

OpenSource/Linux技術文書



GangliaによるHadoopクラスタの監視

日本ヒューレット・パッカード株式会社
2011年4月5日

目次

[本ドキュメントについて].....	4
システム構成	5
Hadoop クラスタの準備.....	6
Ganglia のインストール (NN)	6
Ganglia gmond のインストール (DN)	7
Ganglia の設定 (NN)	7
Ganglia の設定 (DN)	9
Ganglia による Hadoop クラスタの監視 (NN)	12

図表目次

図 1. Cloudera Distribution for Hadoop on HP ProLiant SL6500 のシステム構成例 .	5
図 2. Ganglia による Hadoop クラスタの監視の様子.....	12
図 3. Ganglia による NameNode と DataNode の CPU 負荷状況とメモリ利用率の監視 ...	13
図 4. Ganglia による Hadoop クラスタのハードウェア・リソースと OS の状態監視 ...	13
図 5. Ganglia の GUI 画面から DataNode の 1 つを選択する.....	14
図 6. DataNode の HDFS の状況「dfs.datanode」を確認する.....	14

[本ドキュメントについて]

- 本ドキュメントでは、NameNode サーバーのみでの作業を (NN)、DataNode サーバーのみでの作業を (DN)、NameNode サーバーと DataNode サーバー両方での作業を (NN, DN) と記すことにします。
 - 例 1)
「ファイルをコピーします。(NN)」と記載してあるものは、NameNode だけでファイルをコピーするという意味になります。
 - 例 2)
「rpm コマンドでパッケージをインストールします (NN, DN)」と記載してあるものは、NameNode と DataNode の両方で rpm コマンドを使ってインストールを行うという意味になります。
- コマンドラインでの入力が長く紙面の都合で折り返して記載する場合は、下記のように「¥」記号を挿入して複数行にわたって記載しています。複数行にわたって記載されていても実際には 1 行で入力するものは、その記述の最後に「(実際には 1 行で入力)」を挿入しています。
 - 例 3)

```
# alternatives --install /etc/hadoop-0.20/conf hadoop-0.20-conf ¥  
/etc/hadoop-0.20/conf.hp001 20 (実際には 1 行で入力)
```
- 本ドキュメントの内容については充分チェックをしておりますが、その正確性を保証する物ではありません。また、将来、予告なしに変更することがあります。
- 本ドキュメントの使用で生じるいかなる結果も利用者の責任となります。日本ヒューレット・パッカード株式会社は、本ドキュメントの内容に一切の責任を負いません。
- 本ドキュメントの技術情報は、ハードウェア構成、OS、アプリケーションなど使用環境により大幅に数値が変化する場合がありますので、十分なテストを個別に実施されることを強くお勧め致します。
- 本ドキュメント内で表示・記載されている会社名・サービス名・商品名等は各社の商標又は登録商標です。
- 本ドキュメントで提供する資料は、日本の著作権法、条約及び他国の著作権法にいう著作権により保護されています。

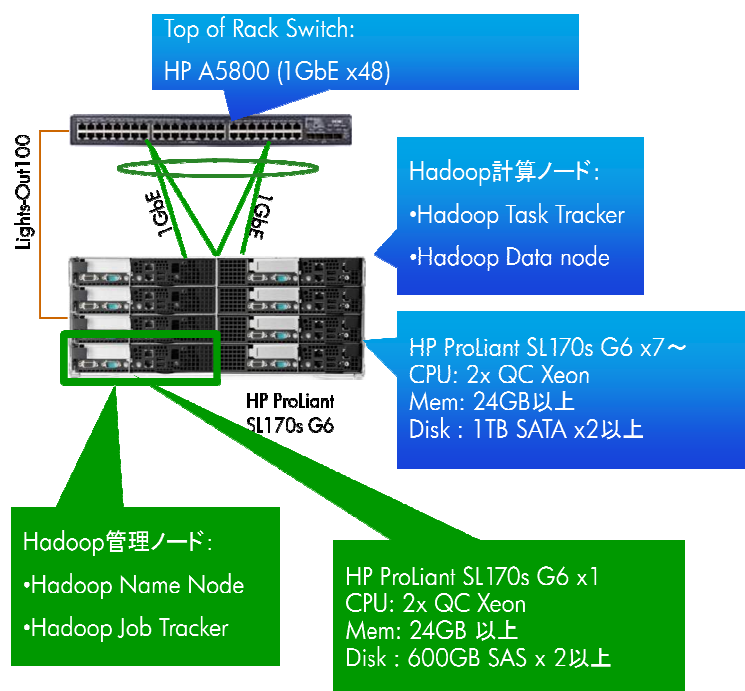
本ドキュメントはHP ProLiantサーバー上で、Cloudera distribution for Hadoop 3 beta 4 (通称CDH3b4)が構築されたシステムにGangliaをインストールし、Hadoopクラスターシステムを監視するためのガイドです。CDH3b4は現時点でベータリリースの最新版です。CDHのベータリリースの本番商用利用については検討が必要となりますのでご注意ください。

システム構成

以下にCDHをインストールする環境を示します。

ハードウェア	: HP ProLiant SL6500 (Server: HP ProLiant SE2170s x8)
OS	: Red Hat Enterprise Linux 5.6 x86-64 (NameNodeおよびDataNode)
JDK	: 1.6u24 x64 (Oracle社が提供するパッケージを利用)
CDH	: CDH3b4 (Red Hat系OSに対応したRPMパッケージを利用)
Ganglia	: 3.1.7

以下にハードウェア外観を示します。今回の構成ではNameNodeの可用性は考慮していないため、NameNodeの障害時のデータロスが発生するSPOFが存在する点にご注意ください。



- Proof Of Concept 構成
- Hadoop の Name Node は 1 台構成
- Name Node は、Job Tracker を兼用
- Name Node の可用性はなし
- システムの可用性は HDFS のレプリカで担保できるが、ロスしてもよいレベルの動作確認用途
- 容量を確保するため、SATA のディスクを搭載
- 障害による交換頻度を下げたい場合は、SAS を検討
- Top of Rack スイッチはギガビットのポート数を多数もつものを用意
- HP A5800 では 48 ポートが利用可能

図 1. Cloudera Distribution for Hadoop on HP ProLiant SL6500 のシステム構成例

Hadoop クラスタの準備

本ドキュメントに記載しているGangliaの設定手順を行う前に、Hadoopクラスタのインストールと事前準備を済ませておく必要があります。Hadoopクラスタの設定については、別紙「Cloudera Distribution for Hadoopインストール手順及び利用例」を参照してセットアップを完了させておいてください。

Ganglia のインストール (NN)

Gangliaの本体のtarballであるganglia-3.1.7.tar.gzを、以下URLに示すダウンロードサイトから入手します。

```
http://downloads.sourceforge.net/project/ganglia/ganglia%20monitoring%20core/3.1.7/ganglia-3.1.7.tar.gz?r=http%3A%2F%2Fsourceforge.net%2Fprojects%2Fganglia%2Ffiles%2Fganglia%2520monitoring%2520core%2F3.1.7%2F&ts=1301892228&use_mirror=iweb
```

Gangliaの関連パッケージを、以下URLに示すダウンロードサイトから入手します。

```
http://apt.sw.be/redhat/el5/en/x86_64/rpmforge/RPMS/rrdtool-1.4.3-3.el5.rf.x86_64.rpm
http://apt.sw.be/redhat/el5/en/x86_64/rpmforge/RPMS/rrdtool-devel-1.4.3-1.el5.rf.x86_64.rpm
http://download.fedora.redhat.com/pub/epel/5/x86_64/libconfuse-2.5-4.el5.x86_64.rpm
http://download.fedora.redhat.com/pub/epel/5/x86_64/libconfuse-devel-2.5-4.el5.x86_64.rpm
http://apt.sw.be/redhat/el5/en/x86_64/rpmforge/RPMS/perl-rrdtool-1.4.4-1.el5.rf.x86_64.rpm
http://download.fedora.redhat.com/pub/epel/5/x86_64/lua-5.1.4-4.el5.x86_64.rpm
```

以下のRPMパッケージをRHEL5.6のDVDから事前にインストールしておきます。(NN)

```
# rpm -vhi php-gd-5.1.6-27.el5_5.3.x86_64.rpm
# rpm -vhi pcre-devel-6.6-6.el5.x86_64.rpm
# rpm -vhi apr-devel-1.2.7-11.el5_5.3.x86_64.rpm
```

上記URLから入手したRPMパッケージをインストールします。(NN)

```
# rpm -vhi libconfuse-devel-2.5-4.el5.x86_64.rpm
# rpm -vhi libconfuse-2.5-4.el5.x86_64.rpm
# rpm -vhi rrdtool-1.4.3-3.el5.rf.x86_64.rpm
# rpm -vhi rrdtool-devel-1.4.3-1.el5.rf.x86_64.rpm
# rpm -vhi perl-rrdtool-1.4.3-1.el5.rf.x86_64.rpm
# rpm -vhi lua-5.1.4-4.el5.x86_64.rpm
```

その他にもexpat-devel、freetype-devel、gcc-c++、libart_lgpl-devel、libpng-devel、python-devel、rpm-build等が必要になりますので、インストールされていない場合は、RHEL5.6のDVDからインストールしておきます。(NN)

上記URLから入手したGangliaの本体のtarballであるganglia-3.1.7.tar.gzからRPMパッケージをビルドします。入手したtarballはNameNodeの/rootに置いてあるものとします。(NN)

```
# rpmbuild -ta --target noarch,x86_64 /root/ganglia-3.1.7.tar.gz
```

GangliaのRPMパッケージが正常にビルドされているかを確認します。

```
# cd /usr/src/redhat/RPMS/x86_64/
# ls -l *ganglia*
ganglia-debuginfo-3.1.7-1.x86_64.rpm
ganglia-devel-3.1.7-1.x86_64.rpm
ganglia-gmetad-3.1.7-1.x86_64.rpm
ganglia-gmond-3.1.7-1.x86_64.rpm
ganglia-gmond-modules-python-3.1.7-1.x86_64.rpm
libganglia-3_1_0-3.1.7-1.x86_64.rpm
```

```
# cd /usr/src/redhat/RPMS/noarch/
# ls -l *ganglia*
ganglia-web-3.1.7-1.noarch.rpm
```

rpmコマンドでGangliaをNameNodeにインストールします。(NN)

```
# cd /usr/src/redhat/RPMS/x86_64/
# rpm -vhi      ganglia-gmetad-3.1.7-1.x86_64.rpm ¥
                ../noarch/ganglia-web-3.1.7-1.noarch.rpm ¥
                libganglia-3_1_0-3.1.7-1.x86_64.rpm   (実際には1行で入力)
# rpm -vhi      ganglia-gmond-3.1.7-1.x86_64.rpm ¥
                ganglia-gmond-modules-python-3.1.7-1.x86_64.rpm (実際には1行で入力)
```

Ganglia gmond のインストール (DN)

Ganglia の gmond を Hadoop の DataNode にインストールします。入手した libconfuse 及び NameNode 上でビルドした ganglia-gmond と ganglia-gmond-modules-python、libganglia を DataNode にコピーし、インストールを行います。(DN)

```
# rpm -vhi      libconfuse-2.5-4.el5.x86_64.rpm
# rpm -vhi      ganglia-gmond-3.1.7-1.x86_64.rpm ¥
                ganglia-gmond-modules-python-3.1.7-1.x86_64.rpm ¥
                libganglia-3_1_0-3.1.7-1.x86_64.rpm   (実際には1行で入力)
```

Ganglia の設定 (NN)

```
# cd /etc/ganglia/
# cp gmetad.conf gmetad.conf.org
# vi gmetad.conf
...
#### Ganglia management node
data_source "Hadoop Name Node" hd01
#### Describe the one of managed nodes.
data_source "Hadoop Data Node" hd02
data_source "Hadoop Data Node" hd03
```

```
data_source "Hadoop Data Node" hd04
data_source "Hadoop Data Node" hd05
data_source "Hadoop Data Node" hd06
data_source "Hadoop Data Node" hd07
data_source "Hadoop Data Node" hd08
```

...

```
gridname "Hadoop Cluster"
```

...

Ganglia の gmetad サービスを起動します。(NN)

```
# chkconfig --list gmetad
# service gmetad start
# /etc/init.d/gmetad status
gmetad (pid 7984) を実行中...
#
```

NameNode の Ganglia 設定ファイル gmond.conf を編集します。ここでは NameNode における Ganglia のマルチキャストアドレスとして 239.2.11.**72** を使用します。(NN)

```
# cd /etc/ganglia/
# cp gmond.conf gmond.conf.org
# vi gmond.conf
```

...

```
cluster {
    name = "Hadoop Name Node"
    owner = "unspecified"
    latlong = "unspecified"
    url = "unspecified"
}
```

...

```
udp_send_channel {
    bind_hostname = yes          # Highly recommended, soon to be default.
                                # This option tells gmond to use a source address
                                # that resolves to the machine's hostname. Without
                                # this, the metrics may appear to come from any
                                # interface and the DNS names associated with
                                # those IPs will be used to create the RRDs.

    mcast_if = eth0
    mcast_join = 239.2.11.72
    port = 8649
    ttl = 1
}
```

...

```
udp_recv_channel {
    mcast_join = 239.2.11.72
    mcast_if = eth0
    port = 8649
    bind = 239.2.11.72
}
```



```
}  
...
```

NameNode 上の/etc/hadoop-0.02/conf.hp001 ディレクトリにある `hadoop-metrics.properties` を編集します。先に指定した NameNode でのマルチキャストアドレス 239.2.11.72 を指定します。(NN)

```
# cd /etc/hadoop-0.20/  
# cd conf.hp001  
# cp hadoop-metrics.properties hadoop-metrics.properties.org  
# vi hadoop-metrics.properties  
...  
dfs.servers=239.2.11.72:8649  
...  
mapred.servers=239.2.11.72:8649  
...  
jvm.servers=239.2.11.72:8649  
...  
fairscheduler.servers=239.2.11.72:8649  
...
```

NameNode 上でマルチキャストアドレスが利用するネットワークインタフェースを明示的に指定します。(NN)

```
# cd /etc/sysconfig/network-scripts/  
# vi route-eth0  
239.2.11.72 dev eth0
```

NameNode 上の Ganglia の gmond サービスを起動します。(NN)

```
# chkconfig gmond --list  
# service gmond start  
# service gmond status
```

NameNode 上で Network サービスを再起動します。(NN)

```
# service network restart  
# netstat -rn
```

Kernel IP routing table

Destination	Gateway	Genmask	Flags	MSS Window	irrt	Iface
239.2.11.72	0.0.0.0	255.255.255.255	UH	0 0	0	eth0
192.168.1.0	0.0.0.0	255.255.255.0	U	0 0	0	eth1
192.168.122.0	0.0.0.0	255.255.255.0	U	0 0	0	virbr0
169.254.0.0	0.0.0.0	255.255.0.0	U	0 0	0	eth1
172.16.0.0	0.0.0.0	255.255.0.0	U	0 0	0	eth0

Ganglia の設定 (DN)

DataNode の Ganglia 設定ファイル `gmond.conf` を編集します。ここでは DataNode における

Ganglia のマルチキャストアドレスとして 239.2.11.71 を使用します。(NN) (DN)

```
# cd /etc/ganglia/
# cp gmond.conf gmond.conf.org
# vi gmond.conf
...
cluster {
    name = "Hadoop Data Node"
    owner = "unspecified"
    latlong = "unspecified"
    url = "unspecified"
}
...
udp_send_channel {
    bind_hostname = yes      # Highly recommended, soon to be default.
                            # This option tells gmond to use a source address
                            # that resolves to the machine's hostname. Without
                            # this, the metrics may appear to come from any
                            # interface and the DNS names associated with
                            # those IPs will be used to create the RRDs.

    mcast_if = eth0
    mcast_join = 239.2.11.71
    port = 8649
    ttl = 1
}
...
udp_rcv_channel {
    mcast_join = 239.2.11.71
    mcast_if = eth0
    port = 8649
    bind = 239.2.11.71
}
...
```

DataNode 上の/etc/hadoop-0.20/conf.hp001 ディレクトリにある hadoop-metrics.properties を編集します。先に指定した DataNode でのマルチキャストアドレス 239.2.11.71 を指定します。(DN)

```
# cd /etc/hadoop-0.20/
# cd conf.hp001
# cp hadoop-metrics.properties hadoop-metrics.properties.org
# vi hadoop-metrics.properties
...
dfs.servers=239.2.11.71:8649
...
mapred.servers=239.2.11.71:8649
...
jvm.servers=239.2.11.71:8649
```

```
...
fairscheduler.servers=239.2.11.71:8649
...
```

DataNode 上で Ganglia の gmond サービスを起動します。(DN)

```
# chkconfig gmond --list
# service gmond start
# service gmond status
```

DataNode 上でマルチキャストアドレスが利用するネットワークインタフェースを明示的に指定します。(DN)

```
# cd /etc/sysconfig/network-scripts/
# vi route-eth0
239.2.11.71 dev eth0
```

NameNode 上で Network サービス、hadoop の各種サービス、Ganglia のサービスを再起動します。(NN)

```
# service network restart
# netstat -rn
Kernel IP routing table
Destination Gateway Genmask Flags MSS Window irtt Iface
239.2.11.72 0.0.0.0 255.255.255.255 UH 0 0 0 eth0
192.168.1.0 0.0.0.0 255.255.255.0 U 0 0 0 eth1
192.168.122.0 0.0.0.0 255.255.255.0 U 0 0 0 virbr0
169.254.0.0 0.0.0.0 255.255.0.0 U 0 0 0 eth1
172.16.0.0 0.0.0.0 255.255.0.0 U 0 0 0 eth0
```

```
# service hadoop-0.20-namenode restart
# service hadoop-0.20-jobtracker restart
# service gmetad restart
# service gmond restart
```

DataNode において、Network サービス、hadoop の各種サービス、Ganglia のサービスを再起動します。(DN)

```
# service network restart
# netstat -rn
Kernel IP routing table
Destination Gateway Genmask Flags MSS Window irtt Iface
239.2.11.71 0.0.0.0 255.255.255.255 UH 0 0 0 eth0
192.168.122.0 0.0.0.0 255.255.255.0 U 0 0 0 virbr0
169.254.0.0 0.0.0.0 255.255.0.0 U 0 0 0 eth0
172.16.0.0 0.0.0.0 255.255.0.0 U 0 0 0 eth0
```

```
# service hadoop-0.20-datanode restart
# service hadoop-0.20-tasktracker restart
# service gmond restart
```

GangliaによるHadoop クラスタの監視(NN)

NameNode で Apache サービスが起動しているか確認します。起動していない場合は Apache サービスを起動します。

```
# service httpd status  
# service httpd restart
```

Ganglia によって、Hadoop 分散ファイルシステム HDFS の状態を確認します。Web ブラウザで NameNode にアクセスします。(NN)

```
# firefox http://hd01/ganglia
```

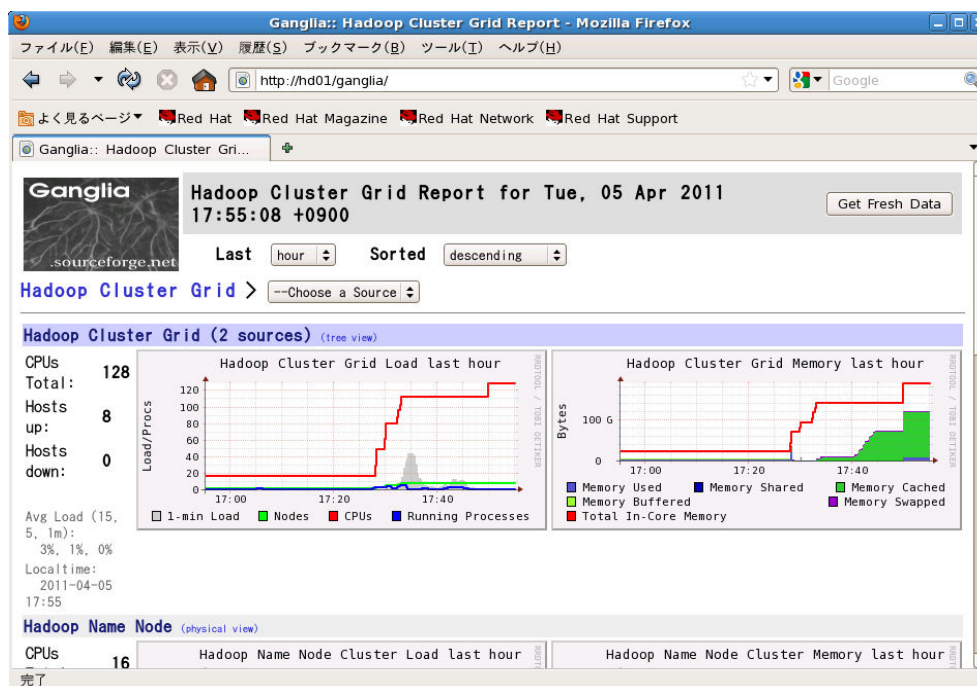


図 2. Ganglia による Hadoop クラスタの監視の様子

NameNode の状態を確認します。Ganglia の画面から「Hadoop Cluster Grid」をクリックします。この Grid 名「Hadoop Cluster」は、NameNode の/etc/ganglia/gmetad.conf ファイルで定義した gridname パラメータの文字列です。「Hadoop Cluster Grid」をクリックすると、NameNode と DataNode の CPU 負荷状況とメモリ使用率を確認することができます。

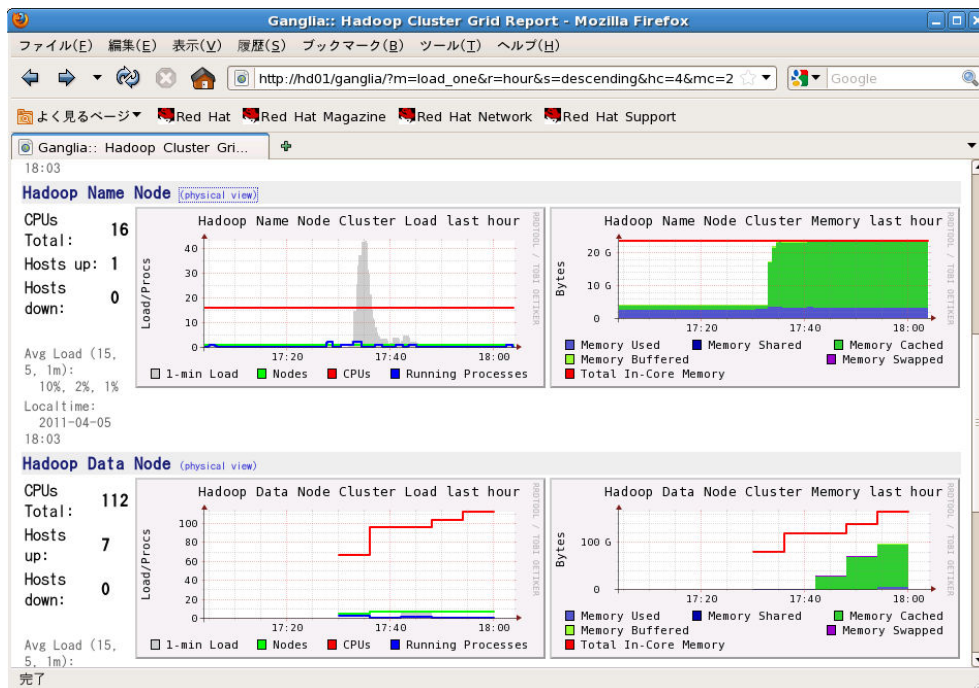


図 3. GangliaによるNameNodeとDataNodeのCPU負荷状況とメモリ利用率の監視

「Hadoop Name Node」の右横に記載されている「(physical view)」をクリックし、さらにNameNodeのホスト名「hd01.jpn.linux.hp.com」をクリックします。すると、NameNodeのハードウェア構成及びインストールされているLinux OSのカーネルバージョン、アーキテクチャ、起動時間等を把握できます。DataNodeについても同様です。

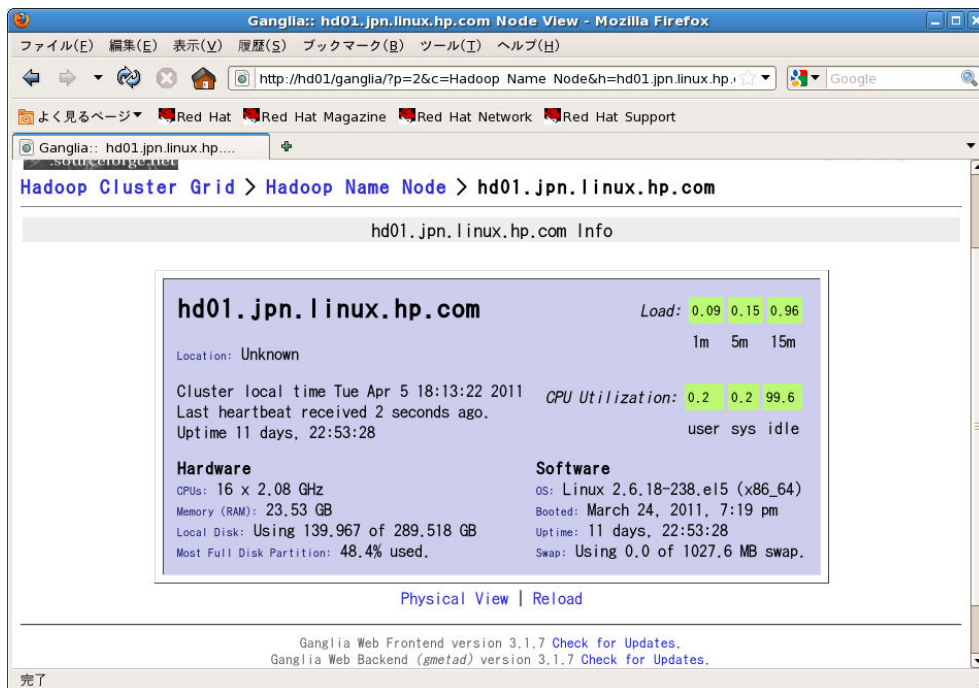


図 4. GangliaによるHadoopクラスタのハードウェア・リソースとOSの状態監視

「Hadoop Cluster Grid」をクリックし、「Hadoop Data Node」をクリックすると、Gangliaに登録したDataNodeの状況を把握できます。ここでは、hd04.jpn.linux.hp.comを選択してみます。

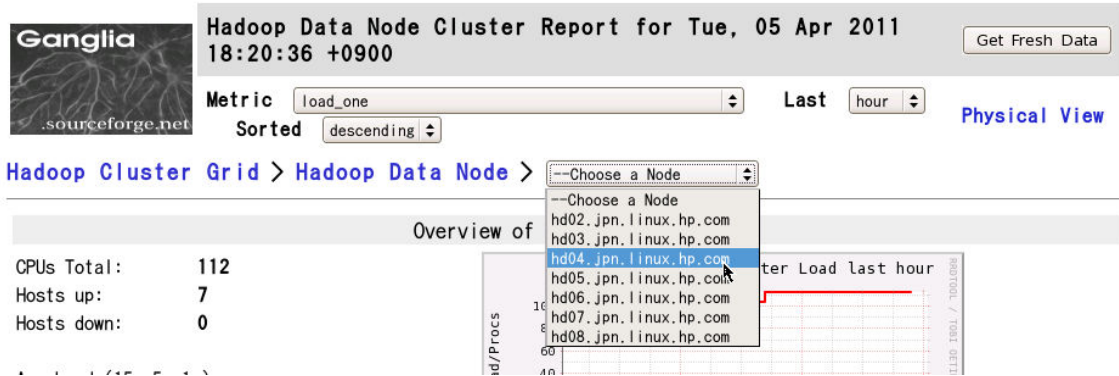


図 5. Ganglia の GUI 画面から DataNode の 1 つを選択する

DataNode の Hadoop 分散ファイルシステムの使用率などを時系列で表示することが可能です。

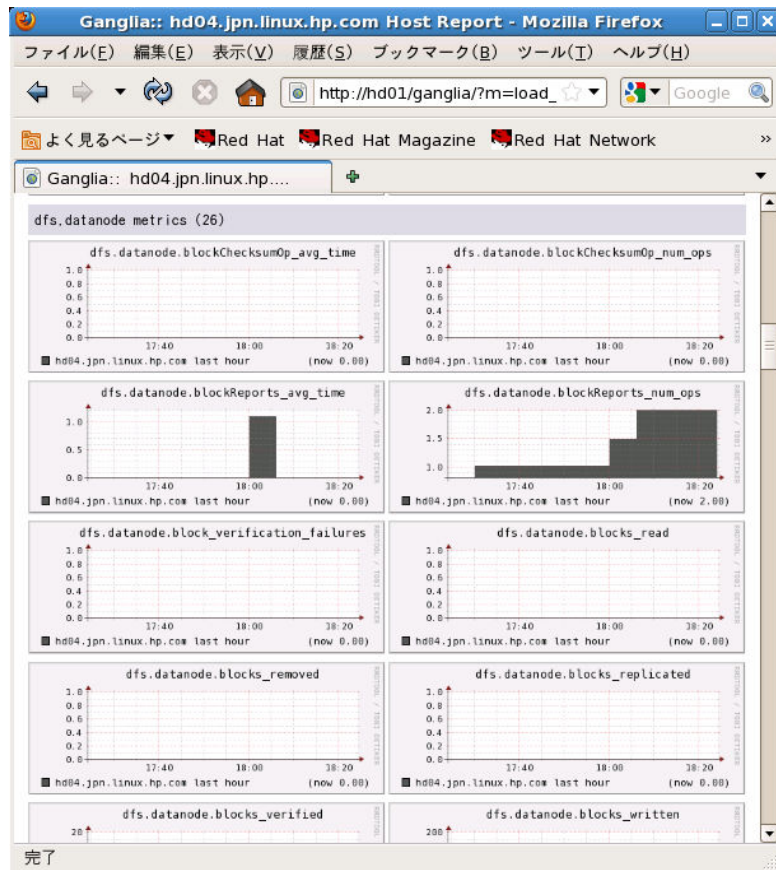


図 6. DataNode の HDFS の状況「dfs.datanode」を確認する

以上