# Pre-requisite

* RHEL 6.5 x86\_64
* Python 2.7.6 installed in /usr/local/bin
* PyVCF 0.67 installed
* GPFS software installed and configured as GPFS cluster client node
* GPFS file system mounted

# Installation

* Logon as root
* Create home directory - /root/pyproj

## Verify python version

# python -V

Python 2.7.6

## Install Redis

cd /root/pyproj

git clone https://github.com/redis/hiredis.git

cd hiredis

make

cd ../

## Unpack Caddy Files

* cd /root/pyproj
* unzip caddy\_rel1.zip
* /root/pyproj

-rw-r--r--. 1 root root 2386 May 15 22:12 cad\_analyze\_vcf.py

-rw-r--r--. 1 root root 514 May 15 23:22 cad\_gpfs\_exec.py

-rw-r--r--. 1 root root 377 May 11 06:49 cad\_gpfs\_ingest.json

-rw-r--r--. 1 root root 6904 May 15 23:22 cad\_gpfs\_ingest.py

-rw-r--r--. 1 root root 642 May 6 11:18 exec.list

drwxr-xr-x. 2 root root 4096 May 15 20:07 meta\_api

-rwxr-xr-x. 1 root root 142326 May 15 20:29 meta\_api.so

-rw-r--r--. 1 root root 510 May 6 15:39 vcf.list

* /root/pyproj/meta\_api

-rw-r--r--. 1 root root 5916 May 13 16:30 libmetaengine.a

-rw-r--r--. 1 root root 4011 May 15 19:03 meta\_api.c

-rwxr-xr-x. 1 root root 142326 May 15 19:03 meta\_api.so

-rw-r--r--. 1 root root 711 May 13 14:02 metaengine.h

-rw-r--r--. 1 root root 207 May 13 16:47 test.py

## Compile meta\_api.c

* cd /root/pyproj/meta\_api
* gcc -fPIC -shared -I/usr/local/include/python2.7 -lpython2.7 -ometa\_api.so meta\_api.c libmetaengine.a ../hiredis/libhiredis.a
* cp meta\_api.so ../

# Testing

## Source Files

* /root/pyproj/cad\_gpfs\_ingest.json
* /root/pyproj/cad\_gpfs\_ingest.py
* /root/pyproj/cad\_gpfs\_exec.py
* /root/pyproj/cad\_analyze\_vcf.py
* /root/pyproj/meta\_api/metaengine.h
* /root/pyproj/meta\_api/meta\_api.c

## Configuration File - cad\_gpfs\_ingest.json

* Create directories in gpfs file system, if not already exist
  + /rsch1/.GlobalWorkDirectory
  + /rsch1/tmp
* Default mmapplypolicy setting does not need to be changed
* Edit/add search list

{

"gwd": "/rsch1/.GlobalWorkDirectory",

"lwd": "/rsch1/tmp",

"maxfiles": "1000",

"threadlevel": "5",

"nodelist": "localhost",

"debuglevel": "1",

"redis\_hostname": "localhost",

"search\_list": [

{

"fsname": "rsch1",

"fileset": "tmp",

"select": "(lower(NAME) like '%.vcf' or lower(NAME) like '%.dcm')"

}

]

}

## Module Test - meta\_api.c

* redis-cli ping
* cd /root/pyproj/meta\_api
* python test.py
* redis-cli keys \*

## Module Test - cad\_analyze\_vcf.py

* cd /root/pyproj
* Edit vcf.list – provide FQDN of vcf files to be parsed
* python cad\_analyze\_vcf.py

## Module Test – cad\_gpfs\_exec.py

* cd /root/pyproj
* python cad\_gpfs\_exec.py LIST exec.list
* python cad\_gpfs\_exec.py TEST /rsch1
  + echo $? (should return 0)
* python cad\_gpfs\_exec.py TEST /foobar
  + echo $? (should return 1)

## Caddy End-to-end Test

* Python cad\_gpfs\_ingest.py (should list all vcf files found)
* Python cad\_gpfs\_ingest.py 1 (should return files with mtime < 1 day)