

# グリッド・コンピューティングとR

西日本電信電話株式会社

山本 寛

福崎 昭伸

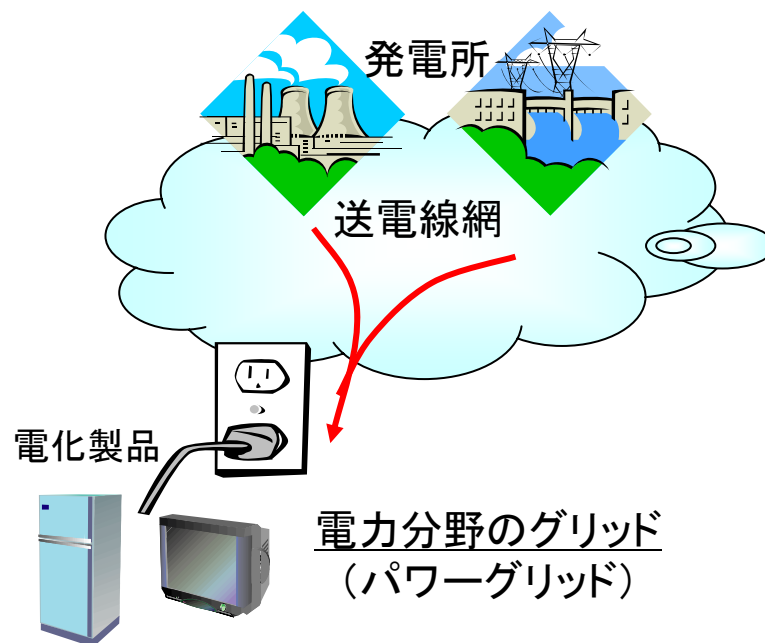
統計数理研究所

中野

純 司

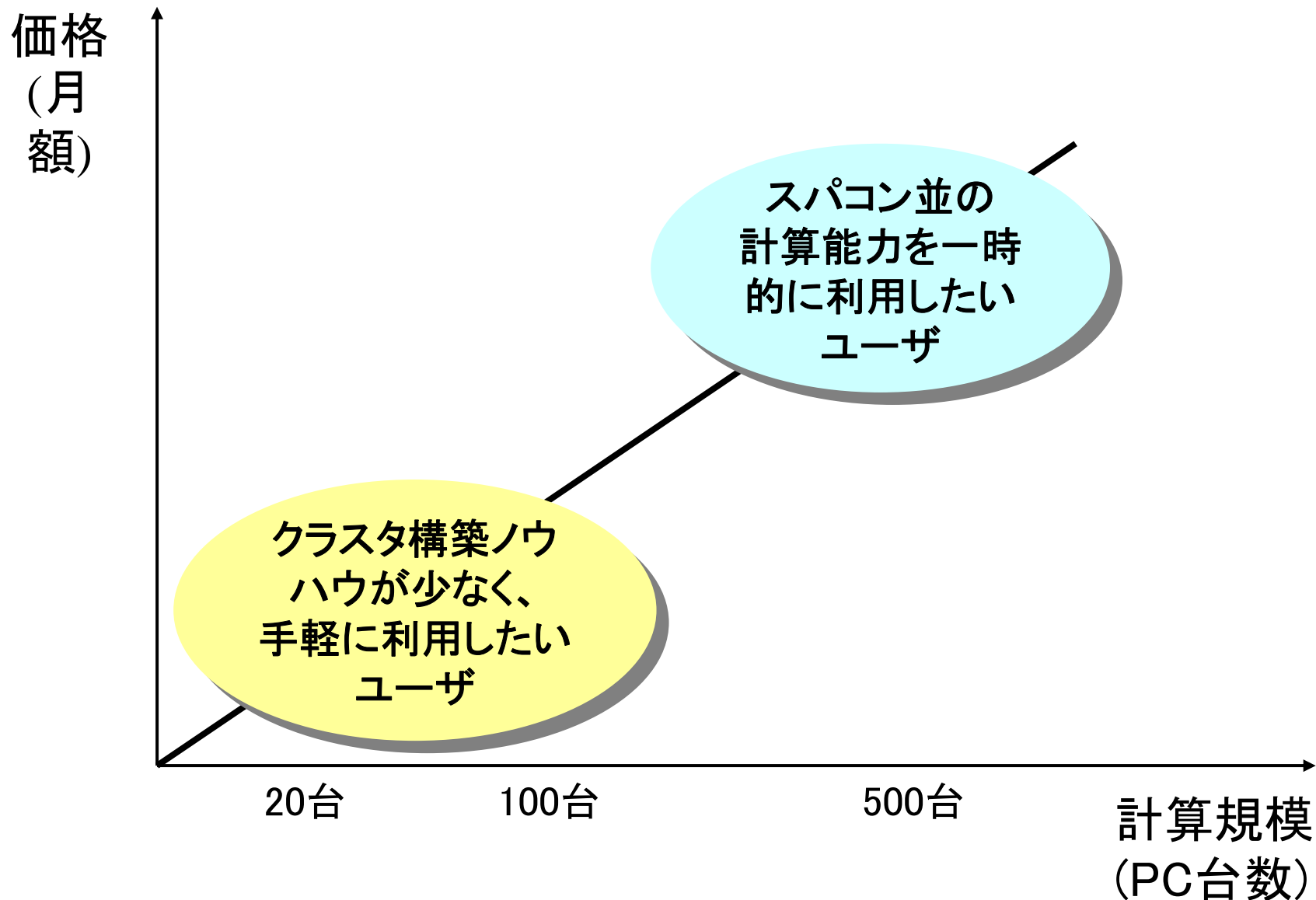
# グリッドの概要

グリッドとは、電気を伝える高圧送電線網(パワーグリッド)に由来しています。コンセントに差し込めばいつでもどこでも必要なだけ電力が得られるように、ネットワークを通して、安全に・安定して・安易に様々なサービス(計算資源、アプリケーションなど)を利用できるようにするというコンセプトを指します。

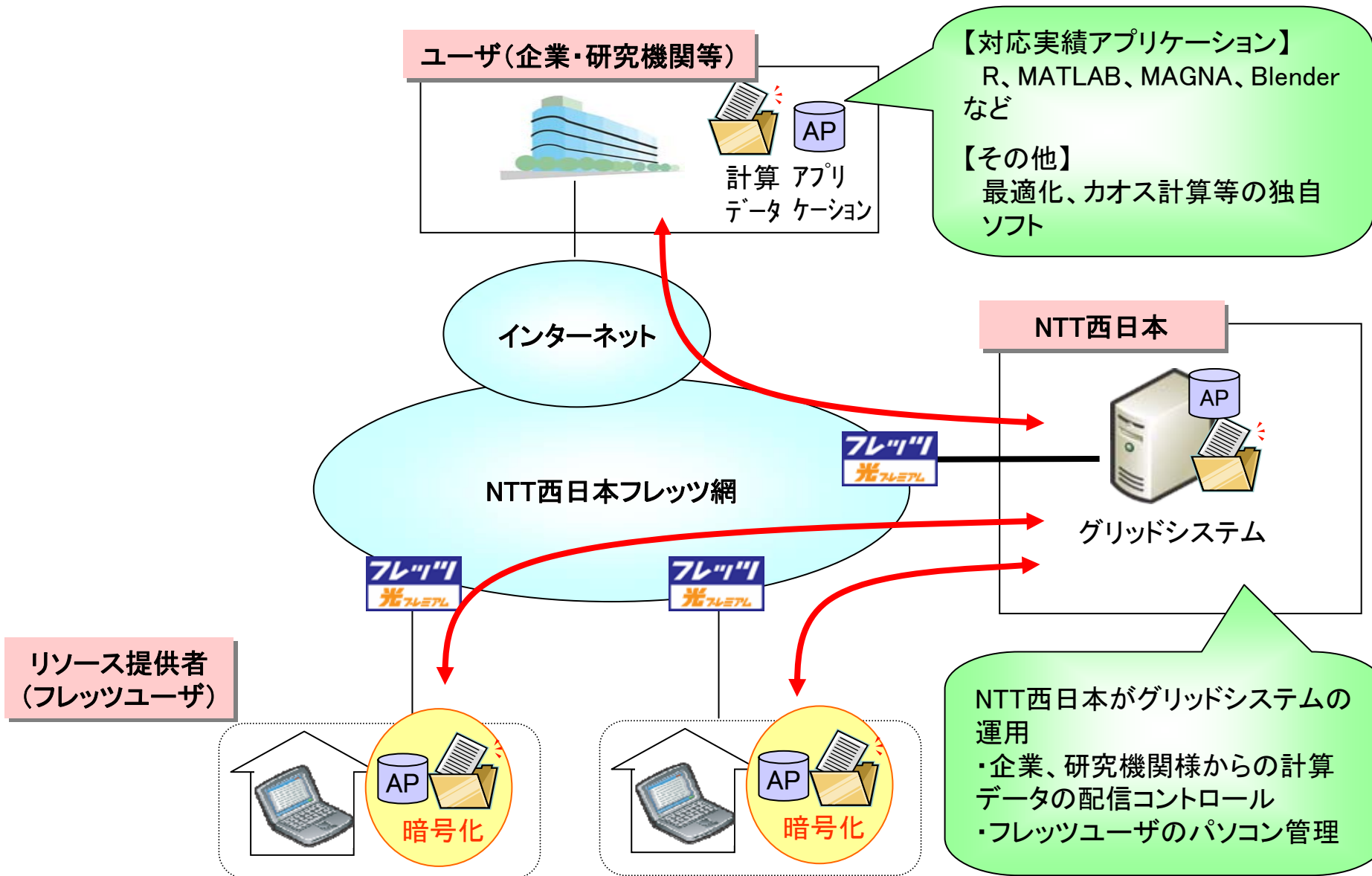




# ひかりグリッドの利用者層 ～規模と価格～



# ひかりグリッド概要



## アプリケーション条件

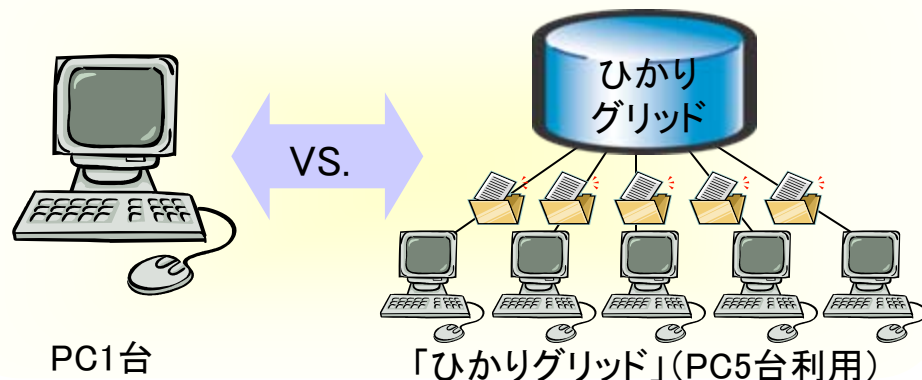
- 大量な計算処理が必要になること
- 計算処理を分割して、独立して処理できること
- WindowsVista、WindowsXP上で動作すること
- 複数のパソコンに配布可能なライセンス条件であること  
(Rなどのフリーソフト)
- 計算結果取得までの厳密なターンアラウンドタイムを必要としないこと
- 個々の計算処理時間が短時間であること(数十分程度)

## 【評価内容】

- ・ MATLABによるモンテカルロ・シミュレーション (モーメント適合法による乱数生成の正規化)
- ・ MATLABコンパイラによる実行モジュールを『ひかりグリッド』に実装し、PC5台で並列計算を実施

## 【環境】

- ・ 使用ソフト:  
モーメント適合法プログラム (exe)、  
MATLABランタイムライブラリ
- ・ 使用パソコン:
  - Core2Duo 1.6GHz (2.0GB メモリ)



※各PCのOSはWindowsXP, MATLABランタイム環境をインストールしています。

## 【結果】

	時間(分)	倍率
PC1台	90	1
「ひかりグリッド」(PC5台利用)	22	4.15

※PC台数を増やせば、4.15倍以上の計算時間の短縮が見込めます。

※計算データの入出力に伴う時間は対象外としています。

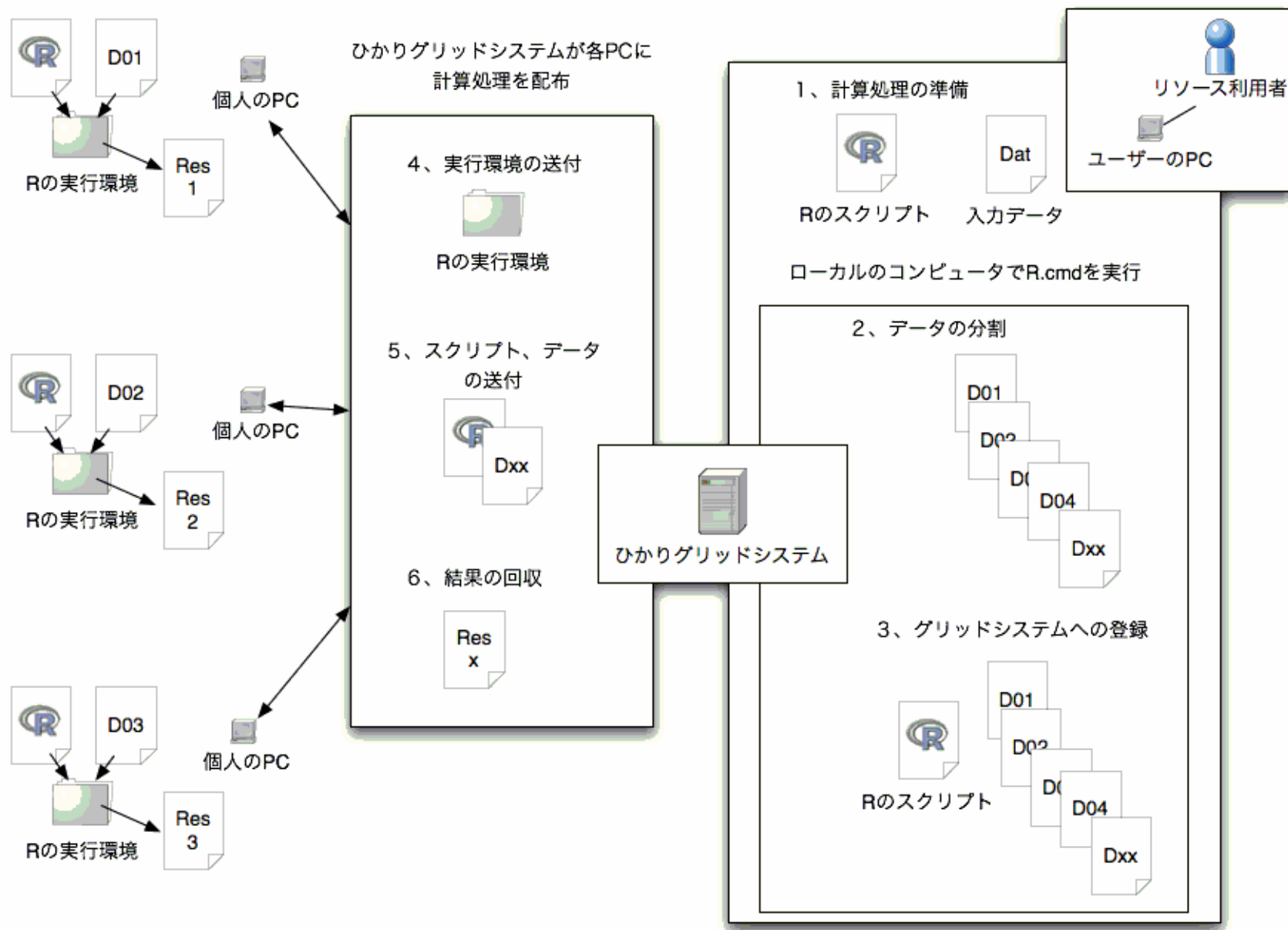


# グリッド版R

- ひかりグリッド版「R」は以下を組み合わせで実装済み
  - 「R」実行環境(ネットで配布されているもの)
  - SDK(ひかりグリッド利用者に提供)
- 「R」の統計計算を簡便に実行するための支援ツール
  - R.cmd(ひかりグリッド利用者に提供)

以下では、サンプルスクリプト、データを用いて実際にジョブを投入し、結果を得てみることにします。

# グリッド版Rの実行の流れ



毎日の気温を記録したデータを配列にし、この配列データを入力して、グラフを描画するスクリプトです。気温のデータは、3地点での各365日分が記録されており、配列はシーケンシャルに展開されているため、入力データファイルはヘッダーをあわせて1096行あります。

```
data <- read.table("temperature.txt", header = T)
len <- nrow(data)
period <- 365
start <- 1
count <- 1
while(start <= len) {
  png(paste("test",count,".png",sep=""))
  end <- min(start+period-1,len)
  tt <- data[start:end,]
  plot(tt$Day, tt$Temp, type = "l",
       ylim = c(-15,30),
       xlab = "Day", ylab = "Temperature")
  start <- start+period
  count <- count+1
  dev.off()
}
```

Day	Temp
1	-10.5
2	-8.7
3	-6.5
4	-7.1
5	-9.1
6	-11
7	-8.1
8	-9.8
9	-11.4
10	-8.2
.	.
.	.
.	.

File: temperature.txt

File: autorun.R

このスクリプトは単独のPCで稼働し、3枚のグラフを描きます。これを支援ツール(R.cmd)を使ってグリッドシステムへ投入してみます。このとき、R.cmdが、データファイルを365行単位(地点別)で自動的に分割し、合計3つの入力データにしてくれるようにパラメータをセットします。

R.cmdは、配列や行列、ファイルからのデータ入力を含む「R」の統計計算を簡便に実行するための支援ツールです。R.cmdは、①「R」のスクリプトと入力データをひかりグリッドシステムの対応形式にし、②グリッド・ジョブとして登録・投入し、③結果を回収する、といった機能を持ちます。R.cmdは、ひかりグリッドの利用者がアカウントとともにNTT西日本から提供を受けることができます。このツールを使うことで、ユーザーはプログラムモジュールやデータパッケージの作成といったひかりグリッドでプログラムをグリッド化する際に本来必要な作業を行なうことなく「R」のスクリプトを簡便に実行できます。以下にR.cmdの簡単な使い方を紹介いたします。

## R.cmd 使用方法

### Addjob (Data-splitting mode):

Rのスクリプト、データを指定して指定の条件でデータを分割し、複数のワークユニット(分割された処理単位)を作り出し、ジョブを投入する。投入時に表示されたjobidを使って後で結果を回収できる。

### Syntax:

```
-----  
R.cmd addjob -dir <directory_name> -script <script_file> -splitfile <file_to_split> [-header <number_of_lines>] [-splitsize  
<splitSize>] -splittype CHUNKS
```

### Options:

```
-----  
<directoryname> スクリプトとデータのあるディレクトリ <script file> スクリプトのファイル名 <file to split> 入力ファイル(分割対象)  
<number of lines> データのヘッダーの行数 <splitSize> データ分割時の1つのファイルごとのデータ数(行数)
```

### Retrieve:

Addjobで投入したジョブの結果をダウンロードし、ローカルのディスクに保存する。Jobidを省略した場合は自分のアカウントで投入したジョブの一覧を表示する。

### Syntax:

```
-----  
R.cmd retrieve [<jobid>]
```

R.cmdが設置されているR用の作業ディレクトリ下にdataディレクトリを作り、その下にhgsampleというディレクトリをさらに作って今回のサンプルスクリプトとデータを配置いたします。コマンドラインでR.cmdのディレクトリに移動し、以下のコマンドを実行します。

```
F:¥UD_R_v5.1.02>R.cmd addjob -dir data¥hgsample -script data¥hgsample¥autorun.R -splitfile
data¥hgsample¥temperature.txt -header 1 -splitsize 365 -splittype CHUNKS
...省略...
```

-----  
Job ID 1506  
-----

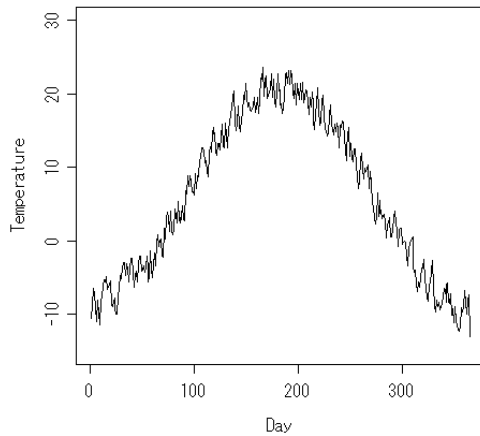
```
Uploading data package 1...done
Uploading data package 2...done
Uploading data package 3...done
Created 3 workunit(s)
...省略...
```

Job ID 1506にて3ワークユニットのジョブが生成され投入されました。しばらく経過したのち、結果を回収します。

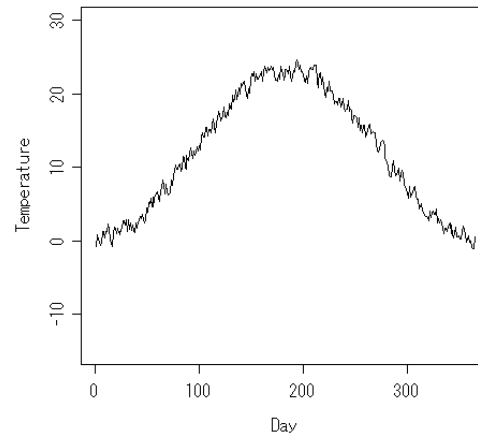
```
F:¥UD_R_v5.1.02>R.cmd retrieve 1506
...省略...
```

現在のディレクトリに1506という名前のディレクトリが作成され、その中にワークユニット毎の結果パッケージ(tarファイル)がディレクトリに分かれて保存されます。これらのtarファイルを展開して結果を確認します。

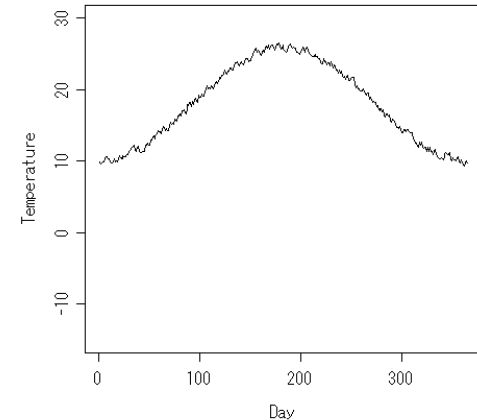
結果の回収によって生成された1506というディレクトリの下には投入されたワークユニットの個数に対応したディレクトリができています。今回はWU\_1、WU\_2、WU\_3の3つです。これらのディレクトリ下にあるresult.tarファイルを展開すると、各ワークユニットの実行結果であるグラフのpngファイルが取り出されます。



WU\_1¥test1.png



WU\_2¥test1.png



WU\_3¥test1.png

以上のようにひかりグリッドとR.cmdを利用することで簡単にRのスク립トを並列に走らせることができます。相互に依存のない独立な計算の部分はこのスキームにすることで数百台以上の計算資源を利用した高速な演算が実現できるようになります。

ただし、個別のアルゴリズムによって効率的なやり方は変わってきますので、ご関心のある方はお気軽にお問い合わせ願います。

## ■お問い合わせ先

西日本電信電話株式会社 サービスクリエーション部

新ビジネス部門

山本 寛

URL : <http://www.nttw-grid.jp/>

E-mail : [hiroshi.yamamoto@west.ntt.co.jp](mailto:hiroshi.yamamoto@west.ntt.co.jp)