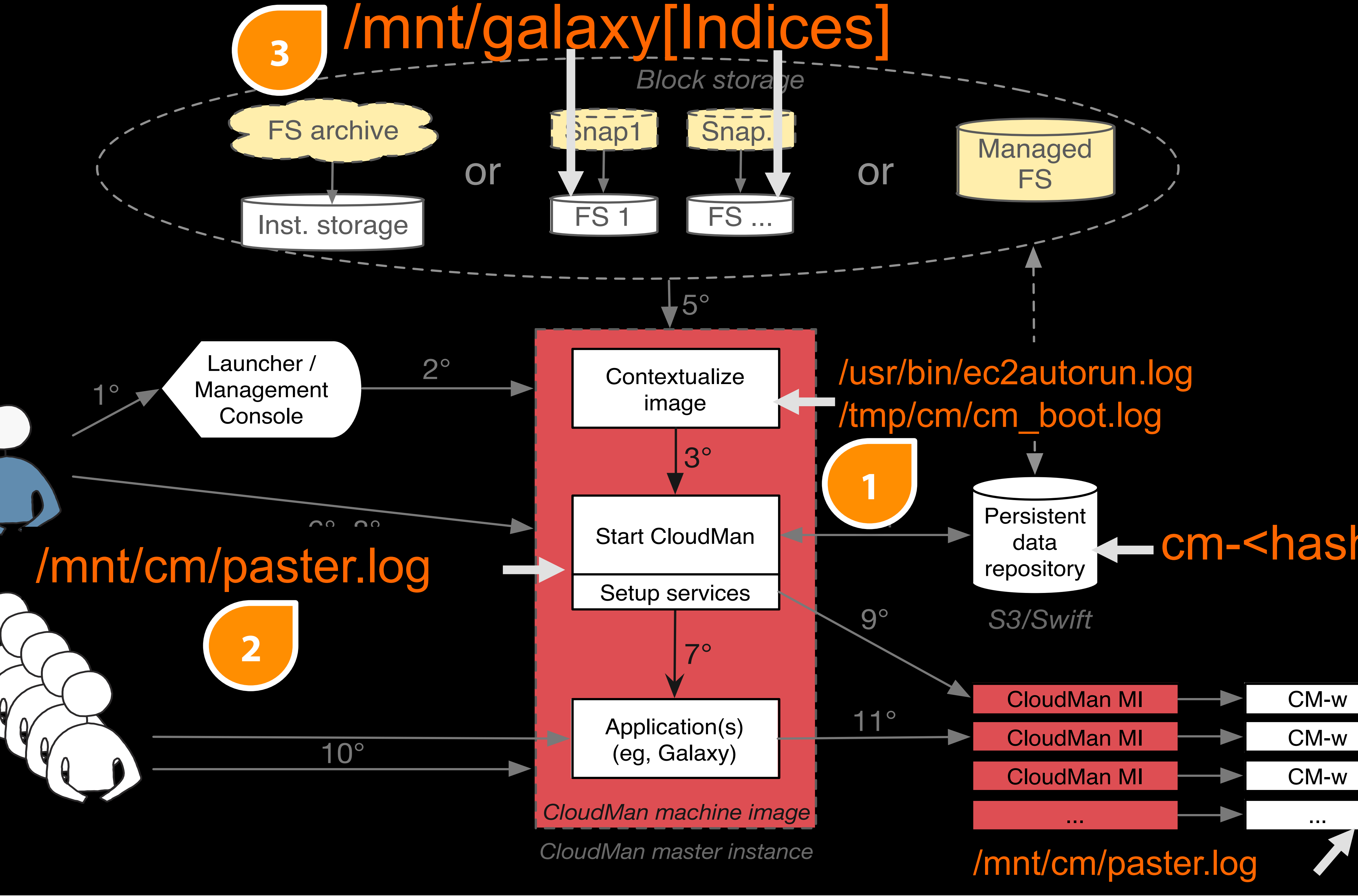
- Leverage CloudBioLinux build framework
- Via GVL flavor
  - Base CloudMan machine image
  - Tools and data
- There are also more specific CBL flavors available
  - cloudman

github.com/afgane/gvl_flavor

Deploying

- Integrated with BioCloudCentral.org app
    - Use the public one, deploy your own or run locally
- BCC supports multiple clouds

Troubleshooting



3 /mnt/galaxy[Indices]

1 /usr/bin/ec2autorun.log
/tmp/cm/cm_boot.log

cm-<hash

2 /mnt/cm/paster.log

/mnt/cm/paster.log