



SCore クラスタ構築法

SCore 5.0版用

2001.03.19版

開発部会
アプリケーション・技術支援部会

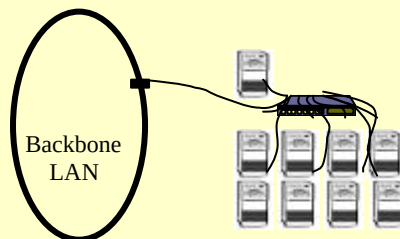
本資料は、会員向けです!!!

本資料に関するコメント、最新版の情報入手は
devel@pccluster.orgまで



SCoreに必要な機材

- ◆ サーバホスト
 - ◆ ファイルサーバ
 - ◆ 各種デーモンの実行
 - ◆ 計算ホストを兼ねることも可能
- ◆ 計算ホスト
 - ◆ ディスク
- ◆ ネットワーク
 - ◆ Ethernet
 - ◆ EEPRO100を推奨
 - ◆ Myrinet-2000
 - ◆ Fibreを推奨





NIC (Network Interface Card)

- ◆ 新情報処理開発機構で実績のあるNIC
 - ◆ 100Mbps Ethernet
 - ◆ EEPRO100
 - ◆ Tulip
 - ◆ 3C905B
 - ◆ ただし、Network Trunking機能はEEPRO100でしか動作確認していない
 - ◆ 1Gbps Ethernet
 - ◆ Syskonnect SK-9843
- ◆ Linux driverがあるからと言って安定かつ性能が出るとは限らない

02/03/20

3



ホスト名およびIPアドレスを定義するもの

- ◆ /etc/sysconfig/network
 - ◆ NETWORKING=yes
 - ◆ HOSTNAME=server.pccluster.org
- ◆ /etc/sysconfig/network-scripts/ifcfg-eth0
 - ◆ BROADCAST=163.220.0.0
 - ◆ IPADDR=163.220.1.100
 - ◆ NETMASK=255.255.248.0
 - ◆ NETWORK=163.220.0.0
 - ◆ ONBOOT=yes
- ◆ /etc/nsswitch.conf
 - ◆ データベース検索順序が定義されている
- ◆ hostnameコマンドでホスト名が表示される。システム起動時、/etc/sysconfig/networkで規定された名前を使ってhostnameコマンドで設定されている。

02/03/20

4





NIS (Network Information name Service)

- ◆ 名前を管理するシステム
 - ◆ パスワード、ホスト名とIPアドレス、グループ名、
- ◆ NIS ドメイン名
 - ◆ NISドメイン名毎にNISサーバがある
- ◆ コマンド群
 - ◆ ypswitch
 - ◆ 使用しているNISサーバを表示
 - ◆ ypmatch
 - ◆ キーワードに対応する値の表示
 - ◆ ypcat
 - ◆ NIS データベースに登録されているエントリ全て表示
- ◆ ファイル群
 - ◆ /etc/sysconfig/network /etc/passwd, /etc/shadow
 - ◆ NISDOMAIN=trc.info /etc/hosts
 - ◆ /var/yp/ /etc/group

02/03/20

5



ドメイン、ドメイン名、DNS

- ◆ ドメイン、ドメイン名、FQDN (Fully Qualified Domain Name)
 - ◆ インターネット上のサイトの名前。ドメインとは、ネットワーク環境でのひとまとまりの管理対象を指す。www.pccluster.org
 - ◆ host名からドメイン名まで全てを記述した名前をFQDNと呼ぶ
- ◆ DNS (Domain Name System)
 - ◆ ホスト名から対応するIPアドレス、IPアドレスから対応するホスト名を得るためのサービスを提供するためのシステム
- ◆ コマンド群
 - ◆ nslookup (今後サポートされなくなる)
 - ◆ host
- ◆ ファイル群
 - ◆ /etc/resolve.conf DNSサーバのリスト

02/03/20

6





NFS

- ◆ ファイル群
 - ◆ /etc/exports
 - ◆ アクセス権限を規定している
 - ◆ /etc/fstab
 - ◆ ファイルシステムの一覧
 - ◆ /etc/mntab
 - ◆ マウントされたファイルシステムの一覧
 - ◆ /var/lib/nfs/xtab
 - ◆ アクセス権限を保存しているファイル。これは、`exportfs -r`することによって/etc/exportsと内容の同期が取られる。
 - ◆ /var/lib/nfs/rmtab
 - ◆ リモートにマウントしているホストメイトマウント場所のリスト
 - ◆ /var/lib/nfs/etab
- コマンド群
 - exportfs
 - アクセス権限の更新



有用なコマンド (ファイル関係)

- ◆ find `find . -name 'score*' -print`
 - ◆ ファイルを調べる
- ◆ grep `find . -name 'score*' -exec grep -i myrinet {} {} \;`
 - ◆ ファイルの中身を調べる
- ◆ rgrep
 - ◆ ディレクトリの下までgrep可能
- ◆ du
 - ◆ ディスク使用量
- ◆ df
 - ◆ 空きディスク領域表示



有用なコマンド (プロセス関係)

- ◆ ps
 - ◆ プロセス状態の表示
 - ◆ `ps -axwl | grep scout`
 - ◆ W オプションを指定することにより、一行の長さ以上に情報を出すことが出来る。これによりコマンドアークメントもわかる
- ◆ top
 - ◆ プロセス実行状態が表示される。
- ◆ kill
 - ◆ プロセスにシグナルを送る。強制終了させたい場合は、
`kill -KILL <process ID>`あるいは
`kill -QUIT <process ID>`



有用なコマンド (ネットワーク関係)

- ◆ ping
 - ◆ TCP/IP ネットワークにおいて、IP パケットが通信先まで届いているかを調べるためのコマンド。ICMP (Internet Control Message Protocol) の echo コマンドを使って実現されている。
- ◆ arp
 - ◆ IP アドレスと Ethernet アドレスの対を表示、設定するコマンド。arp は、TCP/IP において、IP アドレスから Ethernet アドレスを求めるためのプロトコル名。
- ◆ ifconfig
 - ◆ ネットワークインターフェイス
状態表示、設定用コマンド。

```
eth0  Link encap:Ethernet HWaddr 00:02:B3:06:0A:1E
       inet addr:163.220.7.109 Bcast:163.220.0.0 Mask:255.255.248.0
       UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
       RX packets:197193 errors:0 dropped:0 overruns:0 frame:0
       TX packets:39174 errors:0 dropped:0 overruns:0 carrier:0
       collisions:51604 txqueuelen:100
       Interrupt:11 Base address:0x1800
```



ホスト名

- ◆ 計算ホスト名は、英文字の後にホスト番号をつけると良い
- ◆ サーバホスト名も計算ホスト名も FQDN(Fully Qualified Domain Name) としましょう。

eg. comp0.pccluster.org, comp1.pccluster.org,
comp2.pccluster.org



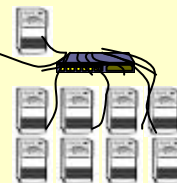
必要なネットワーク情報

サーバIPアドレス
計算ホストIPアドレス
DHCP IPアドレス
ネットマスク
デフォルト・ゲートウェイIPアドレス
ドメイン名
NISドメイン名

注意点！！

- ホスト名はすべてFQDNにすべき。
- DNS, NIS, /etc/hosts, hostnameで統一が取れていることを確認すること。
- DHCP IPアドレスは計算ホストIPアドレスとは別に確保すること。
- DHCP IPアドレスにもホスト名をつけること

Backbone
LAN





クラスタ実演

server IP Address	192.168.0.2
comp0 to comp7 IP Addresses	192.168.0.100 -- 192.168.55.107
Netmask	255.255.255.0
Default Gateway IP address	192.168.0.1
Primary Nameserver IP address	192.168.0.1
Domain Name	pccluster.org
Server Hostname	server.pcccluster.org
NIS domain name	score

02/03/20

13

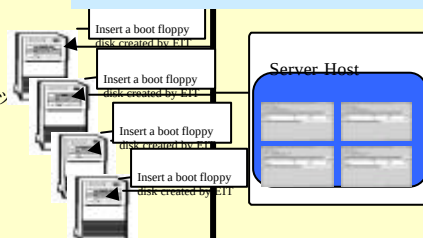


インストールの流れ

- ◆ EITはanacondaを改造して使用しています
- ◆ サーバホストのネットワーク設定
 - ◆ サーバ側にanaconda kickstart fileを作成
- ◆ EITを使ってブートフロッピーディスクを作成
 - ◆ ブートフロッピーは計算ホストに依存しない
- ◆ 計算ホストのインストール
 - ◆ ブートフロッピーディスクを使って起動
 - ◆ EITDからIPアドレスを取得。
 - ◆ IPアドレスをアサインする。
 - ◆ サーバホストのファイルシステムをマウントし、キックスタートファイルを取得する
 - ◆ SCore用に改良した Red Hat anaconda によるネットワークインストール。
 - ◆ SCore用の設定(postprocessing)
 - ◆ リブート
- ◆ サーバホストの SCore設定

ここではネットワークはEthernet
だけを仮定します

サーバホスト側ディレクトリ
/opt/score/etc/ndconf/
ここに各計算ホストの Ethernet address 情報
が格納される。
/opt/score/setup/RedHat/instimage/compconf/
ここに各計算ホスト用のキックスタートファ
イルが置かれる。ファイル名はMACアドレス



02/03/20

14





Powered by Red Hat

サーバの設定

- ◆ Red Hat 7.2をno firewall指定でフルインストール
- ◆ ホスト名を入力するところで、FQDNを入れること！！
- ◆ DNS、NISが運用されているなら、その設定
- ◆ /opt/ディレクトリ下に 500MBytes以上の空きスペースが必要
- ◆ SCore 5.0 CD-ROMをCD-ROMドライブに装着し、

```
# mount /mnt/cdrom
```

```
# cd /mnt/cdrom
```

```
# ./Install
```

02/03/20

15



Powered by Red Hat

サーバ設定の確認

hostnameコマンドを実行してホスト名がFQDNになっていることを確認しましょう！

- ◆ DNSが設定されているなら、ホスト名からIPアドレス、IPアドレスからホスト名が引けるか確認する。

```
# host <ホスト名>
# host <IPアドレス>
```
- ◆ NISが設定されていないなら、サーバ上でNISを立ち上げる。計算ホストのアカウントはこのNISで行うため。
 - ◆ ユーザアカウントを作成する
 - ◆ /etc/hostsに計算ホストおよびDHCPホストを登録する。

```
# configNIS server <NIS domain name>
```
- ◆ NISが正常に稼動しているかどうかのテスト。

02/03/20

16





EITの起動

- ◆ ルートになって、eitを起動する
 - ◆ # /opt/score/bin/eit



02/03/20

17



Keyboard Selection

- ◆ キーボードタイプを決めてください。
- ◆ Bugs
 - ◆ キーボードタイプが、サーバ側のX window systemにも反映されてしまいます。サーバ側は、X window systemを立ち上げ直せば元に戻ります。これはanacondaの問題ですが、どこを変更すれば、これを止めることができるかは不明



02/03/20

18





Powered by VMware

Network Configuration

- ◆ ネットワーク情報の設定です。サーバの設定情報を取り出しています。
- ◆ Bugs
 - ◆ Server NameとDisplayは変更しないでください

The screenshot shows the 'Network Configuration' window of the PCCC Score 5.0 Easy Installation Tool. The window contains the following fields and values:

Field	Value
Server Name	server.pccluster.org
Domain Name	pccluster.org
Netmask	255.255.255.0
Gateway	192.168.0.1
NIS	score
Name Server	192.168.0.1
Mount Point	/mnt/cdrom/
Display	server.pccluster.org

Buttons: Back, Next, Cancel

02/03/20

19



Powered by VMware

NFS Configuration

- ◆ 計算ホスト側のNFSマウントポイントの指定を行います。

The screenshot shows the 'NFS Configuration' window of the PCCC Score 5.0 Easy Installation Tool. The window contains a table for configuring NFS mounts:

Host	Dir	Mount Point
server.pccluster.org	/home	/home
Host	Dir	Mount Point
Host	Dir	Mount Point
Host	Dir	Mount Point
Host	Dir	Mount Point
Host	Dir	Mount Point
Host	Dir	Mount Point
Host	Dir	Mount Point
Host	Dir	Mount Point
Host	Dir	Mount Point

Buttons: Back, Next, Cancel

02/03/20

20





Disk Partition

- ◆ 計算ホストのディスクパーティション設定です。
- ◆ Disk Partition Planでディスクサイズをクリックすると、そのサイズのディスクパーティションの推奨が表示されます。



02/03/20

21



計算ホストの指定

ホストの数

2 way SMPなら
2を指定

計算ホスト名のprefix

桁数、ここでは1を指定する

Addボタンをクリックすると
このようなウインドウが表示される

02/03/20

22





グループ指定



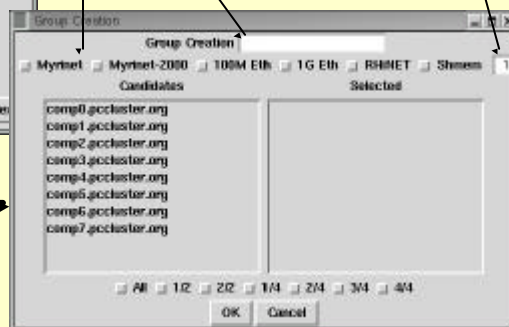
計算ホストの指定で登録された計算ホスト名がCandidates欄に表示されます。各計算ホスト名にマウスをもっていきクリックすると、Select欄にコピーされます。Select欄の計算ホスト名にマウスをもっていきクリックすると削除されます。

AllボタンをクリックするとCandidates欄の全ての計算ホスト名がSelected欄にコピーされます。
1/2は最初の半分、2/2は後半の半分がコピーされます。以下同様

使用するネットワークタイプを指定。

グループ名を入力

SMPの場合、
プロセッサ数
を入力

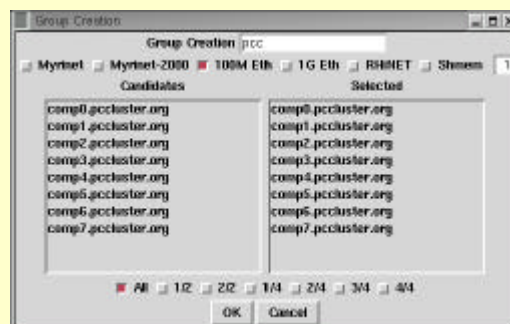


02/03/20

23



グループ指定その2



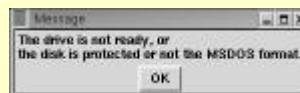
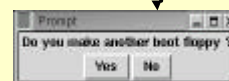
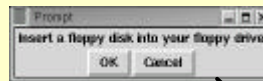
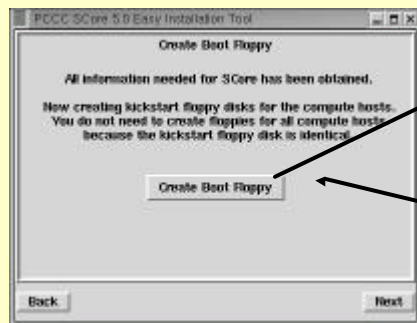
02/03/20

24





ブートフロップの作成



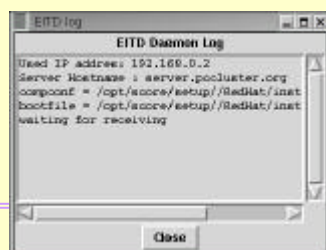
02/03/20

25



Cluster Setup

- ◆ この画面が現れたら、インストールする計算ホストにブートフロッピーを入れて電源を投入してください
- ◆ 右下の画面のようにホスト番号とNICアドレス、IPアドレスが表示されたら、次の計算ホストにブートフロッピーを入れて電源投入して大丈夫です



02/03/20

26





Powered by Xilinx

計算ホスト側の様子

```

NET4: Unix domain sockets 1.0/SMP for Linux NET4.0.
RAMDISK: Compressed image found at block 0
Freeing initrd memory: 724k freed
VFS: Mounted root (ext2 filesystem).
Using ERTS feature
mounting /proc filesystem... done
Testing .....
No dhcp_server specified. Used Broadcast.
setupNetwork cannot set the gateway address
done
NFS mount 192.168.0.2:/opt/score/setup//RedHat/instimage/ /mnt/runtime
NFS mount 192.168.0.2:/opt/score/setup//RedHat/instimage/modules /modules
Testing .....
NFS mount 192.168.0.2:/mnt/cdrom/ /mnt/source2
Mac Address = 00:50:56:40:00:93
LD_LIBRARY_PATH = /lib:/usr/lib:/usr/X11R6/lib
DISPLAY = 192.168.0.2:0.0

PCCC Easy Installation Tool Version 5.0
Copyright 2002 PC Cluster Consortium
Copyright 2001,2000 Real World Computing Partnership
Amescomda

```

02/03/20

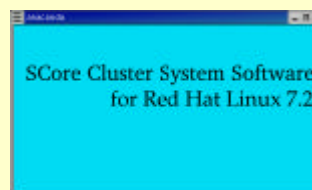
27



Powered by Xilinx

サーバ上の画面

- ◆ 最初右上のようなウインドウが現れます。
- ◆ 次に右下のようなウインドウが現れます。タイトルバーにホスト番号が表示されます。このウインドウ上でNextボタンを押すと実際のインストールが開始されます。
 - ◆ インストール後フロッピーを抜かずに電源投入すると、再びインストールが始まってしまいます。このとき、ここで止めてしまえば再インストールされません。



02/03/20

28





Powered by VMware

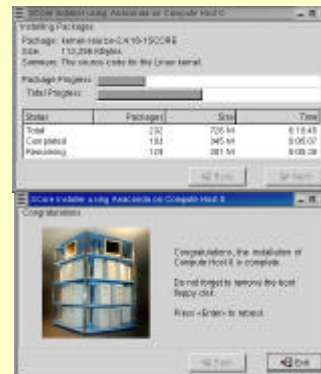
サーバ上の画面

- ◆ 右上のようなウインドウ順次現れます
- ◆ 右真中のウインドウでパッケージがインストールされている様子がわかります。
- ◆ インストールが終了すると右下のようなウインドウになります。フロッピーを抜いたのを確認の上Exitボタンをクリックしてください。自動的にリブートします。

Formatting / filesystem...

Setting up RPM transaction...

Preparing to install...



02/03/20

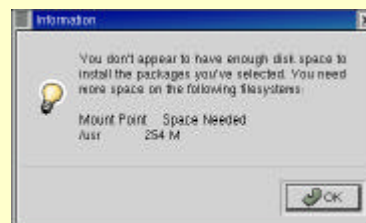
29



Powered by VMware

エラー画面の例

- ◆ インストール中、以下のウインドウが現れたら、ディスクサイズをきりなおしましょう。



02/03/20

30

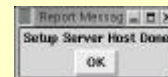




Powered by VMware

サーバの設定

- ◆ 計算ホストがインストールされたら、サーバ側の設定を行います。
- ◆ 終了すると、右下のようなウインドウが現れます



02/03/20

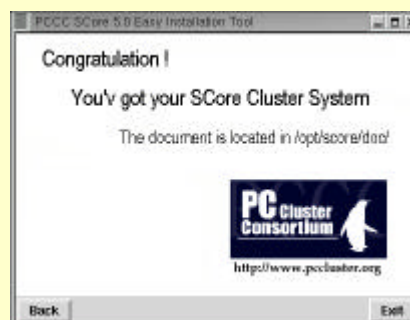
31



Powered by VMware

終了画面

- ◆ これでインストールはおしまいです。



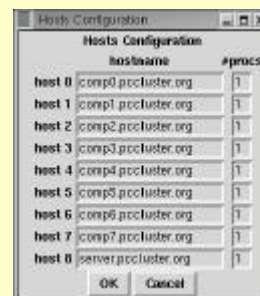
02/03/20

32





サーバホストも計算ホストに追加してみる



02/03/20

33



グループの変更



02/03/20

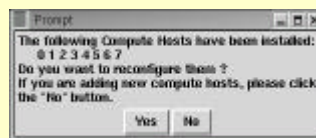
34





Cluster Setup

- ◆ クラスタ設定画面に移るときに、以下のようなウインドウが表示されます。
- ◆ Noボタンをクリックしてください。



SCoreのテスト(1/3)

- ◆ scoreboardのテスト
 - \$ scorehosts -l -g pcc
 - \$ sceptic -v -g pcc
- ◆ msgb実行
 - \$ msgb -group pcc
- ◆ scoutの実行
 - \$ scout -g pcc
 - \$ scout
 - \$ scout date



SCoreのテスト(2/3)

◆ 一対一ネットワークテスト

```
$ cd /opt/score/sbin  
$ ./rpmtest comp1 ethernet -reply
```

```
$ cd /opt/score/sbin  
$ ./rpmtest comp0 ethernet -dest 1 -ping
```

◆ トータルネットワークテスト

```
# cd /opt/score/sbin  
# scout -g pcc  
# ./sctest -network ethernet
```



SCoreのテスト(3/3)

◆ ユーザアプリケーションによるストレステスト

- ◆ 初期故障を見つけるために、必ずやるべき



EIT インターナル(1/3)

- ◆ EITはanacondaを改造して使っています。E IIの主な仕事は、
 - ◆ anacondaキックスタートファイルの作成
 - ◆ scorehosts.dbおよびpm-myrinet.conf, pm-ethernet.confの作成
 - ◆ bininstall -serverの実行
- ◆ ブートフロップのinitプログラムはEIT用に書き直しています。
 - ◆ DHCPプロトコルによるIPアドレスの取得
 - ◆ 必要なファイルのリモートマウント
 - ◆ キックスタートファイルの読み込み



EIT インターナル(2/3)

- ◆ サーバ上のファイル
 - ◆ /opt/score/setup/install/ndpostpro.sh
 - ◆ 計算ホストインストールの最後で起動されるスクリプト。このファイルはEITにより作成される
 - ◆ /opt/score/setup/install/svrpostpro.sh
 - ◆ サーバホストインストールの最後で起動されるスクリプト。このファイルはEITにより作成される
 - ◆ /opt/score/setup/RedHat/instimage/
 - ◆ インストール時のルートとなります
 - ◆ /opt/score/setup/RedHat/instimage/compconf/.conf/
 - ◆ このディレクトリの下にEITにより計算ホスト番号を名前としたキックスタートファイルが作られます。



EIT インターナル(3/3)

◆ サーバ上のファイル

◆ /opt/score/setup/RedHat/instimage/compconf/

- ◆ このディレクトリの下にEthernet MACアドレスを名前としたキックスタートファイルが作られます。これは、EITによって作られた /opt/score/setup/RedHat/instimage/compconf/.conf の下のファイルを rename して移動したものです。



E II カスタマイズ

- ◆ 計算ホストあるいはサーバホストに独自設定を行いたい時は、以下のファイルを修正してEITを出荷すれば良い

◆ /opt/score/lib/tcl/ndpostpro-skelton

◆ /opt/score/lib/tcl/svrpostpro-skelton

上記のファイルは、EIT実行時tclファイルとして読み込まれるので、EITで定義されているtcl変数名を参照することが可能です。変数名が展開されて、

◆ /opt/score/setup/install/ndpostpro.sh

◆ /opt/score/setup/install/svrpostpro.sh

に格納されます。



トラブル解決法

- ◆ 症状
 - ◆ 計算ホストインストール時、DHCP アドレスが取れないという
- ◆ 原因と対処方法
 - ◆ EITDと通信が出来ない
 - ◆ ネットワークが違っている
 - ◆ 計算ホストが2つのEthernetの口を持っている時は、とりあえず、2本ともサーバと同じセグメントに指しておきましょう
 - ◆ ネットワークドライバがサポートされていない
 - ◆ 御社でがんばってください



トラブル解決法

- ◆ 症状
 - ◆ 計算ホストインストール時、サーバのファイルシステムがマウントできない、あるいはマウント後ベルが鳴らずに止まってしまう。
- ◆ 原因と対処方法
 - ◆ サーバをno firewallで設定しましたか？
 - ◆ DNS 上はFQDNになっているが、/etc/hostsやNIS上ではホスト名しか登録されていないければ、統一し、/etc/exportsを綺麗にして再度eitを実行してください。
 - ◆ DNS上での名前の確認
 - ◆ host <hostname>
 - ◆ NIS 上での名前の確認
 - ◆ ypcat hosts | grep <hostname>
 - ◆ /etc/hosts 上での名前の確認
 - ◆ grep <hostname> /etc/hosts
 - ◆ サーバ上でexportfs -ra コマンドを実行した時にエラーが出ませんか？
/etc/exportsを綺麗にして再度eitを実行してください。



トラブル解決法

- ◆ 症状
 - ◆ msgb や scout が No host と言って起動しない
- ◆ 原因と対処法
 - ◆ DNS 上は FQDN になっているが、/etc/hosts や NIS 上ではホスト名しか登録されていない。サーバのホスト名が、DNS や NIS 上のホスト名と違うならば、統一した後に、
 - ◆ /etc/rc.d/init.d/scoreboard reload
 - ◆ /etc/rc.d/init.d/msgbserv stop; /etc/rc.d/init.d/msgbserve start
 - ◆ /opt/score/etc/scoreshosts.db に登録されている計算ホスト名、MSGBSERV ホスト名が、DNS、/etc/hosts、NIS で登録されているホスト名の形式 (FQDN) と異なっているならば、統一した後に、
 - ◆ /etc/rc.d/init.d/scoreboard reload
 - ◆ /etc/rc.d/init.d/msgbserv stop; /etc/rc.d/init.d/msgbserve start

02/03/20

45



トラブル解決法

- ◆ 症状
 - ◆ Disk partition 設定画面が出てくる
- ◆ 原因と対処法
 - ◆ anaconda の disk partition プログラムで対応できないことによるもので、
 - ◆ Partition が 4 つ以上設定できないディスクの場合
 - ◆ Root, /usr, swap だけを設定して、後のパーティションは 0 とする。
 - ◆ 例えば、
 - ◆ Root *
 - ◆ /usr 3072
 - ◆ Swap 128
 - ◆ swap 容量を 2GB 以上に設定した場合
 - ◆ 2GB より小さく設定する

02/03/20

46

