

---

# SQC勉強会(1)

## (QCの概要とソフトウェア(R、Pentaho)の紹介)

### 統計的品質管理とは？

---

- 製品・サービスの品質を
- 計量データと統計的手法に基づき
- 管理する手法

## 品質の管理とは？

2009年11月5日

- 品質:「(モノの)利用における(効用の)適合度」  
(Montgomery, 2005)
- 4要素
  - 対象
    - 何(部門、工程)を対象とするのか
  - 特性
    - 対象のどのような性質を管理したいのか
  - 標準(水準)
    - 許容できる特性の測定量はどの範囲か？
  - 要因
    - 特性の計量値に影響を与える要素は何か？
- SQCでは特に標準の決め方と活動データと標準との関係の判断に関する手法を重視

QC勉強会資料

## 事例データ

2009年11月5日

- あるシステムのレスポンスデータ
- 3台のPCサーバ
- レスポンスの遅さを改善したいが要因(原因)が分からない
- 生データのため扱いにくい
  
- データの加工、はずれ値の判断、要因の特定

QC勉強会資料

## 事例における品質の検討

2009年11月5日

- 管理の4要素
  - 対象
    - システム(=ハードウェア+ソフトウェア+ネットワーク)
    - 特にハードウェア+ソフトウェア
  - 特性
    - レスポンス時間
    - 代替値としての転送容量
  - 標準(水準)
    - まだ分からない。これから検討する
  - 要因
    - サーバ情報、プログラム(URL)

QC勉強会資料

## 今回の実施手順

2009年11月5日

- 生データを加工し、分析用のデータに変換(Pentaho)
- QCツールのいくつかを用い、標準の検討と不良データの要因解析(Rコマンドー)
- 詳細な説明は次回以降

QC勉強会資料

## 1.データの加工

2009年11月5日

- データの必要な部分を抽出(Pentaho)
- 不必要な情報を切り落とす
- 分散した情報をまとめる
- 必要に応じて情報を加工(計算、置き換え)する

OC勉強会資料

-7-

## データ加工 : Pentaho

2009年11月5日

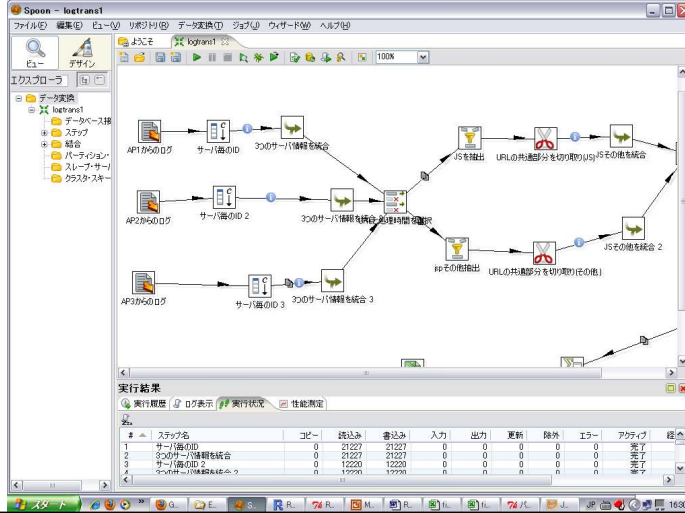
- 商用オープンソースのBIソフト
- データ加工、分析(DWH)、レポート、ダッシュボード、データマイニングのモジュールから構成
- GPLライセンスに基づき使用可能
- 今回はデータ加工(DataIntegration)を利用

OC勉強会資料

# データ加工 : Pentaho DataIntegration

2009年11月5日

- GUIにより変換用テンプレートを定義する



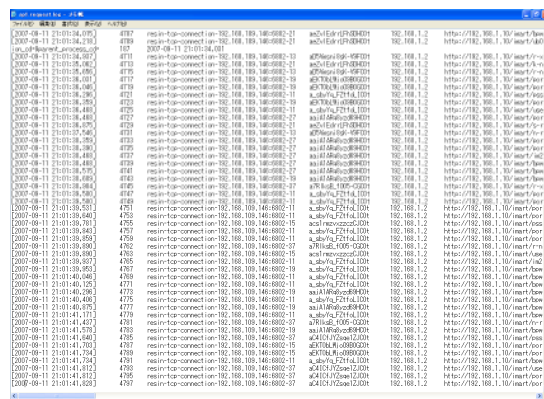
The screenshot displays the Spoon - InetTrans1 application window. The main area shows a data flow diagram with several steps: 'APIからのログ' (Log from API), 'サーバ番号ID' (Server ID), '3つのサーバ情報を読み込み' (Load 3 server info), 'JSONを抽出' (Extract JSON), and 'URLの共通部分を切り取り(JSP)の共通部分を抽出' (Extract common parts of URL (JSP) and common parts of URL). The bottom panel shows the '実行結果' (Execution Results) table.

ステップ名	コネ	読み込み	入力	出力	更新	除外	エラー	アクティブ
1	0	21227	21227	0	0	0	0	完了
2	0	21227	21227	0	0	0	0	完了
3	0	12220	12220	0	0	0	0	完了
4	0	12220	12220	0	0	0	0	完了

OC勉強会資料

## データ

2009年11月5日



The screenshot shows a Notepad window containing a large volume of log entries, all starting with 'res-in-tcp-connect'. The entries include IP addresses, port numbers, and connection details, illustrating a dense stream of raw data.

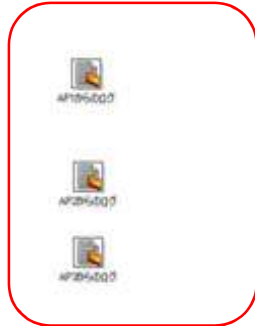
生データのため扱いにくい(データ形式)  
⇒必要な部分のみを取り出してみる

OC勉強会資料

## テキストデータを取り込んでみる

2009年11月5日

- まず、テキストデータを取り込んでみる



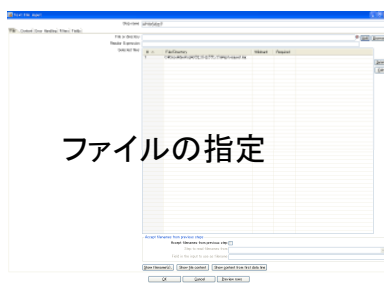
.11.

QC勉強会資料

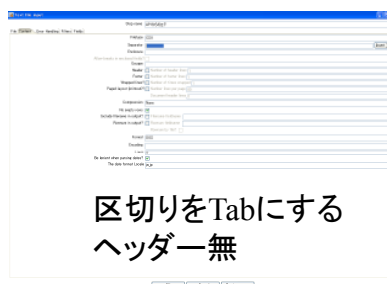
## テキストデータを取り込んでみる

2009年11月5日

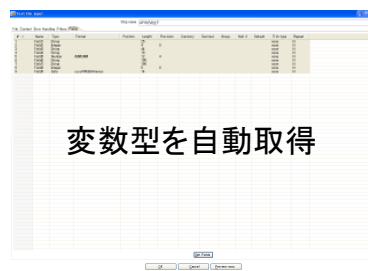
ファイルの指定



区切りをTabにする  
ヘッダー無



変数型を自動取得

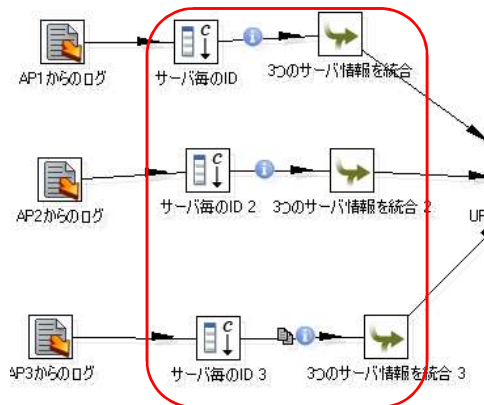


QC勉強会資料

## 3つのサーバからのログを統合する

2009年11月5日

- サーバごとのIDをつける
- 統合の前処理を行う



QC勉強会資料

## 3つのサーバからのログを統合する

2009年11月5日

- サーバごとにIDを付与
- ID用のフィールド「SID」
- サーバごとのID「AP1」「AP2」「AP3」

Add constant values

ステップ名: サーバ毎のID

#	フィールド名	データタイプ	書式	長さ	精度	通貨記号	桁区切り文字	数値囲み文字	値
1	SID	String							AP1

OK(O) キャンセル(C)

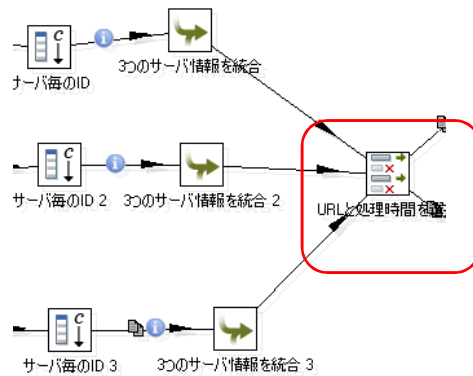
QC勉強会資料

## 必要な部分のみを抽出

2009年11月5日

- 必要な部分のみを抽出

- リクエストURL
- リクエスト元のURL
- 処理時間
- SID

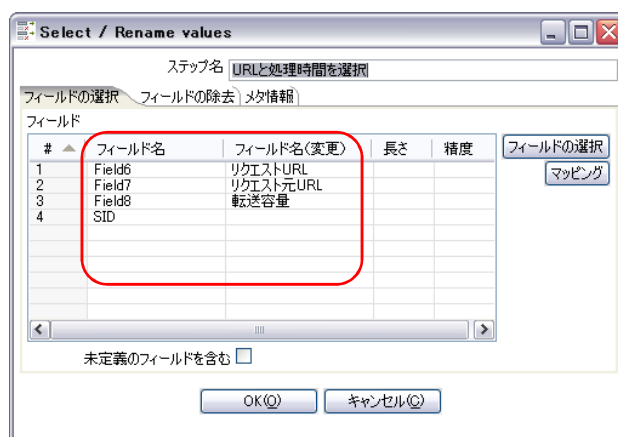


勉強会資料

15.

## 必要な部分のみを抽出

2009年11月5日



変数を選択  
変数名を変更

QC勉強会資料

16.

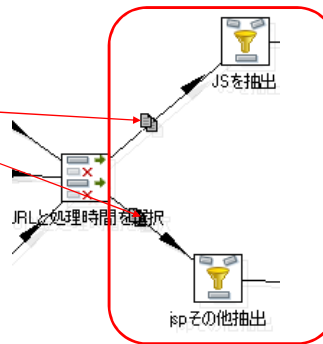


## プログラム形式ごとに分岐

2009年11月5日

- JSプログラム
  - JSPプログラム+その他
- を分岐
- URLの形式が異なるため
- これに対する前処理

同じストリーム(データ)  
をコピーする



QC勉強会資料

## プログラム形式ごとに分岐

2009年11月5日



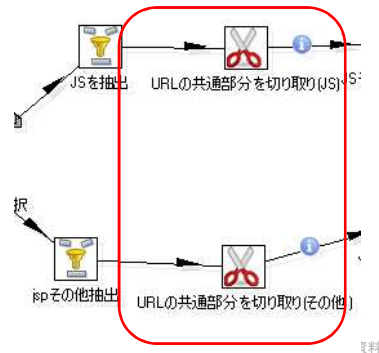
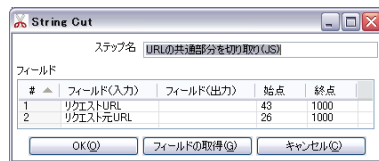
フィルタを行う条件を設定  
AND、OR、NOR NOT等の  
論理演算も可能

QC勉強会資料

# STRING(文字データ)の切り取り

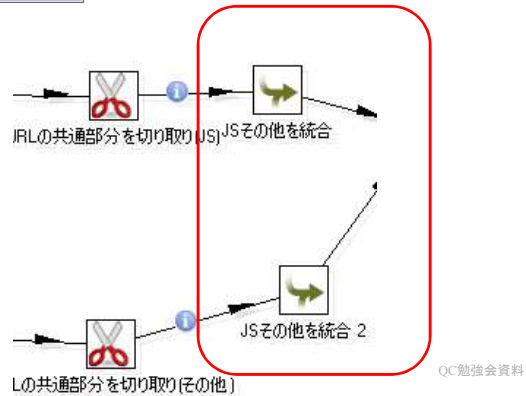
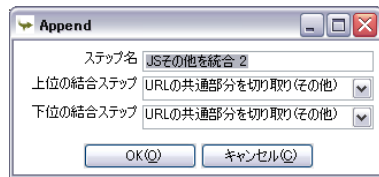
2009年11月5日

- 文字データの一部を切り取り



# 分岐したストリームを再統合

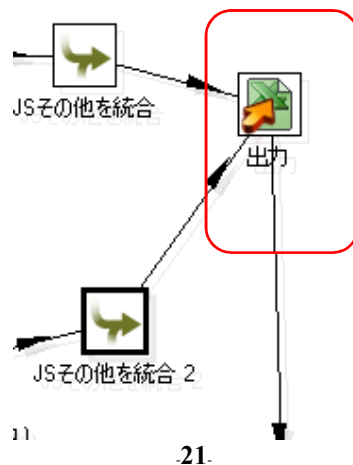
2009年11月5日



## 結果をExcelで出力する

2009年11月5日

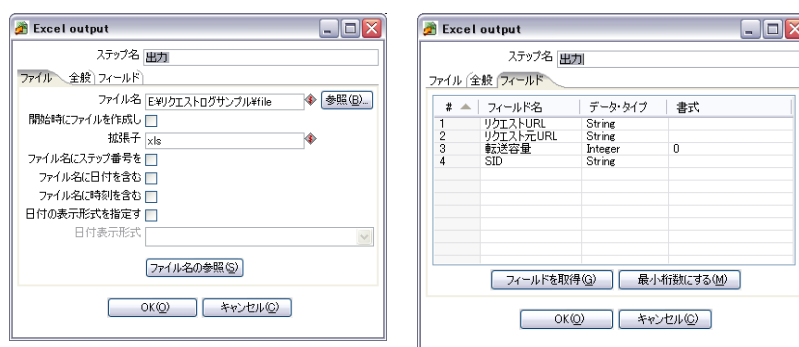
- 加工した結果を出力する
- ログデータの統合→必要な項目の抽出



QC勉強会資料

## 結果をExcelで出力する①

2009年11月5日



- 項目の選択、ファイル名の変更が可能

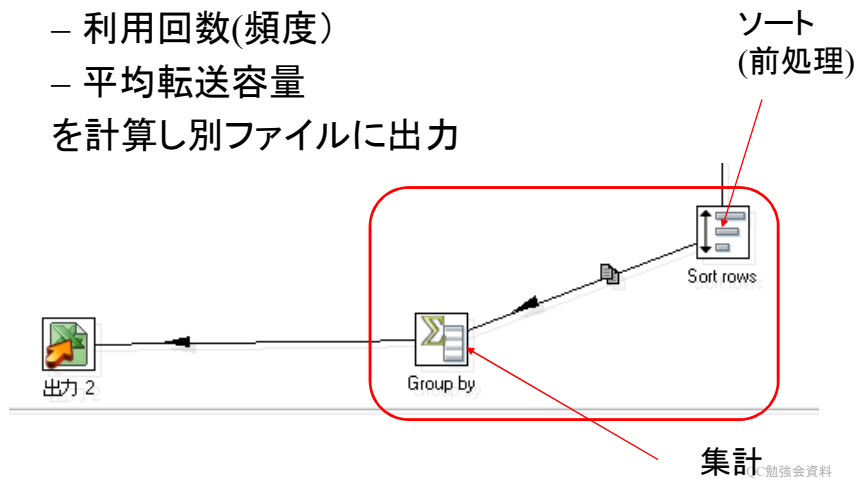
QC勉強会資料

22.

# プログラムごとに集計する

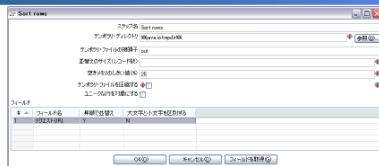
2009年11月5日

- プログラムごとに
  - 利用回数(頻度)
  - 平均転送容量
 を計算し別ファイルに出力

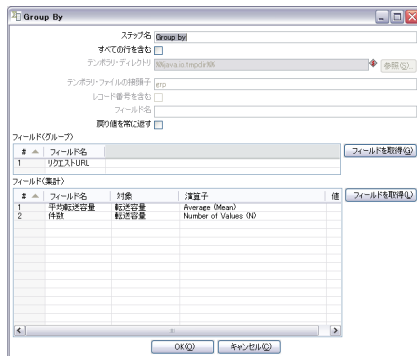


# プログラムごとに集計する

2009年11月5日



- ソートしたい変数
  - リクエストURL



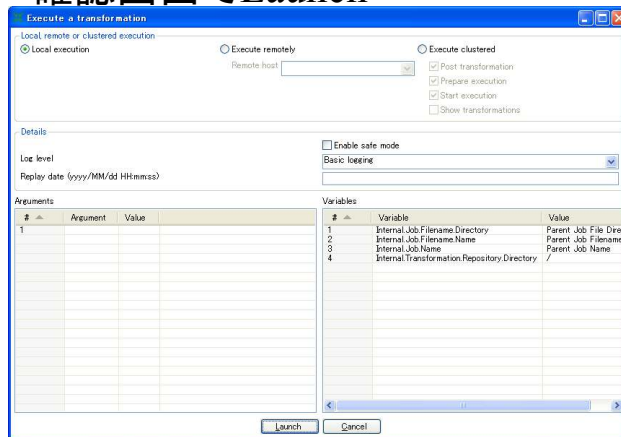
- 集計対象
  - 転送容量の平均
  - (任意の項目)のカウント→頻度

QC勉強会資料

# 実行

2009年11月5日

- メニューバーの「Run」ボタン
- 確認画面でLaunch



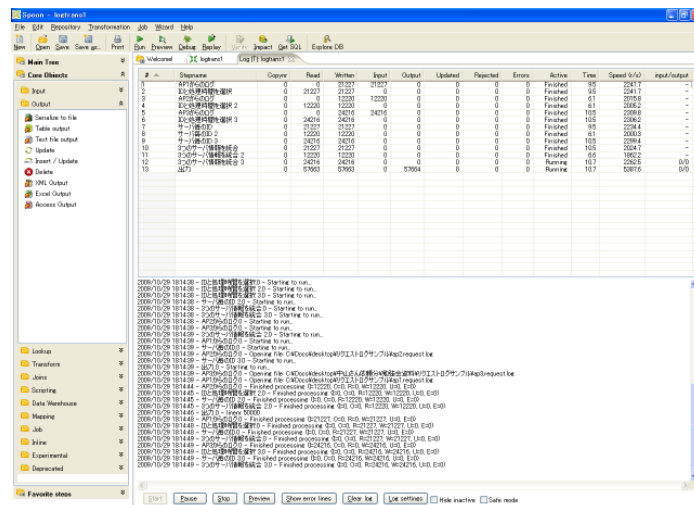
OC勉強会資料

25.

# 実行

2009年11月5日

- 実行ログの表示



# 実行

2009年11月5日

リクエストURL	リクエストURL	転送容量	件数
http://192.168.1.10/mactools/http/192.168.1.10/mactools		5,531.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		17,258.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		4,525.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		6,250.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		3,735.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		17,252.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		5,758.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		6,341.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		4,433.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		6,522.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		7,484.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		7,652.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		5,707.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		3,737.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		9,207.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		3,737.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		6,337.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		6,454.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		9,207.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		6,500.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		9,207.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		12,579.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		8,622.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		6,202.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		3,810.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		5,843.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		4,806.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		6,500.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		6,500.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		3,864.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		3,864.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		4,262.00 AP	1
http://192.168.1.10/mactools/http/192.168.1.10/mactools		4,262.00 AP	1

リクエストURL	平均転送容量	件数
activepsse@qvm@qvm	10781	1
*activepsse@qvm@qvm	2460	1
tol@psse@qvm@qvm21	750	2
lpw-ad@search_call_ser	343	1
lpw-application_apply_f	523	2
lpw-applied_body_call_i	3621	1
lpw-applied_frame_call	62	1
lpw-process_instance	531	6
lpw-process_instance	90	6
lpw-process_instance	283	6
lpw-switch_link_custom	523	2
lpw-switch_link_jpress	422	2
lm2100 menu	130	6
lm200 menu	141	1
/portal-main_menu_servic	653	13
/portal-main_top_service	62	13
/portal-main_top_bar_ser	239	13
/portal-portal_display_ser	239	13
/portal-session_keeper_c	102	13
/portal/portal/portal_frame	198	13
psse portal	198	26
user login	674	13
psse@lpw@qvm@qvm	6070	125
psse@qvm@qvmcommon	594	813
psse@qvm@qvmcommon	2627	634
psse@qvm@qvm2110	2633	639
psse@qvm@qvm2100	714	915



時刻	送信元 IP	送信元ポート	宛先 IP	宛先ポート	プロトコル	転送容量	件数
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	5531	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	17258	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	4525	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	6250	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	3735	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	17252	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	5758	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	6341	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	4433	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	6522	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	7484	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	7652	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	5707	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	3737	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	9207	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	3737	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	6337	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	6454	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	9207	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	6500	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	9207	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	12579	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	8622	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	6202	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	3810	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	5843	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	4806	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	6500	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	6500	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	3864	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	3864	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	4262	1
2009-11-05 11:05:15.15	192.168.1.10	49152	192.168.1.10	80	HTTP	4262	1

- 大量／複数のデータを
- GUIで
- 必要な部分のみ加工

27.

QC勉強会資料

# 全体のヒストグラム

2009年11月5日

- QCツール→QCヒストグラム

**QCヒストグラム**

変数 (1つ選択)

転送容量

層別のプロット...

以下、\*の項目は層別のヒストグラムには適用されません

ヒストグラムの描き方のオプション

規格の上限 (USL) <無し>

規格の下限 (LSL) <無し>

\* 小数点以下の桁数 2

\* データの測定単位 <auto>

\* 区間の幅 <auto>

ヒストグラムへの追加: 選択する場合、チェック

\* 統計量

推定密度のグラフ

正規密度のグラフ

OK キャンセル ヘルプ

QC勉強会資料

## データの解析

2009年11月5日

- Pentahoで加工した2つのファイル
  - アクセスログの統合データ
    - リクエストURL、リクエスト元URL、SID、転送容量
  - プログラムごとの集計データ
    - リクエストURL、平均転送容量、頻度
- Rコマンダー+QCパッケージでの実行
  - はずれ値の判定
  - はずれ値を起こす要因の解析

QC勉強会資料

## Rコマンダーの起動

2009年11月5日

- Binの下にあるRgui.exeを起動
- 設定してあれば、自動的にRコマンダーが起動

QC勉強会資料

## データの取り込み

2009年11月5日

- メニュー 「データ」→「データのインポート」  
→「Excel・・・」を選択
- データセットに名前をつける
  - 「全体」「頻度」
- クリップボードからのインポートも可能

QC勉強会資料

## 全体のヒストグラム

2009年11月5日

- メニュー 「QCツール」→「QCヒストグラム」
  - デフォルトのままOK
  - 層別から「SID」を選択してOK

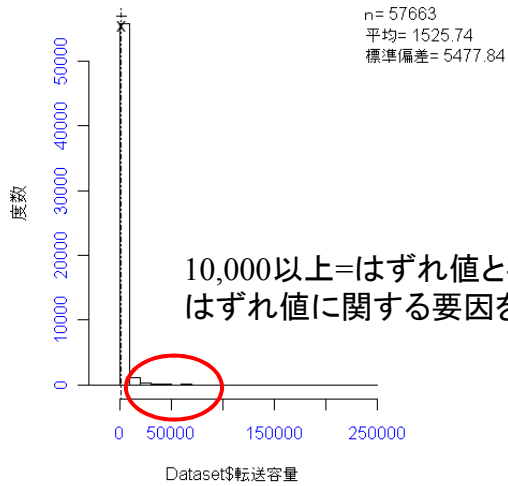
QC勉強会資料



# 全体のヒストグラム

2009年11月5日

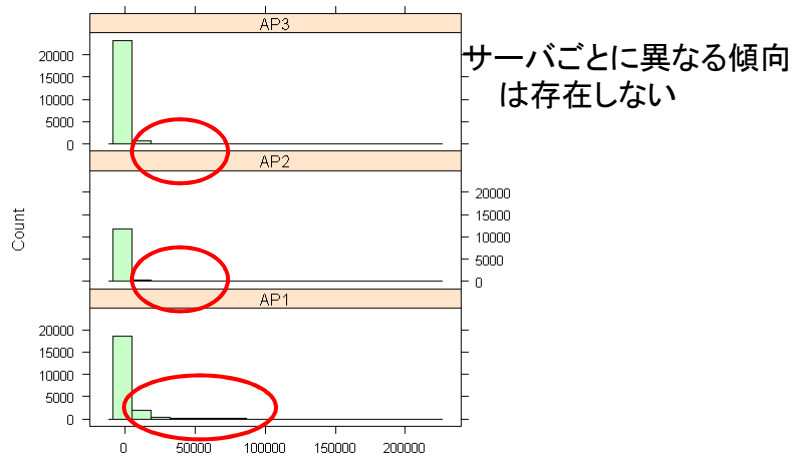
Dataset\$転送容量のヒストグラム



QC勉強会資料

# 全体のヒストグラム一層別

2009年11月5日

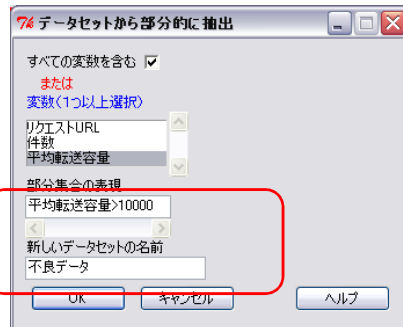


QC勉強会資料

## データの絞込み

2009年11月5日

- 転送容量が10,000以上のデータははずれ値(=不良の発生要因)とみなす
- データセットの中ではずれ値を絞り込む
- 「頻度」データセットの絞りこみ
- メニュー「アクティブデータセット」→「絞込み」

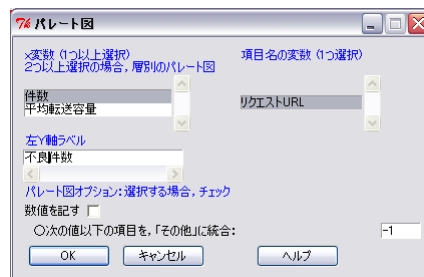


QC勉強会資料

## パレート図の作成

2009年11月5日

- 不良(レスポンスの低下)の発生頻度が多いプログラムは何か？
- メニュー「QCツール」→「パレート図」



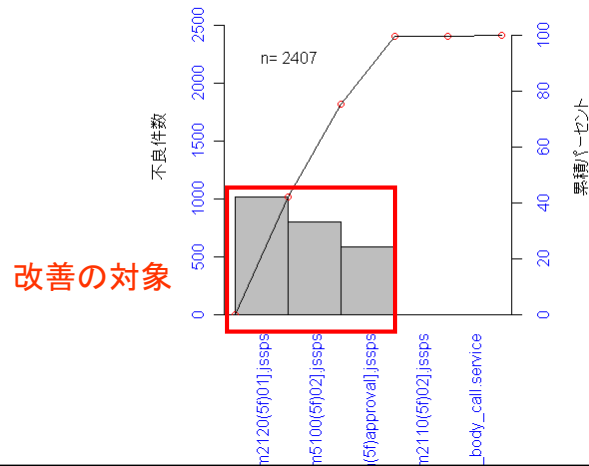
QC勉強会資料

## パレート図の作成

2009年11月5日

- 不良のほぼ100%近くが3つのプログラムによるものだった

件数のパレート図



QC勉強会資料

## まとめ

2009年11月5日

- SQCの概略と管理の4要素
- Pentahoによるデータ加工方法
- Rを用いたSQCの概略(デモ)

QC勉強会資料