Ganglia setup for Hadoop Data Lake

INDEX

Why Monitoring is required:-

Network monitoring is the use of a system that constantly monitors a computer network for slow or failing components and that notifies the network administrator (via email, SMS or other alarms) in case of outages. It is part of network management.

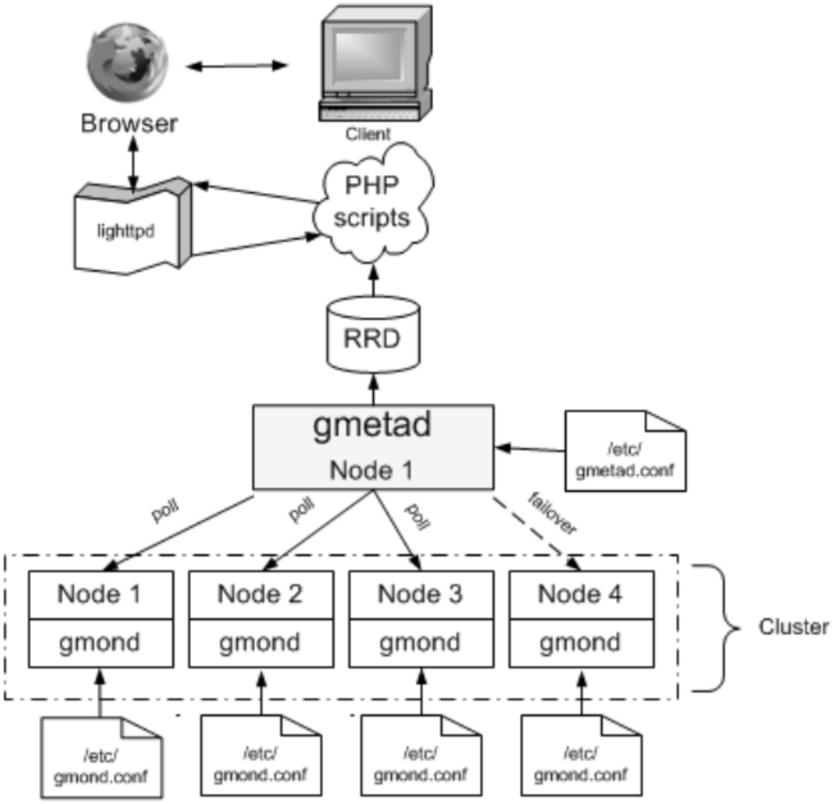
GANGLIA

Ganglia is a scalable distributed monitoring system for high-performance computing systems such as clusters and Grids. It is based on a hierarchical design targeted at federations of clusters. It leverages widely used technologies such as XML for data representation, XDR for compact, portable data transport, and RRDtool for data storage and visualization. It uses carefully engineered data structures and algorithms to achieve very low per-node overheads and high concurrency. The implementation is robust, has been ported to an extensive set of operating systems and processor architectures, and is currently in use on thousands of clusters around the world. It has been used to link clusters across university campuses and around the world and can scale to handle clusters with 2000 nodes.

GANGLIA ARCHITECTURE

The Ganglia architecture (see Figure 1) has the following main components:

- Ganglia monitoring daemon (gmond)
- Ganglia meta daemon (gmetad)
- RRD
- Ganglia frontend



GANGLIA MONITORING DAEMON

gmond is installed on all nodes that need to be monitored. It monitors the changes in host state, listens to other gmond instances over multi-cast/unicast, and is responsible for sending XML, over a TCP connection, to gmetad.. The configuration file will be available at /etc/ganglia/gmond.conf

GANGLIA META DAEMON

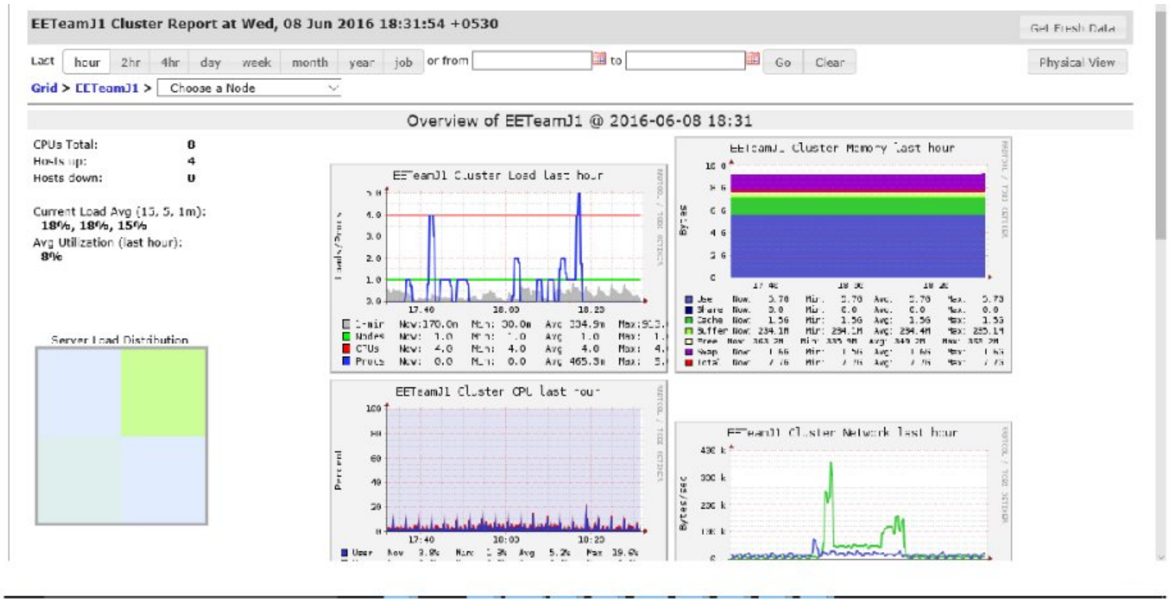
gmetad collects information from multiple gmond or gmetad sources. It saves the information to a local round-robin database, and exports (in XML) a concatenation of all the data sources. Below is a sample configuration. The configuration file will be available at /etc/ganglia/ gmetad.conf

RRD

RRDs are Round-Robin Databases. The data is archived after each hourit is stored in an average form.

GANGLIA PHP WEB FRONTEND

The frontend presents the data collected by gmond and gmetad in a meaningful form.



EETeamJ1 Cluster Report at Wed, 08 Jun 2016 18:31:54 +0530

Overview of EETeamJ1 @ 2016-06-08 18:31

CPUs Total: 8
Hosts up: 4
Hosts down: 0

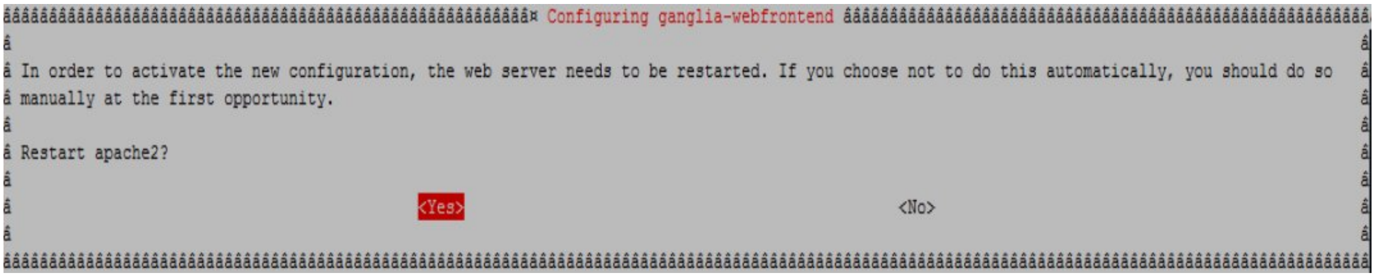
Current Load Avg (15, 5, 1m):
18%, 18%, 15%
Avg Utilization (last hour):
8%

Install Ganglia on Ubuntu 14.04

Open the terminal and run the following command

sudo apt-get install ganglia-monitor rrdtool gmetad ganglia-webfrontend

During the installation you should see similar to the following related to apache2 server restart after installation select Yes and press enter



```
âââââââââââââââââââââââââââââââââââââââââââââââââââââ¤ Configuring ganglia-webfrontend âââââââââââââââââââââââââââââââââââââââââââââââââââââ
â                                                                                                                          â
â In order to activate the new configuration, the web server needs to be restarted. If you choose not to do this automatically, you should do so   â
â manually at the first opportunity.                                                                                       â
â                                                                                                                          â
â Restart apache2?                                                                                                         â
â                                                                                                                          â
â                                   <Yes>                                              <No>                                 â
â                                                                                                                          â
ââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââââ
```

CONFIGURING GANGLIA

Now we have completed the basic installation so we need to configure the required steps

CONFIGURE MASTER NODE

Now copy ganglia configuration file /etc/ganglia-webfrontend/apache.conf to /etc/apache2/sites-enabled/ directory as shown below.

```
$ sudo cp /etc/ganglia-webfrontend/apache.conf /etc/apache2/sites-enabled/ganglia.conf
```

Then edit file /etc/ganglia/gmetad.conf,

```
$ sudo nano /etc/ganglia/gmetad.conf
```

Find the following line and modify as shown below.

```
# data_source "my cluster" 50 192.168.1.101:8649
data_source "EETeamJ1" 20 172.26.60.16:8649 172.26.60.17:8649 172.26.60.18:8649 172.26.60.19:8649
```

As per the above line, the logs will be collected from each node every 20 seconds. Here, we assign a name our client “EETeamJ1” groups. Here 172.26.60.16 is my master node IP address.

Save and close the file.

Edit file /etc/ganglia/gmond.conf,

```
$ sudo nano /etc/ganglia/gmond.conf
```

Find the following sections and modify them with your values.

```
[...]
cluster {
  name = "EETeamJ1"  ## Name assigned to the client groups
  owner = "EETeamJ1"
  latlong = "unspecified"
  url = "unspecified"
}
[...]
udp_send_channel   {
#mcast_join = 239.2.11.71 ## Comment
  host = 172.26.60.16   ## Master node IP address
  port = 8649
  ttl = 1
}
[...]
udp_recv_channel {
  port = 8649
}
/* You can specify as many tcp_accept_channels as you like to share
   an xml description of the state of the cluster */
tcp_accept_channel {
  port = 8649
}
[...]
```

The changes in the above configuration file show that the master node which has IP address 172.26.60.16 will collect data from all nodes on tcp and udp port 8649.

Save and close the file. Then start ganglia-monitor, gmetad and apache services.

```
$ sudo /etc/init.d/ganglia-monitor start
$ sudo /etc/init.d/gmetad start
$ sudo /etc/init.d/apache2 restart
```

now at master node run

```
gstat –a
```

This should hopefully report the current CPU / load stats of all the nodes in the ganglia cluster.

GANGLIA INSTALLATION ON CLIENTS

Install the following package for each client you want to monitor.

```
$ sudo apt-get install ganglia-monitor
```

CONFIGURE CLIENTS

Edit file /etc/ganglia/gmond.conf,

```
$ sudo nano /etc/ganglia/gmond.conf
```

Make the changes as shown below.

```
[...]
cluster {
  name = "EETeamJ1"     ## Cluster name
  owner = "EETeamJ1"
  latlong = "unspecified"
  url = "unspecified"
[...]
udp_send_channel {
  #mcast_join = 239.2.11.71   ## Comment
  host = 172.26.60.16   ## IP address of master node
  port = 8649
  ttl = 1
}
## Comment the whole section
/* You can specify as many udp_recv_channels as you like as well.
udp_recv_channel {
  mcast_join = 239.2.11.71
  port = 8649
  bind = 239.2.11.71
}
*/
tcp_accept_channel {
  port = 8649
}
[...]
```

Save and close the file. Next, restart ganglia-monitor service.

```
$ sudo /etc/init.d/ganglia-monitor restart
```

Hadoop settings (all Hadoop nodes)

To enable monitoring Hadoop with ganglia uncomment the necessary lines in $HADOOP_HOME/conf/hadoop-metrics2.properties file.

```
*.sink.ganglia.class=org.apache.hadoop.metrics2.sink.ganglia.GangliaSink31
*.sink.ganglia.period=10

# default for supportsparse is false
*.sink.ganglia.supportsparse=true
*.sink.ganglia.slope=jvm.metrics.gcCount=zero,jvm.metrics.memHeapUsedM=both
*.sink.ganglia.dmax=jvm.metrics.threadsBlocked=70,jvm
```

Now restart Hadoop cluster and ganglia services.

ACCESS GANGLIA WEB FRONTEND

Now point your web browser with URL http://172.26.60.16/ganglia.