分散処理アプリ演習 APPENDIX edubase cloud環境 セットアップ

(株)NTTデータ

内容



- 1. はじめに 今回の環境の説明
- 2. セットアップ1 クラウドクライアント初期設定
- 3. セットアップ2 セキィリティグループ作成
- 4. セットアップ3 仮想マシン起動
- 5. セットアップ4 VNCでの接続
- 6. セットアップ5 環境設定シェルスクリプト実行

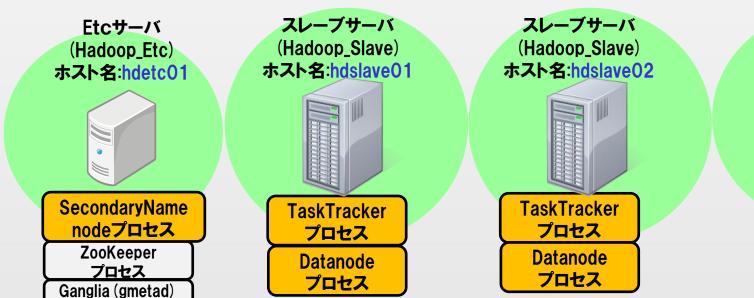
プロセス

EDUCATION PROGRAM FOR TOP SOFTWARE ENGINEERS

はじめに 今回の環境の説明









4 CHILERS EDUCATION PRODUCTION OF THE PRODUCTION

セットアップ1 クラウドクライアント初期設定

シンクライアント起動

- ログインID/パスワードは、シンクライアント貼付の白いシールに記入されている「tucl00xx」を小文字で使う
- クラウドクライアント初期設定
 - ■「スタート」-「すべてのプログラム」-「クラウドクライアント」-「CloudClient」を 起動
 - ■「ファイル」-「設定」-「クラウドクライアント」に、「接続URL」、 「ログインID」、「パスワード」、「プロジェクトチームID」、を設定する
 - ■「プロジェクトチーム名取得」、「認証情報取得」の順で実行する
 - ■「ファイル」-「設定」-「クラウドクライアント」-「キーペア」の名前リスト領域でマウス右メニューから「新規キーペア作成」を行う
 - キーの名前は各自がわかりやすい名前を設定「苗字」数字」等
 - 例: takahashi_01, saito_02(苗字がかぶる可能性があるので何か適当な数字も付加)
 - → 各自、6VMづつ起動することになるので見つけやすい名前がよい。

※次ページ以降でキャプチャー画面による流れを説明

セットアップ1 クラウドクライアント初期設定

5 SUMERS EDUCATION DA PROPERTIES OF THE PROPERTI

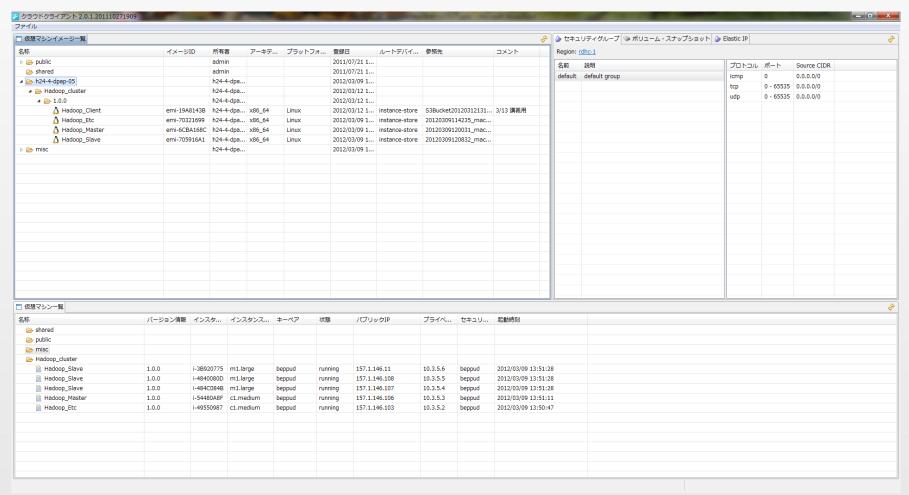
■ クラウドクライアント起動

接続URL	プロジェクトチームID	ログインID		
https://vclc0017.ecloud.nii.ac.jp:8773/services/RDHC	h24-4-dpap-a	tucl0001	tucl0002	tucl0003
	h24-4-dpap-b	tucl0004	tucl0005	tucl0006
	h24-4-dpap-c	tucl0007	tucl0008	tucl0009
	h24-4-dpap-d	tucl0010	tucl0011	tucl0012
	h24-4-dpap-e	tucl0013	tucl0014	tucl0015
	h24-4-dpap-f	tucl0016	tucl0017	tucl0018
	h24-4-dpap-g	tucl0019	tucl0020	tucl0021
	h24-4-dpap-h	遠隔受講者1	遠 隔受講者 2	遠隔受講者 3
	h24-4-dpap-i	遠隔受講者4	遠隔受講者 5	遠隔受講者6

6 SUMERS EDUCATION PRODUCTION OF THE NGINEERS OF THE NGINEERS

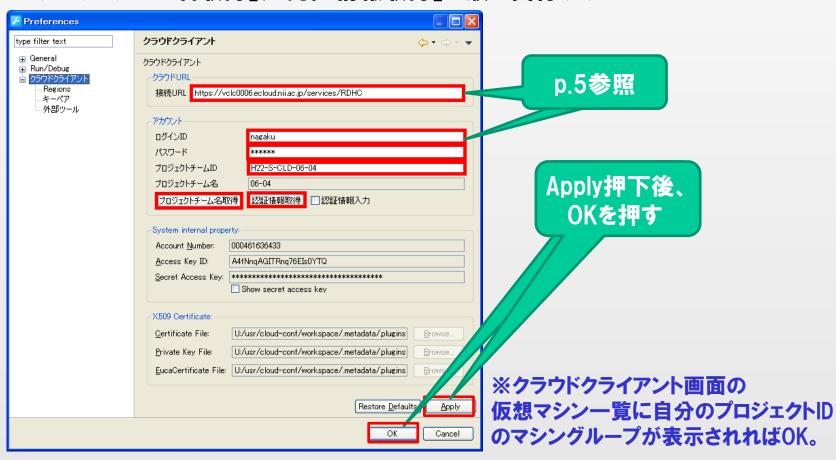
セットアップ1 クラウドクライアント初期設定

- クラウドクライアント初期設定
 - ■「スタート」-「すべてのプログラム」-「クラウドクライアント」-「CloudClient」を起動



セットアップ1 クラウドクライアント初期設定

- クラウドクライアント初期設定
 - ■「ファイル」-「設定」-「クラウドクライアント」に、「接続URL」、 「ログインID」、「パスワード」、「プロジェクトチームID」、を設定する
 - ■「プロジェクトチーム名取得」、「認証情報取得」の順で実行する

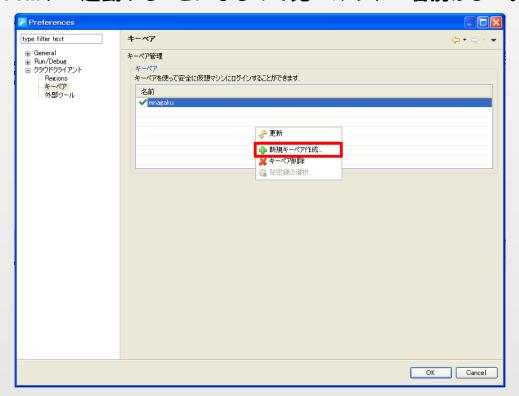




8 SUNTERS EDUCATION PROPERTY OF THE NGINEERS O

セットアップ1 クラウドクライアント初期設定

- クラウドクライアント初期設定
 - ■「ファイル」-「設定」-「クラウドクライアント」-「キーペア」の名前リスト領域でマウス右 メニューから「新規キーペア作成」を行う
 - キーの名前は各自がわかりやすい名前を設定「苗字」数字」等
 - 例: takahashi_01, saito_02 (苗字がかぶる可能性があるので何か適当な数字も付加)
 - → 各自、6VMずつ起動することになるので見つけやすい名前がよい。



BOT THARE PRODUCTION P

セットアップ2 セキィリティグループ作成

- セキュリティグループの作成の流れ
 - ■「セキュリティグループ」の「名前‐説明」リスト領域でマウス右メニューから「新規グループ」を選ぶ
 - セキュリティグループ名はキーペア名と同様各自わかりやすい名前を設定、説明を適 当に付けて、セキュリティグループを作成する
 - ■「名前-説明」リスト領域で作成したキー名を選択状態とし、「プロトコル-ポート-SourceCIDR」リスト領域でマウス右メニューから「パーミッションの追加」を選ぶと、セキュリティグループの設定が出来る
 - 下記、表の設定を行う

プロトコル	ポート	SourceSIDR
icmp	0	0.0.0.0/0
tcp	0-65535	0.0.0.0/0
udp	0-65535	0.0.0.0/0

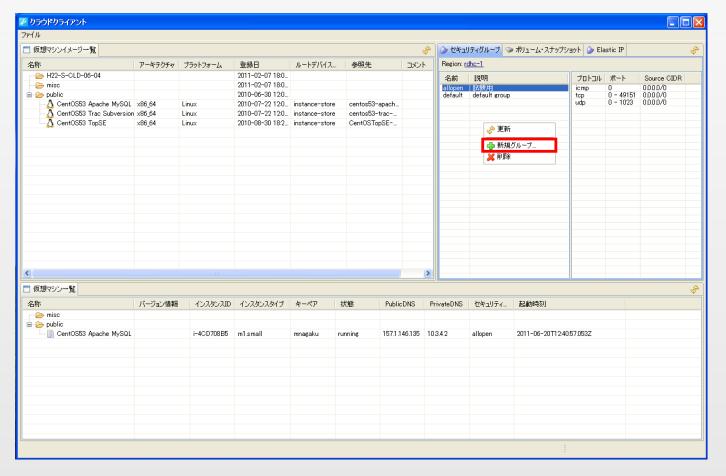
※次ページ以降でキャプチャー画面による流れを説明

10

EDUCATION PROGRAM FOR TOP SOFTWARE ENGINEERS

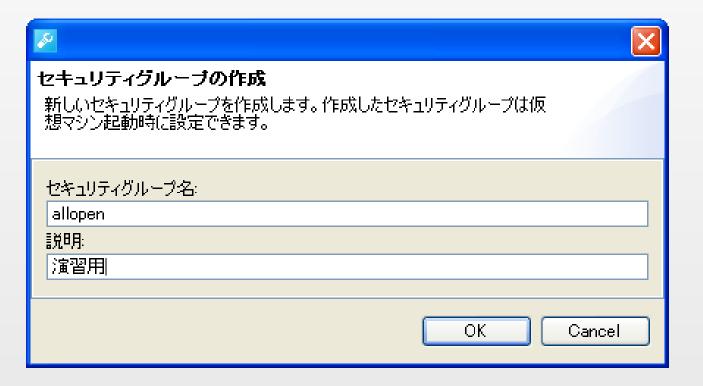
セットアップ2 セキィリティグループ作成

- セキュリティグループの作成
 - ■「セキュリティグループ」の「名前‐説明」リスト領域でマウス右メニューから「新規グループ」を選ぶ



セットアップ2 セキィリティグループ作成

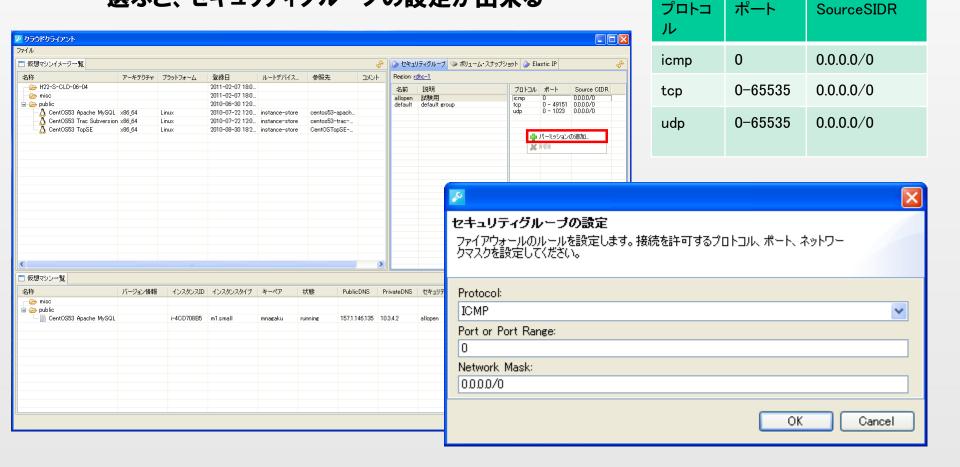
- セキュリティグループの作成
 - セキュリティグループ名を、キーペアと同様にわかりやすい名前を付ける。
 - 説明を適当に付けて、セキュリティグループを作成する





セットアップ2 セキィリティグループ作成

- セキュリティグループの作成
 - ■「名前-説明」リスト領域で作成したセキュリティグループを選択状態とし、「プロトコルーポート-SourceCIDR」リスト領域でマウス右メニューから「パーミッションの追加」を選ぶと、セキュリティグループの設定が出来る





セットアップ3 仮想マシン起動

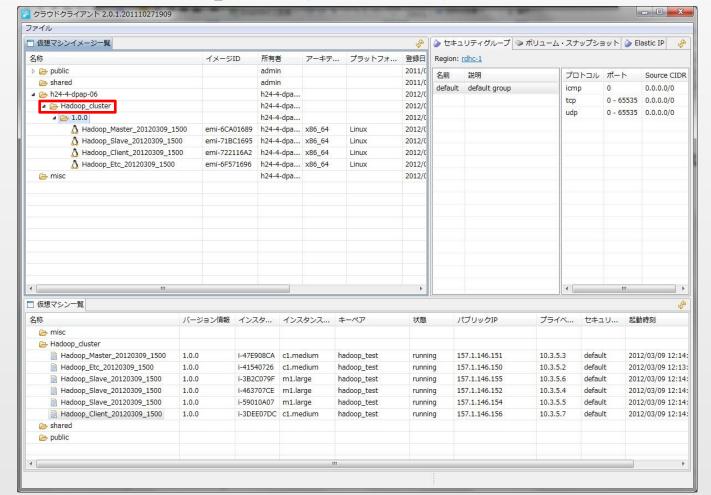
仮想マシン起動

- ■「仮想マシンイメージ一覧」-「shared」-「h24-4-dpap-a~i」のマウス右メニューから「仮想マシン起動」を選ぶ(グループ起動)
- 下記表のとおりに6VMを起動する
- ■「仮想マシン一覧」-「Hadoop_cluster」にインスタンスが起動する
- 自分が起動したインスタンスを見分けるには、キーペアを見る
- 10分ほど待って「仮想マシン一覧」の更新を行うと、状態が「running」となり、仮想マシンの起動を確認出来る ※次ページ以降でキャプチャー画面による流れを説明

仮想マシン名	バージョン	インスタンス タイプ	キーペア	インスタンス数	セキュリティグ ループ
Hadoop_Slave	1.0.0	m1.large	自分で設定し たもの	3	自分で設定し たもの
Hadoop_Etc	1.0.0	c1.midium	自分で設定し たもの	1	自分で設定し たもの
Hadoop_Master	1.0.0	c1.midium	自分で設定し たもの	1	自分で設定し たもの
Hadoop_Client	1.0.0	c1.midium	自分で設定し たもの	1	自分で設定し たもの

セットアップ3 仮想マシン起動

- 仮想マシン起動
 - ■「仮想マシンイメージ一覧」-「shared」-「h24-4-dpap-a~i」のマウス右メニューから「仮想マシン起動」を選ぶ(グループ起動)





セットアップ3 仮想マシン起動

- 仮想マシン起動
 - インスタンスタイプ等をp.12の表にしたがって入力、起動する。
 - 起動に10分程度かかる。





セットアップ3 仮想マシン起動

CFTWARE

■ 仮想マシンの操作

- ■「仮想マシン一覧」-「shared」に起動した、自分のキーペアを持ったインスタンス6つが、「running」状態であることを確認する
- <u>自分のキーペア持ったインスタンスのIPを見つけ、覚えておいてください。</u> プライベートIP:6VM分 ※インスタンスを右クリックでIPのコピーが可能

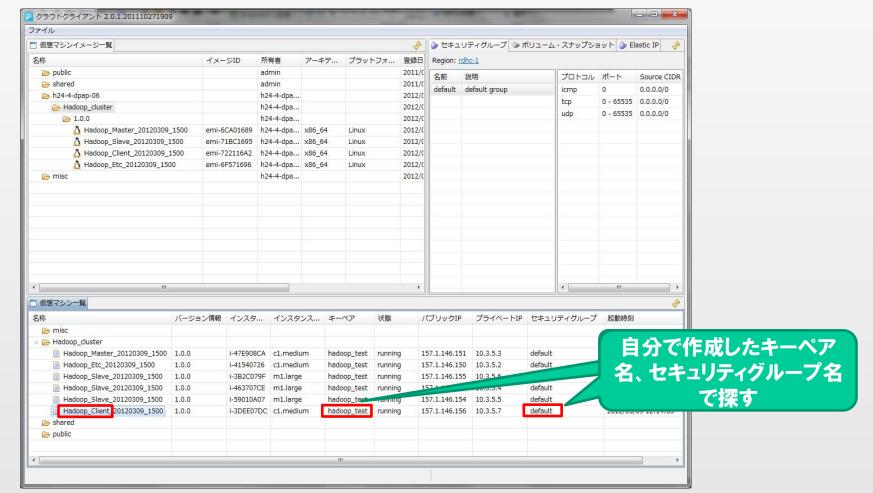
パブリックIP: Clientの分

- 自分が起動したインスタンス「Hadoop_Client」で、マウス右メニューから「シェルの起動」を選ぶ
- 作成したキーペアを始めて使い時には、鍵の保存が必要となる ※ポップアップで出る鍵の保存場所を各自記憶しておくこと
- 鍵の保存後、「PuTTY Key Generator」を「×」で閉じる
- 鍵を保存したら、再度「Hadoop_Client」で「シェルの起動」を行う
- 始めてアクセスするIPアドレスに対しては確認を求められる
- ssh接続で仮想マシンが操作出来る

※次ページ以降でキャプチャー画面による流れを説明

セットアップ3 仮想マシン起動

- 仮想マシンの操作
 - ■「仮想マシン一覧」-「shared」の中に自分が起動したインスタンスの「Hadoop_Client」のマウス右メニューから「シェルの起動」を選ぶ。



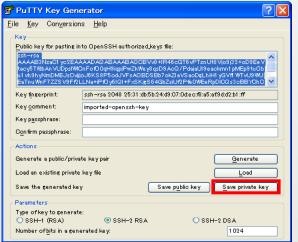


セットアップ3 仮想マシンの操作の流れ

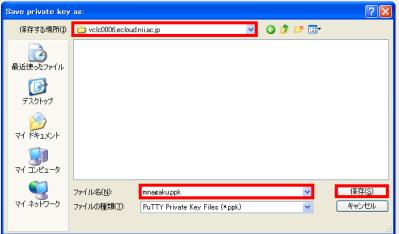
- 仮想マシンの操作
 - 作成したキーペアを始めて使う時には、鍵の保存が必要となる



(3) (4)





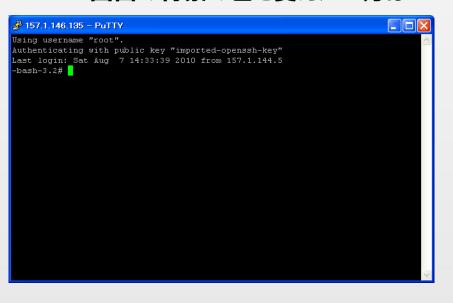


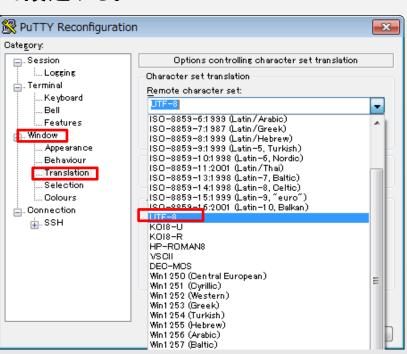
18

※(1)で表示された場所に 自分で作った「キーペア名.ppk」で保存

セットアップ3 仮想マシン起動

- 仮想マシンの操作
 - 始めてアクセスするIPアドレスに対しては確認を求められる
 - ssh接続で仮想マシンが操作出来る
- PuTTYの設定
 - PuTTYの画面のタイトルバーで右クリック →「Change Settings...」→「Window-Translation」→「UTF-8」を設定する。
 - 画面の背景の色を変えたい方は「Colours」で設定する。









キャンセル

いいえ(<u>N</u>)

20

EDUCATION PROGRAM FOR TOP SOFTWARE ENGINEERS

セットアップ3 仮想マシン起動



■ /root/hadoop_exercise/ 以下に演習資材があるので確認する

```
$ ls /root/hadoop_exercise
01 02 04 06 07 11 12 14
```



セットアップ4 VNCでの接続



■ hdclient01のパブリックIPアドレス:5901を入力

例: 157.1.145.146:5901

- パスワードは「hadoop」
- hdclient01にグラフィカルログイン

- デスクトップにeclipseを起動するアイコンがあるか確認する
 - eclipseを利用する演習あり
- デスクトップにHadoopのAPIドキュメントへのリンクがあるか確認する
 - HadoopAPIDocs内のindex.htmlをクリック



セットアップ5 環境設定シェルスクリプト実行

- **■** Hadoopを使用するために
 - Hadoopシステムを使用するためには、HDFS・MapReduce用領域の作成や、 NameNodeのフォーマット、そしてHadoopデーモン(NameNode、DataNode、 JobTracker、TaskTracker)を起動しておく必要がある
 - 今回の授業では、シェルスクリプトを使用してHadoopシステムの初期設定を行う



23

EDUCATION PROGRAM FOR TOP SOFTWARE ENGINEERS

セットアップ5 環境設定シェルスクリプト実行



- Hadoop初期設定
 - hdclient01のシェルを起動したら、/root/shellディレクトリへ移動し、node.txtを編集 する

```
$ cd /root/shell
$ vi node.txt
```

■ node.txtには、以下の内容が書かれている

```
#IP_master
#IP_slave
#IP_slave
#IP_slave
#IP_client
#IP_etc
```

セットアップ5 環境設定シェルスクリプト実行



■ Hadoop初期設定

■ node.txtの内容を消去し、自分が起動したHadoopクラスタのプライベートIPアドレス を記入し、保存する(各ノードのプライベートIPアドレスはCloudClient上の「仮想マシン一覧ビュー」に記載がある。自分で設定したキーペアを持つノードを探し、プライベートIPアドレスを抜き出す)



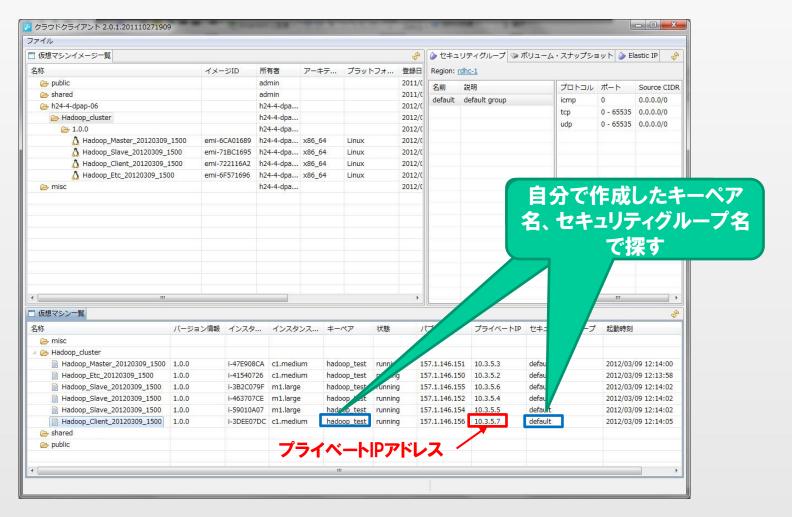
1. マスターのIPアドレス
2. スレーブのIPアドレス
3. スレーブのIPアドレス
4. スレーブのIPアドレス
5. クライアントのIPアドレス
6. ETCのIPアドレス
の順に1行目から書く
※順番厳守

- ※IPアドレスは自分が起動した仮想マシンのものを記入すること
- ※IPアドレスは必ず続き番号(順不同)になっているので、保存する前に確認すること

セットアップ5 環境設定シェルスクリプト実行

■ Hadoop初期設定

(参考1)プライベートIPアドレスの記載場所(パブリックIPアドレスと間違えないこと)





26

EDUCATION PROGRAM FOR TOP SOFTWARE ENGINEERS

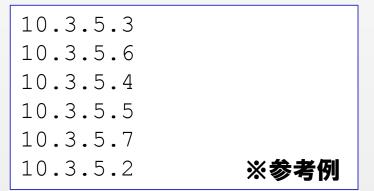
セットアップ5 環境設定シェルスクリプト実行

■ Hadoop初期設定 (参考2)node.txtの書き換え

・node.txt編集前

```
#IP_master
#IP_slave
#IP_slave
#IP_slave
#IP_client
#IP_etc
```

・node.txt編集後



内容を消去して、書き換える

AND THE PROPERTY OF THE PROPER

セットアップ5 環境設定シェルスクリプト実行

- Hadoop初期設定
 - 00_ping_check.shを実行し、6つのインスタンスすべてにpingが通ることを確認する

```
$ sh 00_ping_check.sh
```

■ 01_setting_first_hadoop.shを実行する

```
$ sh -x 01_setting_first_hadoop.sh
```

■ 02_starting_first_hadoop.shを実行する

```
$ sh -x 02_starting_first_hadoop.sh
```

■ 03_starting_first_ganglia.shを実行する

```
$ sh -x 03_starting_first_ganglia.sh
```

※ 01_setting_first_hadoop.sh使用時に "ERROR:Some servers are unconnected."と表示されたら… しばらくしてから再度、01_setting_first_hadoop.shを実行してください。

セットアップ5 環境設定シェルスクリプト実行



- Hadoop初期設定
 - DataNodeがNameNodeに接続されHDFSのメンバーに含まれたかをdfsadminコマンドより確認する(Datanodes availableの行を確認し、3と表示されていればOK)
 - \$ hadoop dfsadmin -report
 - TaskTrackerがJobTrackerに接続されMapReduceのメンバーに含まれたかをjobコマンドで確認する(3と表示されればOK)
 - \$ hadoop job -list-active-trackers | wc -l
 - 以上で、Hadoopの初期設定が終了し、Hadoopデーモン(NameNode、DataNode、 JobTracker、TaskTracker)が起動したことを確認できる

(仮想マシン立ち上げ時には、Hadoop初期設定を毎回行うこと)

VNCのブラウザで

http://hdmaster01:50070/

http://hdmaster01:50030/ と入力し、 Hadoopデーモンの起動を確認する方法もある

デフォルトからの主な変更点

- Hadoop用領域の場所
 - /mnt/hadoop/data
- logの出力場所
 - /var/log/hadoop-0.20/ … Hadoop関連ログ
 - /var/log/hive/ … hive関連ログ
 - /var/log/hbase/ ··· HBase関連ログ
- パーミッション
 - 演習の関係上、HDFSのパーミッションを無効に設定してある。設定場所は /etc/hadoop-0.20/conf.hdsol/hdfs-site.xml

```
<name>dfs.permissions</name>
    <value>false</value>
```



(付録) 環境設定シェルスクリプト実行

- Hadoop起動(2回目以降)
 - /root/shell/hadoop_operation/ディレクトリに移動する
 - hadoop_start.shを実行する

```
$ cd /root/shell/hadoop_operation/
$ sh hadoop_start.sh
```

■ 以上で、Hadoopデーモン(NameNode、DataNode、JobTracker、TaskTracker)が 起動する



(付録) 環境設定シェルスクリプト実行

- Hadoop停止
 - /root/shell/hadoop_operation/ディレクトリに移動する
 - hadoop_stop.shを実行する

```
$ cd /root/shell/hadoop_operation/
$ sh hadoop_stop.sh
```

■ 以上で、Hadoopデーモン(NameNode、DataNode、JobTracker、TaskTracker)が 停止する



TOP SOLUTION OF THE PROPERTY O

(付録) その他の操作について ※演習では利用しません

- スナップショットの保存
 - 自分が起動したインスタンスの「名称」のマウス右メニューから 「イメージの追加」を選ぶ
 - 他の人と重複しない適当な「登録名」、仮想マシングループ名「misc」、仮想マシン ディスクサイズ「6144」として「Finish」する
 - 10分ほどでスナップショットが作成される
- 仮想マシンの停止
 - 自分が起動したインスタンスの「名称」のマウス右メニューから 「インスタンス停止」を選ぶ
- スナップショットの確認
 - ■「仮想マシンイメージ一覧」-「misc」で自分の作成したスナップショットのマウス右メニューから「仮想マシン起動」を選ぶ