Contenido

[Dataset Description: SCF-1 3](#_Toc387158263)

[1 Introduction 3](#_Toc387158264)

[1.1 Dataset family context 3](#_Toc387158265)

[1.2 List of data sets in family 3](#_Toc387158266)

[2 Dataset 1: SCF-1-1-raw 4](#_Toc387158267)

[2.1 Context 4](#_Toc387158268)

[2.2 Overview 4](#_Toc387158269)

[2.2.1 Internet folder 5](#_Toc387158270)

[2.2.2 Intranet folder 7](#_Toc387158271)

[2.3 Content 9](#_Toc387158272)

[2.3.1 CPU monitoring file 9](#_Toc387158273)

[2.3.2 HD monitoring file 9](#_Toc387158274)

[2.3.3 Interface monitoring file 9](#_Toc387158275)

[2.3.4 Event log file 10](#_Toc387158276)

[2.4 Format examples 12](#_Toc387158277)

[2.4.1 CPU.txt 12](#_Toc387158278)

[2.4.2 HD.txt 12](#_Toc387158279)

[2.4.3 INT.txt 13](#_Toc387158280)

[2.4.4 MEM.txt 13](#_Toc387158281)

[2.4.5 EV.csv 13](#_Toc387158282)

[2.5 Issues 14](#_Toc387158283)

[2.5.1 XLSX files 14](#_Toc387158284)

[2.5.2 List of empty xlsx files 14](#_Toc387158285)

[2.5.3 Issue 1b: Hidden XSLX 15](#_Toc387158286)

[2.5.4 Solution 15](#_Toc387158287)

[3 Dataset 2: SCF-1-2-aggrDateFolder 17](#_Toc387158288)

[3.1 Context 17](#_Toc387158289)

[3.2 Overview 17](#_Toc387158290)

[3.3 Content 17](#_Toc387158291)

[3.4 Format 17](#_Toc387158292)

[3.5 Generation 17](#_Toc387158293)

[3.6 Comments 17](#_Toc387158294)

[4 Dataset 3: SCF-1-3-aggrDateFiles 18](#_Toc387158295)

[4.1 Context 18](#_Toc387158296)

[4.2 Overview 18](#_Toc387158297)

[4.3 Content 19](#_Toc387158298)

[4.4 Format 19](#_Toc387158299)

[4.4.1 Event log format 19](#_Toc387158300)

[4.5 Generation 20](#_Toc387158301)

[4.6 [Comments] 23](#_Toc387158302)

[5 Dataset 4: SCF-1-4-visualization-tables 24](#_Toc387158303)

[5.1 Context 24](#_Toc387158304)

[5.2 Overview 24](#_Toc387158305)

[5.3 Content 24](#_Toc387158306)

[5.4 Format 24](#_Toc387158307)

[5.5 [Generation] 24](#_Toc387158308)

[5.6 [Comments] 26](#_Toc387158309)

[5.6.1 Parsing trouble 26](#_Toc387158310)

[5.6.2 Issue: strange resources? 27](#_Toc387158311)

[5.6.3 Future code cleanup 29](#_Toc387158312)

[6 Dataset N 31](#_Toc387158313)

[6.1 Context 31](#_Toc387158314)

[6.2 Overview 31](#_Toc387158315)

[6.3 Content 31](#_Toc387158316)

[6.4 Format 31](#_Toc387158317)

[6.5 [Generation] 31](#_Toc387158318)

[6.6 [Comments] 31](#_Toc387158319)

[7 Dataset analysis 32](#_Toc387158320)

[7.1 List of failure events 32](#_Toc387158321)

[7.1.1 Subset 1: 27/11 to 13/12 32](#_Toc387158322)

[7.2 Failure analysis 35](#_Toc387158323)

[7.2.1 Analysis exploration 1 35](#_Toc387158324)

[7.2.2 Analysis Exploration 2 - Failure list, 15 days 37](#_Toc387158325)

[7.3 Issues 46](#_Toc387158326)

[7.3.1 Hypothesis 46](#_Toc387158327)

# Dataset Description: SCF-1

1 Introduction

1.1 Dataset family context

1.2 List of data sets in family

2 Dataset description

2.1 Dataset 1

2.1.1 Context

2.1.2 Overview

2.1.3 Content

2.1.4 Format

2.1.5 [Generation]

2.1.6 [Comments]

2.2 Dataset N

3 Dataset analysis

# Introduction

Dataset family SCF-1 consists of a list of log files from Santander Consumer Finance, received from 01/02/2014. Last updated in 20/02/2014

## Dataset family context

The scenario for this dataset family is described in the document "SCF-1-Scenario" (TODO)

## List of data sets in family

* **SCF-1-1-raw** - raw data sample, regularly updated (starting on 01/02/2014)
  + **SCF-1-2-aggrDateFolder** - raw data, aggregating all date folders into single folder tree.
* **SCF-1-3-aggrDateFile** - processed data, aggregating files from different dates, and generating tidy tables

# Dataset 1: SCF-1-1-raw

Dataset Totta-1-0-raw includes the original raw files provided for the scenario Totta-1. Starts on 01/02/2014

## Context

The scenario for this dataset family is described in the document "SCF-1-Scenario".

## Overview

Data is available on Drive folder: L3P3/SC Europa

There is a separate data folder for each sampled day. The description below correponds to folder on 01/02/2014, but the same structure is shared across all folders.

2 data subfolders, Internet, Intranet

<https://drive.google.com/a/centeropenmiddleware.com/?usp=folder#folders/0ByYvzC04rW27MldHSldzRlVmZ1k>

### Internet folder

includes

* 94 files
* 8,21 MB total

Each file records 24 hours of data.

List of files:

[.] DSWPEUWT02.sceu.corp\_EV\_010214.csv

[..] DSWPEUWT02.sceu.corp\_INT\_010214.csv

ASWPEUIB01.sceu.corp\_CPU\_010214.csv DSWPEUWT02.sceu.corp\_MEM\_010214.csv

ASWPEUIB01.sceu.corp\_EV\_010214.csv DSWPEUWT03.sceu.corp\_CPU\_010214.csv

ASWPEUIB01.sceu.corp\_INT\_010214.csv DSWPEUWT03.sceu.corp\_EV\_010214.csv

ASWPEUIB01.sceu.corp\_MEM\_010214.csv DSWPEUWT03.sceu.corp\_INT\_010214.csv

ASWPEUIB02.sceu.corp\_CPU\_010214.csv DSWPEUWT03.sceu.corp\_MEM\_010214.csv

ASWPEUIB02.sceu.corp\_EV\_010214.csv DSWPEUWT04.sceu.corp\_CPU\_010214.csv

ASWPEUIB02.sceu.corp\_INT\_010214.csv DSWPEUWT04.sceu.corp\_EV\_010214.csv

ASWPEUIB02.sceu.corp\_MEM\_010214.csv DSWPEUWT04.sceu.corp\_INT\_010214.csv

ASWPEUIB03.sceu.corp\_CPU\_010214.csv DSWPEUWT04.sceu.corp\_MEM\_010214.csv

ASWPEUIB03.sceu.corp\_EV\_010214.csv FWPEUIB01\_CPU\_010214.csv

ASWPEUIB03.sceu.corp\_INT\_010214.csv FWPEUIB01\_EV\_010214.csv

ASWPEUIB03.sceu.corp\_MEM\_010214.csv FWPEUIB01\_HD\_010214.csv

ASWPEUIB04.sceu.corp\_CPU\_010214.csv FWPEUIB01\_INT\_010214.csv

ASWPEUIB04.sceu.corp\_EV\_010214.csv FWPEUIB01\_MEM\_010214.csv

ASWPEUIB04.sceu.corp\_INT\_010214.csv FWPEUIB02\_CPU\_010214.csv

ASWPEUIB04.sceu.corp\_MEM\_010214.csv FWPEUIB02\_EV\_010214.csv

ASWPEUID01.sceu.corp\_CPU\_010214.csv FWPEUIB02\_HD\_010214.csv

ASWPEUID01.sceu.corp\_EV\_010214.csv FWPEUIB02\_INT\_010214.csv

ASWPEUID01.sceu.corp\_INT\_010214.csv FWPEUIB02\_MEM\_010214.csv

ASWPEUID01.sceu.corp\_MEM\_010214.csv FWPEUID01\_CPU\_010214.csv

ASWPEUID02.sceu.corp\_CPU\_010214.csv FWPEUID01\_EV\_010214.csv

ASWPEUID02.sceu.corp\_EV\_010214.csv FWPEUID01\_HD\_010214.csv

ASWPEUID02.sceu.corp\_INT\_010214.csv FWPEUID01\_INT\_010214.csv

ASWPEUID02.sceu.corp\_MEM\_010214.csv FWPEUID01\_MEM\_010214.csv

ASWPEUID03.sceu.corp\_CPU\_010214.csv FWPEUID02\_CPU\_010214.csv

ASWPEUID03.sceu.corp\_EV\_010214.csv FWPEUID02\_EV\_010214.csv

ASWPEUID03.sceu.corp\_INT\_010214.csv FWPEUID02\_HD\_010214.csv

ASWPEUID03.sceu.corp\_MEM\_010214.csv FWPEUID02\_INT\_010214.csv

ASWPEUID04.sceu.corp\_CPU\_010214.csv FWPEUID02\_MEM\_010214.csv

ASWPEUID04.sceu.corp\_EV\_010214.csv lbpeuib01.sceu.corp\_CPU\_010214.csv

ASWPEUID04.sceu.corp\_INT\_010214.csv lbpeuib01.sceu.corp\_EV\_010214.csv

ASWPEUID04.sceu.corp\_MEM\_010214.csv lbpeuib01.sceu.corp\_INT\_010214.csv

dnsautpeuid01.sceu.corp\_CPU\_010214.csv lbpeuib01.sceu.corp\_MEM\_010214.csv

dnsautpeuid01.sceu.corp\_EV\_010214.csv lbpeuib02.sceu.corp\_CPU\_010214.csv

dnsautpeuid01.sceu.corp\_HD\_010214.csv lbpeuib02.sceu.corp\_EV\_010214.csv

dnsautpeuid01.sceu.corp\_INT\_010214.csv lbpeuib02.sceu.corp\_INT\_010214.csv

dnsautpeuid01.sceu.corp\_MEM\_010214.csv lbpeuib02.sceu.corp\_MEM\_010214.csv

dnsautpeuid02.sceu.corp\_CPU\_010214.csv lbpeuid01.sceu.corp\_CPU\_010214.csv

dnsautpeuid02.sceu.corp\_EV\_010214.csv lbpeuid01.sceu.corp\_EV\_010214.csv

dnsautpeuid02.sceu.corp\_HD\_010214.csv lbpeuid01.sceu.corp\_INT\_010214.csv

dnsautpeuid02.sceu.corp\_INT\_010214.csv lbpeuid01.sceu.corp\_MEM\_010214.csv

dnsautpeuid02.sceu.corp\_MEM\_010214.csv lbpeuid02.sceu.corp\_CPU\_010214.csv

DSWPEUWT01.sceu.corp\_CPU\_010214.csv lbpeuid02.sceu.corp\_EV\_010214.csv

DSWPEUWT01.sceu.corp\_EV\_010214.csv lbpeuid02.sceu.corp\_INT\_010214.csv

DSWPEUWT01.sceu.corp\_INT\_010214.csv lbpeuid02.sceu.corp\_MEM\_010214.csv

DSWPEUWT01.sceu.corp\_MEM\_010214.csv

DSWPEUWT02.sceu.corp\_CPU\_010214.csv

Structured list of files: 22 Nodes in 8 categories

* ASWPEUIB
  + 01.sceu.corp: CPU, EV, INT, MEM
  + 02.sceu.corp: CPU, EV, INT, MEM
  + 03.sceu.corp: CPU, EV, INT, MEM
  + 04.sceu.corp: CPU, EV, INT, MEM
* ASWPEUID
  + 01.sceu.corp: CPU, EV, INT, MEM
  + 02.sceu.corp: CPU, EV, INT, MEM
  + 03.sceu.corp: CPU, EV, INT, MEM
  + 04.sceu.corp: CPU, EV, INT, MEM
* dnsautpeuid
  + 01.sceu.corp: CPU, EV, HD, INT, MEM
  + 02.sceu.corp: CPU, EV, HD, INT, MEM
* DSWPEUWT
  + 01.sceu.corp: CPU, EV, INT, MEM
  + 02.sceu.corp: CPU, EV, INT, MEM
  + 03.sceu.corp: CPU, EV, INT, MEM
  + 04.sceu.corp: CPU, EV, INT, MEM
* FWPEUIB
  + 01: CPU, EV, HD, INT, MEM
  + 02: CPU, EV, HD, INT, MEM
* FWPEUID01
  + 01: CPU, EV, HD, INT, MEM
  + 02: CPU, EV, HD, INT, MEM
* lbpeuib
  + 01.sceu.corp: CPU, EV, HD, INT, MEM
  + 02.sceu.corp: CPU, EV, HD, INT, MEM
* lbpeuid
  + 01.sceu.corp: CPU, EV, HD, INT, MEM
  + 02.sceu.corp: CPU, EV, HD, INT, MEM

For each node, the following five files are provided:

* CPU\_151113.csv
* EV\_151113.csv
* HD\_151113.csv // May be missing
* INT\_151113.csv
* MEM\_151113.csv

### Intranet folder

Data is available on Drive folder: L3P3/NET

The folder includes

* 66 files
* 5.61MB total

[.] dnsrespeuin01.sceu.corp\_EV\_010214.xlsx

[..] dnsrespeuin01.sceu.corp\_HD\_010214.csv

ASWPEUIN01.sceu.corp\_CPU\_010214.csv dnsrespeuin01.sceu.corp\_INT\_010214.csv

ASWPEUIN01.sceu.corp\_EV\_010214.xlsx dnsrespeuin01.sceu.corp\_MEM\_010214.csv

ASWPEUIN01.sceu.corp\_INT\_010214.csv dnsrespeuin02.sceu.corp\_CPU\_010214.csv

ASWPEUIN01.sceu.corp\_MEM\_010214.csv dnsrespeuin02.sceu.corp\_EV\_010214.xlsx

ASWPEUIN02.sceu.corp\_CPU\_010214.csv dnsrespeuin02.sceu.corp\_HD\_010214.csv

ASWPEUIN02.sceu.corp\_EV\_010214.xlsx dnsrespeuin02.sceu.corp\_INT\_010214.csv

ASWPEUIN02.sceu.corp\_INT\_010214.csv dnsrespeuin02.sceu.corp\_MEM\_010214.csv

ASWPEUIN02.sceu.corp\_MEM\_010214.csv DSWPEUIN01.sceu.corp\_CPU\_010214.csv

ASWPEUIN03.sceu.corp\_CPU\_010214.csv DSWPEUIN01.sceu.corp\_EV\_010214.xlsx

ASWPEUIN03.sceu.corp\_EV\_010214.xlsx DSWPEUIN01.sceu.corp\_INT\_010214.csv

ASWPEUIN03.sceu.corp\_INT\_010214.csv DSWPEUIN01.sceu.corp\_MEM\_010214.csv

ASWPEUIN03.sceu.corp\_MEM\_010214.csv DSWPEUIN02.sceu.corp\_CPU\_010214.csv

ASWPEUIN04.sceu.corp\_CPU\_010214.csv DSWPEUIN02.sceu.corp\_EV\_010214.xlsx

ASWPEUIN04.sceu.corp\_EV\_010214.xlsx DSWPEUIN02.sceu.corp\_INT\_010214.csv

ASWPEUIN04.sceu.corp\_INT\_010214.csv DSWPEUIN02.sceu.corp\_MEM\_010214.csv

ASWPEUIN04.sceu.corp\_MEM\_010214.csv FWPEUIN01\_CPU\_010214.csv

ASWPEUIN05.sceu.corp\_CPU\_010214.csv FWPEUIN01\_EV\_010214.xlsx

ASWPEUIN05.sceu.corp\_EV\_010214.xlsx FWPEUIN01\_INT\_010214.csv

ASWPEUIN05.sceu.corp\_INT\_010214.csv FWPEUIN01\_MEM\_010214.csv

ASWPEUIN05.sceu.corp\_MEM\_010214.csv FWPEUIN02\_CPU\_010214.csv

ASWPEUIN06.sceu.corp\_CPU\_010214.csv FWPEUIN02\_EV\_010214.xlsx

ASWPEUIN06.sceu.corp\_EV\_010214.xlsx FWPEUIN02\_INT\_010214.csv

ASWPEUIN06.sceu.corp\_INT\_010214.csv FWPEUIN02\_MEM\_010214.csv

ASWPEUIN06.sceu.corp\_MEM\_010214.csv lbpeuin01.sceu.corp\_CPU\_010214.csv

ASWPEUIN07.sceu.corp\_CPU\_010214.csv lbpeuin01.sceu.corp\_EV\_010214.csv

ASWPEUIN07.sceu.corp\_EV\_010214.xlsx lbpeuin01.sceu.corp\_INT\_010214.csv

ASWPEUIN07.sceu.corp\_INT\_010214.csv lbpeuin01.sceu.corp\_MEM\_010214.csv

ASWPEUIN07.sceu.corp\_MEM\_010214.csv lbpeuin02.sceu.corp\_CPU\_010214.csv

ASWPEUIN08.sceu.corp\_CPU\_010214.csv lbpeuin02.sceu.corp\_EV\_010214.csv

ASWPEUIN08.sceu.corp\_EV\_010214.xlsx lbpeuin02.sceu.corp\_INT\_010214.csv

ASWPEUIN08.sceu.corp\_INT\_010214.csv lbpeuin02.sceu.corp\_MEM\_010214.csv

ASWPEUIN08.sceu.corp\_MEM\_010214.csv

dnsrespeuin01.sceu.corp\_CPU\_010214.csv

Structured list of files: 16 Nodes in 5 categories

* ASWPEUIN
  + 01.sceu.corp: CPU, EV, INT, MEM
  + 02.sceu.corp: CPU, EV, INT, MEM
  + 03.sceu.corp: CPU, EV, INT, MEM
  + 04.sceu.corp: CPU, EV, INT, MEM
  + 05.sceu.corp: CPU, EV, INT, MEM
  + 06.sceu.corp: CPU, EV, INT, MEM
  + 07.sceu.corp: CPU, EV, INT, MEM
  + 08.sceu.corp: CPU, EV, INT, MEM
* dnsrespeuin01
  + 01.sceu.corp: CPU, EV, HD, INT, MEM
  + 02.sceu.corp: CPU, EV, HD, INT, MEM
* DSWPEUIN01
  + 01.sceu.corp: CPU, EV, INT, MEM
  + 02.sceu.corp: CPU, EV, INT, MEM
* FWPEUIN01.
  + 01.sceu.corp: CPU, EV, INT, MEM
  + 02.sceu.corp: CPU, EV, INT, MEM
* lbpeuin01
  + 01.sceu.corp: CPU, EV, INT, MEM
  + 02.sceu.corp: CPU, EV, INT, MEM

For each node, the following five files are provided:

* CPU\_151113.csv
* EV\_151113.csv
* HD\_151113.csv // May be missing
* INT\_151113.csv
* MEM\_151113.csv

## Content

Five file types

* cpu: cpu monitoring
* mem: memory monitoring
* int: interface monitoring
* hd: hd monitoring
* ev: event logs

### CPU monitoring file

Several csv tables on file.

Common headers

* report period
* view title
* node id
* resolution

Table 1: Average usage - Data fields

* date
* average cpu usage %

Tables 2-52: Baseline Average - Data fields

* date
* average cpu usage % compared to baseline

### HD monitoring file

Several csv tables on file, one per partition.

Optional file, missing for some nodes (which lack a HD).

Common headers

* report period
* view title
* node id
* resolution

Table N (partition id) - Data fields

* date
* average hd usage %

### Interface monitoring file

Several csv tables on file, two per interface (in/out)

* In
* Out

Common headers

* report period
* view title
* node id

Interface Headers

* interface ID
* resolution

Table N - Data fields

* date
* average interface usage %

### Event log file

Single csv table per file

Data fields

* Severity
* Created On
* Name
* Event
* Created By
* Cleared On
* Cleared By
* Event Type
* Model.Type.Name
* Event.Precedence

#### Severity

Factor variable. Indicates importance of event.

Can have 4 values:

* ""
* Minor
* Major
* Critical

**Events with Severity "Minor" "Major", or "Critical" are considered failure events.**

#### Created On

Date variable. Indicates event timestamp.

#### Name

Factor variable. Indicates name of node generating the event

#### Event

Text variable. Textual description of event.

#### Created By

Factor variable. Indicates name of module generating the event.

Can have values:

* System
* spectrum

#### Cleared On

Date variable. Indicates date for clearing of event, if any. Can be "".

#### Cleared By

Factor variable. Indicates name of module clearing the event. Can be "".

Possible values:

* ""
* System
* cncc@vmwtca01
* cncc@vmwtca02
* x025606@vmwtca02
* x097520@vmwtca02
* x099978@vmwtca01
* x099978@vmwtca02

#### Event Type

Factor variable. Indicates numeric identifier for type of event, in hexadecimal.

Optional variable, missing in some files.

Possible values (37):

* "BLANK"
* "t.0x1001c"
* "t.0x1001d"
* "t.0x10150"
* "t.0x10301"
* "t.0x10302"
* "t.0x10d00"
* [8] "t.0x10d0c"
* "t.0x10d30"
* "t.0x10d35"
* "t.0x10daa"
* "t.0x10f02"
* "t.0x10f03"
* "t.0x10fab"
* [15] "t.0x110a03"
* "t.0x110a04"
* t.0x210017"
* "t.0x210026"
* "t.0x2104c7"
* "t.0x220001"
* "t.0x220002"
* [22] "t.0x220101"
* "t.0x220102"
* "t.0x3b7000a"
* "t.0x3b7000b"
* "t.0x3b70012"
* "t.0x3b70013"
* "t.0x3b70014"
* [29] "t.0x3b70015"
* "t.0x3b70016"
* "t.0x3b70017"
* "t.0x3b70019"
* "t.0x3b7002d"
* "t.0x480016"
* "t.0x5420102"
* [36] "t.0x5420103"
* "t.0xd0002"

#### Model.Type.Name

Factor variable, missing in most lines. Can have the following values

* ""
* F5BigIPDev
* Host\_Device
* Rtr\_Cisco
* SwCiscoIOS

#### Event.Precedence

Numeric variable, missing in most lines. Can have the following values

* ""
* 10

## Format examples

### CPU.txt

From bvtolestrup / 01.totta.corp / CPU

// Headers

"Report Period:,""0:00:00 - 23:59:00 November 15 2013 Central European Time"""

"View Title:,""Average CPU Utilization Trend with Events"""

"Subtitle:,""Server: bvtolestrup01.totta.corp"""

5-Minute Resolution +/- 1.0 Std Dev

"Utilization - Average,""The average value of the operational metric Utilization"""

"Utilization - Baseline Average,""The baseline measurement records the average (arithmetic mean) of the baseline series values across similar time periods. Baselines track the change in performance over time, which can be used to predict future performance."""

(x51)

// Data

"[Time Zone: CET],""Utilization - Average"""

"11/15/2013 0:00 CET,""4 Percent"""

Comments: The observed file includes 51 blocks. The first block shows the average value per 5 minutes. The remaining blocks show baselane averages, with a sample per hour. Each of the 50 baseline averages has the same structure and shows no unique identifiers, but has different data over the same time period. We will request clarification about the meaning of this.

### HD.txt

From bvtolestrup / 01.totta.corp /HD

."Report Period:,""0:00:00 - 23:59:00 November 15 2013 Central European Time"""

"View Title:,""IM MultiTrend (Device) - Partition - Percent Used - Average - Descending"""

"Subtitle:,""Server: bvtolestrup01.totta.corp"""

289 Samples (5-Minute Resolution)

"bvtolestrup01.totta.corp : Partition: 25,""The average value of the operational metric Percent Used"""

// Data

"[Time Zone: CET],""bvtolestrup01.totta.corp : Partition: 25"""

"11/15/2013 0:00 CET,""76 Percent"""

Comments: This file records HD usage, on a per partition basis, sampled every 5 minutes. There are separate data blocks for each partition. this particular file records 12 partitions.

### INT.txt

From bvtolestrup / 01.totta.corp /INT

"Report Period:,""0:00:00 - 23:59:00 November 15 2013 Central European Time"""

"View Title:,""Top Interface Utilization - MultiView (Trend) - Utilization by Percentage - In and Out - Descending"""

"Subtitle:,""Server: bvtolestrup01.totta.corp"""

eth0

289 Samples (5-Minute Resolution) (4.3 Gbps-4.3 Gbps)

eth0

"Percent Utilization In,""The average value of the operational metric Utilization In"""

"TimeStamp,""Percent Utilization In"""

"11/15/2013 0:00 CET,""< 0.01 Percent"""

"Percent Utilization Out,""The average value of the operational metric Utilization Out"""

"TimeStamp,""Percent Utilization Out"""

"11/15/2013 0:00 CET,""< 0.01 Percent"""

eth1

Comments: 2 interfaces, eth0, eth1. This file records interface usage, as a percentage, separated by in and out traffic.

### MEM.txt

From bvtolestrup / 01.totta.corp /MEM

"Report Period:,""0:00:00 - 23:59:00 November 15 2013 Central European Time"""

"View Title:,""Average Memory Utilization Trend with Events"""

"Subtitle:,""Server: bvtolestrup01.totta.corp"""

5-Minute Resolution +/- 1.0 Std Dev

"Utilization - Average,""The average value of the operational metric Utilization"""

"[Time Zone: CET],""Utilization - Average"""

"11/15/2013 0:00 CET,""18.43 Percent"""

"Utilization - Baseline Average,""The baseline measurement records the average (arithmetic mean) of the baseline series values across similar time periods. Baselines track the change in performance over time, which can be used to predict future performance."""

"[Time Zone: CET],""Utilization - Baseline Average"""

Comments: The observed file includes 51 blocks. The first block shows the average value per 5 minutes. The remaining blocks show baselane averages, with a sample per hour. Each of the 50 baseline averages has the same structure and shows no unique identifiers, but has different data over the same time period. We will request clarification about the meaning of this.

### EV.csv

The event file format was the same used for event files in the totta dataset. We reproduce an example from totta.

#### bvtowpocwp01.central.rinterna.local\_EV\_151113.csv

"Severity","Created On","Name","Event","Created By","Cleared On","Cleared By"

"Major","15-nov-2013 5:26:32 CET","bvtowpocwp01.central.rinterna.local","High CPU Utilization. The utilization of 98% for CPU instance 2 named 'CPU: 2' has exceeded the 85% threshold on model bvtowpocwp01.central.rinterna.local for more than the acceptable time period.","System","15-nov-2013 5:26:32 CET","System"

"","15-nov-2013 10:23:44 CET","bvtowpocwp01.central.rinterna.local","Security string for model bvtowpocwp01.central.rinterna.local, has been changed from ADMIN to BOASIS.","System","",""

"","15-nov-2013 5:26:32 CET","bvtowpocwp01.central.rinterna.local","Normal CPU Utilization. The CPU Utilization for all CPU instances is now below the 70% reset threshold for model bvtowpocwp01.central.rinterna.local.","System","",""

## Issues

### XLSX files

Some event files (between 05/02 and 12/02) use the .xlsx format rather than .csv. We need to convert them to .csv.

* This affects
  + 94 files out of a total of 3k files (out of 675 event files)
  + Only event files (not monitoring)
  + Only **empty files**: 9KB files without content
    - 8.62KB
    - 8829 bytes

In Windows, can search using the pattern:

* tamaño:8829bytes ".csv"
* No false positives found

Suggested solution: replace them with empty .csv's with same file name

R code

file.info("file")$size

### List of empty xlsx files

The search shows 243 files with the pattern

> destFiles[which(file.info(files)$size==8829)]

[1] "internet/aswpeuib01.sceu.corp\_ev\_010214.csv" "internet/aswpeuib02.sceu.corp\_ev\_010214.csv"

[3] "internet/aswpeuib03.sceu.corp\_ev\_010214.csv" "internet/aswpeuib04.sceu.corp\_ev\_010214.csv"

[5] "internet/aswpeuid01.sceu.corp\_ev\_010214.csv" "internet/aswpeuid02.sceu.corp\_ev\_010214.csv"

[7] "internet/aswpeuid03.sceu.corp\_ev\_010214.csv" "internet/aswpeuid04.sceu.corp\_ev\_010214.csv"

[9] "internet/dnsautpeuid01.sceu.corp\_ev\_010214.csv" "internet/dnsautpeuid02.sceu.corp\_ev\_010214.csv"

[11] "internet/dswpeuwt01.sceu.corp\_ev\_010214.csv" "internet/dswpeuwt02.sceu.corp\_ev\_010214.csv"

[13] "internet/dswpeuwt03.sceu.corp\_ev\_010214.csv" "internet/dswpeuwt04.sceu.corp\_ev\_010214.csv"

[15] "internet/lbpeuib01.sceu.corp\_ev\_010214.csv" "internet/lbpeuib02.sceu.corp\_ev\_010214.csv"

[17] "intranet/aswpeuin01.sceu.corp\_ev\_010214.xlsx" "intranet/aswpeuin02.sceu.corp\_ev\_010214.xlsx"

[19] "intranet/aswpeuin03.sceu.corp\_ev\_010214.xlsx" "intranet/aswpeuin04.sceu.corp\_ev\_010214.xlsx"

[21] "intranet/aswpeuin05.sceu.corp\_ev\_010214.xlsx" "intranet/aswpeuin06.sceu.corp\_ev\_010214.xlsx"

[23] "intranet/aswpeuin07.sceu.corp\_ev\_010214.xlsx" "intranet/aswpeuin08.sceu.corp\_ev\_010214.xlsx"

[25] "intranet/dnsrespeuin01.sceu.corp\_ev\_010214.xlsx" "intranet/dnsrespeuin02.sceu.corp\_ev\_010214.xlsx"

[27] "intranet/dswpeuin01.sceu.corp\_ev\_010214.xlsx" "intranet/dswpeuin02.sceu.corp\_ev\_010214.xlsx"

[29] "intranet/fwpeuin01\_ev\_010214.xlsx" "intranet/fwpeuin02\_ev\_010214.xlsx"

[31] "internet/aswpeuib01.sceu.corp\_ev\_020214.xlsx" "internet/aswpeuib02.sceu.corp\_ev\_020214.xlsx"

[33] "internet/aswpeuib03.sceu.corp\_ev\_020214.xlsx" "internet/aswpeuib04.sceu.corp\_ev\_020214.xlsx"

[35] "internet/aswpeuid01.sceu.corp\_ev\_020214.xlsx" "internet/aswpeuid02.sceu.corp\_ev\_020214.xlsx"

[37] "internet/aswpeuid03.sceu.corp\_ev\_020214.xlsx" "internet/aswpeuid04.sceu.corp\_ev\_020214.xlsx"

[39] "internet/dnsautpeuid01.sceu.corp\_ev\_020214.xlsx" "internet/dnsautpeuid02.sceu.corp\_ev\_020214.csv"

[41] "internet/dswpeuwt01.sceu.corp\_ev\_020214.xlsx" "internet/dswpeuwt02.sceu.corp\_ev\_020214.xlsx"

[43] "internet/dswpeuwt03.sceu.corp\_ev\_020214.xlsx" "internet/dswpeuwt04.sceu.corp\_ev\_020214.xlsx"

[45] "internet/lbpeuib01.sceu.corp\_ev\_020214.xlsx" "internet/lbpeuib02.sceu.corp\_ev\_020214.xlsx"

[47] "intranet/aswpeuin01.sceu.corp\_ev\_020214.xlsx" "intranet/aswpeuin02.sceu.corp\_ev\_020214.xlsx"

[49] "intranet/aswpeuin03.sceu.corp\_ev\_020214.xlsx" "intranet/aswpeuin04.sceu.corp\_ev\_020214.xlsx"

[51] "intranet/aswpeuin05.sceu.corp\_ev\_020214.xlsx" "intranet/aswpeuin06.sceu.corp\_ev\_020214.xlsx"

[53] "intranet/aswpeuin07.sceu.corp\_ev\_020214.xlsx" "intranet/aswpeuin08.sceu.corp\_ev\_020214.xlsx"

[55] "intranet/dnsrespeuin01.sceu.corp\_ev\_020214.xlsx" "intranet/dnsrespeuin02.sceu.corp\_ev\_020214.xlsx"

[57] "intranet/dswpeuin01.sceu.corp\_ev\_020214.xlsx" "intranet/dswpeuin02.sceu.corp\_ev\_020214.xlsx"

[59] "intranet/fwpeuin01\_ev\_020214.xlsx" "intranet/fwpeuin02\_ev\_020214.xlsx"

[61] "internet/aswpeuib01.sceu.corp\_ev\_030214.xlsx" "internet/aswpeuib02.sceu.corp\_ev\_030214.xlsx"

[63] "internet/aswpeuib03.sceu.corp\_ev\_030214.xlsx" "internet/aswpeuib04.sceu.corp\_ev\_030214.xlsx"

[65] "internet/aswpeuid01.sceu.corp\_ev\_030214.xlsx" "internet/aswpeuid02.sceu.corp\_ev\_030214.xlsx"

[67] "internet/aswpeuid03.sceu.corp\_ev\_030214.xlsx" "internet/aswpeuid04.sceu.corp\_ev\_030214.xlsx"

[69] "internet/dnsautpeuid01.sceu.corp\_ev\_030214.xlsx" "internet/dnsautpeuid02.sceu.corp\_ev\_030214.csv"

[71] "internet/dswpeuwt01.sceu.corp\_ev\_030214.xlsx" "internet/dswpeuwt02.sceu.corp\_ev\_030214.xlsx"

[73] "internet/dswpeuwt03.sceu.corp\_ev\_030214.xlsx" "internet/dswpeuwt04.sceu.corp\_ev\_030214.xlsx"

[75] "internet/lbpeuib01.sceu.corp\_ev\_030214.xlsx" "internet/lbpeuib02.sceu.corp\_ev\_030214.xlsx"

[77] "intranet/aswpeuin01.sceu.corp\_ev\_030214.xlsx" "intranet/aswpeuin02.sceu.corp\_ev\_030214.xlsx"

[79] "intranet/aswpeuin03.sceu.corp\_ev\_030214.xlsx" "intranet/aswpeuin04.sceu.corp\_ev\_030214.xlsx"

[81] "intranet/aswpeuin05.sceu.corp\_ev\_030214.xlsx" "intranet/aswpeuin06.sceu.corp\_ev\_030214.xlsx"

[83] "intranet/aswpeuin07.sceu.corp\_ev\_030214.xlsx" "intranet/aswpeuin08.sceu.corp\_ev\_030214.xlsx"

[85] "intranet/dnsrespeuin01.sceu.corp\_ev\_030214.xlsx" "intranet/dnsrespeuin02.sceu.corp\_ev\_030214.xlsx"

[87] "intranet/dswpeuin01.sceu.corp\_ev\_030214.xlsx" "intranet/dswpeuin02.sceu.corp\_ev\_030214.xlsx"

[89] "intranet/fwpeuin01\_ev\_030214.xlsx" "intranet/fwpeuin02\_ev\_030214.xlsx"

[91] "internet/aswpeuib01.sceu.corp\_ev\_040214.xlsx" "internet/aswpeuib02.sceu.corp\_ev\_040214.xlsx"

[93] "internet/aswpeuib03.sceu.corp\_ev\_040214.xlsx" "internet/aswpeuib04.sceu.corp\_ev\_040214.xlsx"

[95] "internet/aswpeuid01.sceu.corp\_ev\_040214.xlsx" "internet/aswpeuid02.sceu.corp\_ev\_040214.xlsx"

[97] "internet/aswpeuid03.sceu.corp\_ev\_040214.xlsx" "internet/aswpeuid04.sceu.corp\_ev\_040214.xlsx"

[99] "internet/dnsautpeuid01.sceu.corp\_ev\_040214.xlsx" "internet/dnsautpeuid02.sceu.corp\_ev\_040214.csv"

[101] "internet/dswpeuwt02.sceu.corp\_ev\_040214.xlsx" "internet/dswpeuwt03.sceu.corp\_ev\_040214.xlsx"

[103] "internet/dswpeuwt04.sceu.corp\_ev\_040214.xlsx" "internet/lbpeuib01.sceu.corp\_ev\_040214.xlsx"

[105] "internet/lbpeuib02.sceu.corp\_ev\_040214.csv" "intranet/aswpeuin01.sceu.corp\_ev\_040214.xlsx"

[107] "intranet/aswpeuin02.sceu.corp\_ev\_040214.xlsx" "intranet/aswpeuin03.sceu.corp\_ev\_040214.xlsx"

[109] "intranet/aswpeuin04.sceu.corp\_ev\_040214.xlsx" "intranet/aswpeuin05.sceu.corp\_ev\_040214.xlsx"

[111] "intranet/aswpeuin06.sceu.corp\_ev\_040214.xlsx" "intranet/aswpeuin07.sceu.corp\_ev\_040214.xlsx"

[113] "intranet/aswpeuin08.sceu.corp\_ev\_040214.xlsx" "intranet/dnsrespeuin01.sceu.corp\_ev\_040214.xlsx"

[115] "intranet/dnsrespeuin02.sceu.corp\_ev\_040214.xlsx" "intranet/dswpeuin01.sceu.corp\_ev\_040214.xlsx"

[117] "intranet/dswpeuin02.sceu.corp\_ev\_040214.xlsx" "intranet/fwpeuin01\_ev\_040214.xlsx"

[119] "intranet/fwpeuin02\_ev\_040214.xlsx" "internet/aswpeuib01.sceu.corp\_ev\_050214.xlsx"

[121] "internet/aswpeuib02.sceu.corp\_ev\_050214.xlsx" "internet/aswpeuib03.sceu.corp\_ev\_050214.xlsx"

[123] "internet/aswpeuib04.sceu.corp\_ev\_050214.xlsx" "internet/aswpeuid01.sceu.corp\_ev\_050214.xlsx"

[125] "internet/aswpeuid02.sceu.corp\_ev\_050214.xlsx" "internet/aswpeuid03.sceu.corp\_ev\_050214.xlsx"

[127] "internet/aswpeuid04.sceu.corp\_ev\_050214.xlsx" "internet/dnsautpeuid01.sceu.corp\_ev\_050214.csv"

[129] "internet/dnsautpeuid02.sceu.corp\_ev\_050214.csv" "internet/lbpeuib01.sceu.corp\_ev\_050214.xlsx"

[131] "internet/lbpeuib02.sceu.corp\_ev\_050214.xlsx" "intranet/aswpeuin01.sceu.corp\_ev\_050214.xlsx"

[133] "intranet/aswpeuin02.sceu.corp\_ev\_050214.xlsx" "intranet/aswpeuin03.sceu.corp\_ev\_050214.xlsx"

[135] "intranet/aswpeuin04.sceu.corp\_ev\_050214.xlsx" "intranet/aswpeuin05.sceu.corp\_ev\_050214.xlsx"

[137] "intranet/aswpeuin06.sceu.corp\_ev\_050214.xlsx" "intranet/aswpeuin07.sceu.corp\_ev\_050214.xlsx"

[139] "intranet/aswpeuin08.sceu.corp\_ev\_050214.xlsx" "intranet/dnsrespeuin01.sceu.corp\_ev\_050214.xlsx"

[141] "intranet/dnsrespeuin02.sceu.corp\_ev\_050214.xlsx" "intranet/dswpeuin02.sceu.corp\_ev\_050214.xlsx"

[143] "intranet/fwpeuin01\_ev\_050214.xlsx" "intranet/fwpeuin02\_ev\_050214.xlsx"

[145] "internet/aswpeuib01.sceu.corp\_ev\_060214.xlsx" "internet/aswpeuib02.sceu.corp\_ev\_060214.xlsx"

[147] "internet/aswpeuib03.sceu.corp\_ev\_060214.xlsx" "internet/aswpeuib04.sceu.corp\_ev\_060214.xlsx"

[149] "internet/aswpeuid01.sceu.corp\_ev\_060214.xlsx" "internet/aswpeuid02.sceu.corp\_ev\_060214.xlsx"

[151] "internet/aswpeuid03.sceu.corp\_ev\_060214.xlsx" "internet/aswpeuid04.sceu.corp\_ev\_060214.xlsx"

[153] "internet/dnsautpeuid01.sceu.corp\_ev\_060214.csv" "internet/dnsautpeuid02.sceu.corp\_ev\_060214.csv"

[155] "internet/lbpeuib01.sceu.corp\_ev\_060214.xlsx" "internet/lbpeuib02.sceu.corp\_ev\_060214.xlsx"

[157] "intranet/aswpeuin01.sceu.corp\_ev\_060214.xlsx" "intranet/aswpeuin02.sceu.corp\_ev\_060214.xlsx"

[159] "intranet/aswpeuin03.sceu.corp\_ev\_060214.xlsx" "intranet/aswpeuin04.sceu.corp\_ev\_060214.xlsx"

[161] "intranet/aswpeuin05.sceu.corp\_ev\_060214.xlsx" "intranet/aswpeuin06.sceu.corp\_ev\_060214.xlsx"

[163] "intranet/aswpeuin07.sceu.corp\_ev\_060214.xlsx" "intranet/aswpeuin08.sceu.corp\_ev\_060214.xlsx"

[165] "intranet/dnsrespeuin01.sceu.corp\_ev\_060214.xlsx" "intranet/dnsrespeuin02.sceu.corp\_ev\_060214.xlsx"

[167] "intranet/dswpeuin01.sceu.corp\_ev\_060214.xlsx" "intranet/dswpeuin02.sceu.corp\_ev\_060214.xlsx"

[169] "intranet/fwpeuin01\_ev\_060214.xlsx" "intranet/fwpeuin02\_ev\_060214.xlsx"

[171] "internet/lbpeuib01.sceu.corp\_ev\_070214.xlsx" "internet/lbpeuib02.sceu.corp\_ev\_070214.xlsx"

[173] "intranet/aswpeuin01.sceu.corp\_ev\_070214.xlsx" "intranet/aswpeuin02.sceu.corp\_ev\_070214.xlsx"

[175] "intranet/aswpeuin03.sceu.corp\_ev\_070214.xlsx" "intranet/aswpeuin04.sceu.corp\_ev\_070214.xlsx"

[177] "intranet/aswpeuin05.sceu.corp\_ev\_070214.xlsx" "intranet/aswpeuin06.sceu.corp\_ev\_070214.xlsx"

[179] "intranet/aswpeuin07.sceu.corp\_ev\_070214.xlsx" "intranet/aswpeuin08.sceu.corp\_ev\_070214.xlsx"

[181] "intranet/dnsrespeuin01.sceu.corp\_ev\_070214.xlsx" "intranet/dnsrespeuin02.sceu.corp\_ev\_070214.xlsx"

[183] "intranet/dswpeuin01.sceu.corp\_ev\_070214.xlsx" "intranet/dswpeuin02.sceu.corp\_ev\_070214.xlsx"

[185] "intranet/aswpeuin01.sceu.corp\_ev\_080214.xlsx" "intranet/aswpeuin02.sceu.corp\_ev\_080214.xlsx"

[187] "intranet/aswpeuin03.sceu.corp\_ev\_080214.xlsx" "intranet/aswpeuin04.sceu.corp\_ev\_080214.xlsx"

[189] "intranet/aswpeuin05.sceu.corp\_ev\_080214.xlsx" "intranet/aswpeuin06.sceu.corp\_ev\_080214.xlsx"

[191] "intranet/aswpeuin07.sceu.corp\_ev\_080214.xlsx" "intranet/aswpeuin08.sceu.corp\_ev\_080214.xlsx"

[193] "intranet/dnsrespeuin01.sceu.corp\_ev\_080214.xlsx" "intranet/dnsrespeuin02.sceu.corp\_ev\_080214.xlsx"

[195] "intranet/dswpeuin01.sceu.corp\_ev\_080214.xlsx" "intranet/dswpeuin02.sceu.corp\_ev\_080214.xlsx"

[197] "intranet/aswpeuin01.sceu.corp\_ev\_090214.xlsx" "intranet/aswpeuin02.sceu.corp\_ev\_090214.xlsx"

[199] "intranet/aswpeuin03.sceu.corp\_ev\_090214.xlsx" "intranet/aswpeuin04.sceu.corp\_ev\_090214.xlsx"

[201] "intranet/aswpeuin05.sceu.corp\_ev\_090214.xlsx" "intranet/aswpeuin06.sceu.corp\_ev\_090214.xlsx"

[203] "intranet/aswpeuin07.sceu.corp\_ev\_090214.xlsx" "intranet/aswpeuin08.sceu.corp\_ev\_090214.xlsx"

[205] "intranet/dnsrespeuin01.sceu.corp\_ev\_090214.xlsx" "intranet/dnsrespeuin02.sceu.corp\_ev\_090214.xlsx"

[207] "intranet/dswpeuin01.sceu.corp\_ev\_090214.xlsx" "intranet/dswpeuin02.sceu.corp\_ev\_090214.xlsx"

[209] "intranet/aswpeuin01.sceu.corp\_ev\_100214.xlsx" "intranet/aswpeuin02.sceu.corp\_ev\_100214.xlsx"

[211] "intranet/aswpeuin03.sceu.corp\_ev\_100214.xlsx" "intranet/aswpeuin04.sceu.corp\_ev\_100214.xlsx"

[213] "intranet/aswpeuin05.sceu.corp\_ev\_100214.xlsx" "intranet/aswpeuin06.sceu.corp\_ev\_100214.xlsx"

[215] "intranet/aswpeuin07.sceu.corp\_ev\_100214.xlsx" "intranet/aswpeuin08.sceu.corp\_ev\_100214.xlsx"

[217] "intranet/dnsrespeuin01.sceu.corp\_ev\_100214.xlsx" "intranet/dnsrespeuin02.sceu.corp\_ev\_100214.xlsx"

[219] "intranet/dswpeuin01.sceu.corp\_ev\_100214.xlsx" "intranet/dswpeuin02.sceu.corp\_ev\_100214.xlsx"

[221] "intranet/aswpeuin01.sceu.corp\_ev\_110214.xlsx" "intranet/aswpeuin02.sceu.corp\_ev\_110214.xlsx"

[223] "intranet/aswpeuin03.sceu.corp\_ev\_110214.xlsx" "intranet/aswpeuin04.sceu.corp\_ev\_110214.xlsx"

[225] "intranet/aswpeuin05.sceu.corp\_ev\_110214.xlsx" "intranet/aswpeuin06.sceu.corp\_ev\_110214.xlsx"

[227] "intranet/aswpeuin07.sceu.corp\_ev\_110214.xlsx" "intranet/aswpeuin08.sceu.corp\_ev\_110214.xlsx"

[229] "intranet/dnsrespeuin01.sceu.corp\_ev\_110214.xlsx" "intranet/dnsrespeuin02.sceu.corp\_ev\_110214.xlsx"

[231] "intranet/dswpeuin01.sceu.corp\_ev\_110214.xlsx" "intranet/dswpeuin02.sceu.corp\_ev\_110214.xlsx"

[233] "intranet/aswpeuin01.sceu.corp\_ev\_120214.xlsx" "intranet/aswpeuin02.sceu.corp\_ev\_120214.xlsx"

[235] "intranet/aswpeuin03.sceu.corp\_ev\_120214.xlsx" "intranet/aswpeuin04.sceu.corp\_ev\_120214.xlsx"

[237] "intranet/aswpeuin05.sceu.corp\_ev\_120214.xlsx" "intranet/aswpeuin06.sceu.corp\_ev\_120214.xlsx"

[239] "intranet/aswpeuin07.sceu.corp\_ev\_120214.xlsx" "intranet/aswpeuin08.sceu.corp\_ev\_120214.xlsx"

[241] "intranet/dnsrespeuin01.sceu.corp\_ev\_120214.xlsx" "intranet/dnsrespeuin02.sceu.corp\_ev\_120214.xlsx"

[243] "intranet/dswpeuin01.sceu.corp\_ev\_120214.xlsx" "intranet/dswpeuin02.sceu.corp\_ev\_120214.xlsx"

### Issue 1b: Hidden XSLX

Some event files have .csv extension, but still have .xlsx format.

These files are between 01/02 and 10/02

### Solution

This solution is applied as part of the process to generate dataset 2 (aggregated by folder)

Modify the script to aggregate files by date folder, with the following lines

# - XLSX removal: Sometimes event files are in the form of empty, 8829bytes XLSX files

## This can also include XLSX files with csv extension

## Solution: Identify these files, replace them with empty .csv files

...

#code to address empty xlsx files

emptyXlsxIndices<-which(file.info(files)$size==8829)

dotXlsxIndices<-which(grepl(".xlsx",files))

nonEmptyXlsxInd<-which(!dotXlsxIndices %in% emptyXlsxIndices) #should be integer(0)

destFiles<-gsub(".xlsx",".csv", destFiles)

#create temporary empty file

emptyFileName<-file.path(aggrDateFolder,"emptyFile.txt")

writeLines("",emptyFileName)

#replace names of empty xlsx files with emptyfilename

files[emptyXlsxIndices]<-emptyFileName

# Dataset 2: SCF-1-2-aggrDateFolder

## Context

Derived from dataset SCF-1-1-raw. Includes the same files as that dataset, but under a new folder structure, aggregating all files from different dates under a single common folder tree.

## Overview

## Content

## Format

## Generation

Generated from dataset SCF-1-1-raw, using script totta-1-1-aggr-1-aggregateDateFolders.R

Code 1SCF-1-1-aggr-1-aggregateDateFolders.R

#This script aggregates all different date folders for data into a single one

# In addition, it performs the following operations not needed for the totta dataset

# - XLSX removal: Sometimes event files are in the form of empty, 8829bytes XLSX files

## This can also include XLSX files with csv extension

## Solution: Identify these files, replace them with empty .csv files

baseDataPath <- "C:/Users/capelastegui/workspace/OFP/SCF/1-Data"

rawFolder <- file.path(baseDataPath,"1-Raw/1-Current")

aggrDateFolder <- file.path(baseDataPath,"2-AggrDateFolder")

aggrDateFile <- file.path(baseDataPath,"3-AggrDateFile")

folderList<-list.dirs(rawFolder, recursive=FALSE)

folders<-character(0)

files<-character(0)

destFiles<-character(0)

subFolders<-character(0)

for(f in folderList)

{

folders<- unique(c(folders,list.dirs(f,recursive=TRUE,full.names=FALSE)))

files<-c(files,list.files(f,recursive=TRUE, include.dirs=FALSE,full.names=TRUE))

destFiles<-tolower(c(destFiles,list.files(f,recursive=TRUE, include.dirs=FALSE,full.names=FALSE)))

subFolders<-c(subFolders,list.files(f,recursive=TRUE, include.dirs=FALSE,full.names=FALSE))

}

#create folders to copy

folders1<-file.path(aggrDateFolder,tolower(folders))

tmp<-lapply(folders1,dir.create,showWarnings=FALSE)

folders2<-file.path(aggrDateFile,folders)

tmp<-lapply(folders2,dir.create,showWarnings=FALSE)

#code to address empty xlsx files

emptyXlsxIndices<-which(file.info(files)$size==8829)

dotXlsxIndices<-which(grepl(".xlsx",files))

nonEmptyXlsxInd<-which(!dotXlsxIndices %in% emptyXlsxIndices) #should be integer(0)

destFiles<-gsub(".xlsx",".csv", destFiles)

#create temporary empty file

emptyFileName<-file.path(aggrDateFolder,"emptyFile.txt")

writeLines("",emptyFileName)

#replace names of empty xlsx files with emptyfilename

files[emptyXlsxIndices]<-emptyFileName

#copy all files to aggregated dir

tmp<-mapply(file.copy,files,file.path(aggrDateFolder,destFiles), overwrite=FALSE)

## Comments

This dataset is not used directly. It is an intermediate processing step towards dataset 3.

# Dataset 3: SCF-1-3-aggrDateFiles

## Context

Derived from dataset SCF-1-2-aggrDateFolder. Parses all data files, aggregates all files from a same node and different dates into a single table, and generates a new file with that table.

## Overview

Data for this dataset is written into folder "3-AggrDateFile". For each node, all data from each data source for all sampled dates is aggregated into a single file. The current file list reads as follows.

(... TO DO)

For R processing purposes, data from these files is structured as follows:

* **logTable**, a list composed of **node** sublists
  + A **node** list is composed of **resource** data frames
    - A **resource** dataframe includes resource monitoring data for a given node
    - Currently supported resources: CPU,MEM,HD,INT
* **eventTable,** a dataframe including all log event data
* **debugTable,** a list with node sublists and resource elements, including information about missing data and data that could not be parsed.

## Content

The content of files matches that of analogous files in Dataset 2. However, for each (node,resource) pair, data is provided in a single file. In addition, data format has been adapted, as we describe in the following section.

## Format

The format of tables and files for monitoring and event data has been modified with respect to that of raw data from Dataset 1. The new format and changes are detailed here.

### Event log format

Each event log line now has 8 fields:

* date : POSIXct date object
* severity : Factor with 4 levels: BLANK, Critical, Major, Minor
* type : Factor
* node : Factor
* createdBy : Factor
* clearedOn : POSIXct date object
* clearedBy : Factor
* msg : Factor

These match the fields in the raw log, with the following changes:

* For factor variables, a "BLANK" level has been introduced, to aggregate previously missing values, "" values, and otherwise NA values.
* All 8 fields are now always present. Previously, some were optional. Previously missing values are now labeled as "BLANK".
* Renamed fields for clarity
  + **Name** is now **node**
  + **Created On** is now **date**
  + **Severity** is now **severity**
  + **Event** is now **msg**
  + **Event.Type** is now **type**
* Fields have been reordered to improve readability.

## Generation

This dataset is generated using 2 R scripts: "SCF-1-0-raw-2-resourceParser.R" and "SCF-1-2-aggr-1-aggregateDateFiles.R". The scripts are applied over dataset 2, SCF-1-2-aggrDateFolder

This script provides a reusable parser for monitoring data and event logs.

Code 2 SCF-1-0-raw-2-resourceParser.R

library(stringr)

library(gsubfn)

library(plyr)

csvToDf <- function (csvLines, skip=0)

{

#regex "([^"]\*)" identifies blocks delimited by "

regex <- "\"([^\\\"]\*)\""

if(skip>0)csvLines<-csvLines[-skip]

csvLines<-csvLines[csvLines!=""]

if(length(csvLines)<1) {return(data.frame())}

df<-data.frame(t(strapplyc(csvLines,regex,simplify=TRUE)))

names(df)<-laply(df[1,],as.character)

df<-df[-1,]

return(df)

}

resourceParser <- function(resource)

{

if(is.null(resource)|length(resource)<2){return(data.frame())}

breaks<-cut(1:length(resource),which(nchar(resource)==0))

splitResource<-split(resource,breaks)

dfList<-llply(splitResource, csvToDf, skip=1)

dfList<-dfList[-1]

#check dfList includes data frames, filter empty ones

dfList<-dfList[laply(dfList,is.data.frame)]

dfListDims<-laply(dfList,dim)

if(!is.null(dim(dfListDims)))

{dfList<-dfList[dfListDims[,1]!=0]} else

{return (data.frame())}

#{dfList<-dfList[0]}

#llply(dfList,head)

#for partition column, remove all but last word from column name

prevNameList<-list(0)

for(i in 1:length(dfList))

{

#name<-tail(unlist(strsplit(names(dfList[[i]])," ")),1)

nameList<-laply(strsplit(names(dfList[[i]][-1])," "),tail,n=1)

for(name in nameList)

{

if(any(prevNameList==name)){name<-paste(name,i,sep="\_")}

prevNameList[[name]]<-name

}

names(dfList[[i]])[-1]<-name

names(dfList[[i]])[1]<-"date"

}

#join- plyr function to merge 2 dataframes

#Reduce - function to apply binary function (join) to vector

df<-Reduce(function(d1,d2){join(d1,d2,by="date")},dfList)

df$date<-strptime(df$date,format="%m/%d/%Y %H:%M")

for(i in 2:length(df))

{

fac<-df[,i]

levels(fac)<-gsub(" ","",levels(fac))

levels(fac)<-gsub("Percent","",levels(fac))

levels(fac)<-gsub("<","",levels(fac))

df[,i]<-as.numeric(as.character(fac))

}

return(df)

}

# Parses a totta event log file. Returns a table with non-factored character vectors.

eventParser <- function(file\_name)

{

#events<-read.csv(file\_name,stringsAsFactors=FALSE)

events<-read.table(file\_name, header=TRUE, sep=",", colClasses="character")

#If not enough columns, return (should give error)

##TODO: find a way to send an error here!

#if(ncol(events)<4){return(NULL)}

if(ncol(events)<4){return(data.frame(parseError=file\_name))}

if(nrow(events)<1){return(data.frame(emptyLog=file\_name))}

Sys.setlocale("LC\_TIME","Spanish\_Spain.1252")

date<-as.POSIXct(strptime(events$Created.On,format="%d-%b-%Y %H:%M:%S"))

dateClear<-as.POSIXct(strptime(events$Cleared.On,format="%d-%b-%Y %H:%M:%S"))

#check if valid dates

if(length(date)==nrow(events) || any(is.na(date))==FALSE)

{events$Created.On<-date} else

{return(data.frame(parseError=file\_name))}

if(length(dateClear)==nrow(events)){events$Cleared.On<-dateClear}

if(!is.null(events$Event.Type)) {events$Event.Type<- paste("t.",(events$Event.Type),sep="")}

return(events)

}

This scripts applies the previously-defined parsers over dataset 2, to generate dataset 3.

Code 3 totta-1-1-aggr-1-aggregateDateFiles.R

#This script aggregates all different date folders for data into a single one

# In addition, it performs the following operations not needed for the totta dataset

# - Multi-line events:

## Event files use LF instead of CRLF for line feed

## Also, messages can include line feeds and unescaped quotes

library(plyr)

library(gsubfn)

library(doParallel)

library(foreach)

baseRPath <- "C:/Users/capelastegui/workspace/OFP/SCF/2-R/preprocess"

resourceParserPath <- file.path(baseRPath,"SCF-1-0-raw-2-resourceParser.R")

source(resourceParserPath)

# For some reason related to scope, we need to redeclare this here, or parallel calls from plyr will crash

# Function to take totta log/event files from a single folder tree, parse them, and save the aggregated results

aggregateDataFiles <- function(doLogs=FALSE, doEvents=TRUE, saveToFile=FALSE)

{

# File path variables. Modify as needed.

baseDataPath <- "C:/Users/capelastegui/workspace/OFP/SCF/1-Data"

rawFolder <- file.path(baseDataPath,"1-Raw/1-Current")

aggrDateFolder <- file.path(baseDataPath,"2-AggrDateFolder")

aggrDateFile <- file.path(baseDataPath,"3-AggrDateFile")

eventTableFile <- file.path(aggrDateFile,"eventTable.txt")

eventRawFile <- file.path(aggrDateFile,"eventRaw.txt")

# Generate filesTable, a dataframe listing log files, and their associated nodes and resources

files<-list.files(aggrDateFolder,recursive=TRUE, include.dirs=FALSE,full.names=TRUE)

filesCropped<-gsub(aggrDateFolder, "", files)

# regex to split node, resource type, and date component of file name (has form node\_type\_ddmmyy.csv)

regex <- "(.\*)\_([^\_]\*)\_[^\_]\*$"

filesTable<-data.frame(strapplyc(filesCropped,regex, simplify = "rbind"))

names(filesTable)<-c("node","resource")

filesTable$file<-files[grep(regex,files)]

#for debugging purposes, list of files with names not matching expected pattern

wrongNameFiles<-files[-grep(regex,files)]

# Generate rawDataList, eventDataList

## rawDataList: list tree with raw data from all log files

## eventDataList: list tree with raw data from all event files

# function to apply readlines over an array of file names

# return list with file contents; names of list elements are file names

applyReadLines<-function(l)

{

myReadLines<-function(file)

{

lines<-readLines(file)

#remove content if single, empty line

#if(length(lines)==1&nchar(lines[1]==0)){lines<-NULL}

#lines

}

tmpList<-alply(l,1,.fun=myReadLines)

names(tmpList)<-l

tmpList

}

# function to cut data frame by resource type, get fileName columns, apply applyReadLines over them

#return list with one sublist per resource type;

## each sublist element has name=fileName, value=fileContent

getNodeDataFiles<-function(df){dlply(df,.(resource),.fun=function(df){applyReadLines(df$file)})}

# cut data frame by node name, apply getNodeDataFiles for each node

# return list with one sublist per node

## each sublist element is a subsublist with name=resource, value=(list of file contents)

rawDataList <- dlply(filesTable,.(node),.fun=getNodeDataFiles)

#remove event data from rawDataList, add to eventDataList

eventDataList <- llply(rawDataList,.fun=function(l){l[names(l)=="ev"]})

rawDataList <- llply(rawDataList,.fun=function(l){l[names(l)!="ev"]})

#some nodes can have empty event lists, which make eventParser crash

#remove those empty event lists

# structure of eventDataList

## node:list()

## - ev:list()

## --file: char(n)

#remove files with null file content

# eventDataList<-laply(eventDataList,function(l)

# {llply(l,function(l2)

# {l2[nchar(l2)>0]})})

#remove nodes without events from eventDataList

eventDataList <-eventDataList[laply(eventDataList,length)>0]

#SCF: code to deal with multi-line event messages

#

# Generate logTable and eventTable, dataframes with parsed data from log/event files

logTable<-list()

eventTable<-list()

eventRawTable<-character(0)

parseErrorTable<-list()

emptyLogTable<-list()

debugTable<-list()

if(doLogs)

{# apply resource parser over log file groups for each (node, resource) pair. Loop parallelized for performance.

#set up parallel runtime configuration

cl<-makeCluster(8)

registerDoParallel(cl)

#parOptions<-list(.packages = c("plyr","gsubfn","stringr"),.export = c("resourceParser","csvToDf", "tmpEventParser") )

parOptions<-list(.packages = c("plyr","gsubfn","stringr", "resourceParser"))

#csvToDf<-csvToDf

logTable<-llply (rawDataList,

.fun=function(node)

{llply(node,function(resource)

{ldply(resource,resourceParser)})},

.parallel=TRUE,.paropts=parOptions)

stopCluster(cl)

}

if(doEvents)

{

# load eventParser function into scope of this function

# function to apply eventParser over all files of a given node.

applyEventParser <- function(node)

{

tmpNames<-laply(node,names)

nonEmptyIndices<-laply(node,function(l){nchar(l)>0})

tmpEventSubtable<-ldply(tmpNames[nonEmptyIndices],eventParser)

}

applyGetRawEvents<- function(node)

{

node1<-node[[1]]

tmpNames<-names(node1)

nonEmptyIndices<-nchar(node1)>0

node1<-node1[nonEmptyIndices]

tmpRawEvents<-character(0)

if(length(node1)<1)(return(tmpRawEvents))

for (i in 1:length(node1))

{tmpRawEvents<-c(tmpRawEvents,names(node1)[i],node1[[i]])}

tmpRawEvents

}

# apply event parser over all event files for each node. Loop parallelized for performance.

#eventTableList<-llply (eventDataList ,.fun=applyEventParser,.parallel=TRUE,.paropts=parOptions )

eventTableList<-llply (eventDataList ,.fun=applyEventParser)

eventRawTable<-llply (eventDataList ,.fun=applyGetRawEvents)

eventRawTable<-c(eventRawTable,recursive=TRUE)

# remove empty list elements, or the following Reduce+merge fails (merge won't work with empty dataframes)

emptyEventFileList<-eventTableList[laply(eventTableList,nrow)<=0]

eventTableList<-eventTableList[laply(eventTableList,nrow)>0]

# combine list into single dataframe

eventTable<-Reduce( function(d1,d2) {merge(d1,d2, all.x=TRUE, all.y=TRUE)}, eventTableList)

if(!is.null(eventTable$parseError))

{parseErrorTable<-eventTable$parseError[!is.na(eventTable$parseError)]

eventTable<-eventTable[is.na(eventTable$parseError),]

eventTable$parseError<-NULL}

if(!is.null(eventTable$emptyLog))

{emptyLogTable<-eventTable$emptyLog[!is.na(eventTable$emptyLog)]

eventTable<-eventTable[is.na(eventTable$emptyLog),]

eventTable$emptyLog<-NULL}

eventTable$.id<-NULL

# function to factor a character vector, converting ""'s and NAs into "BLANK" levels

factorNoBlanks <- function (f)

{

f[f==""|is.na(f)]<-"BLANK"

factor(f)

}

eventColumnClasses<-laply(eventTable,function(column){class(column)[1]})

eventTable[eventColumnClasses=="character"]<-llply(eventTable[eventColumnClasses=="character"],factorNoBlanks)

# rename table labels

eventTable<-rename(eventTable, c("Created.On"="date", "Name"="node","Event"="msg","Event.Type"="type",

"Severity"="severity", "Created.By"="createdBy","Cleared.On"="clearedOn",

"Cleared.By"="clearedBy"),warn\_missing = FALSE)

# reorder by event date

#eventTable<-eventTable[order(eventTable$date),]

eventTable<-arrange(eventTable,date)

# rearrange table fields

knownFields<-c("date", "severity", "type", "node", "createdBy","clearedOn","clearedBy","msg")

unknownFields<-names(eventTable)[!names(eventTable) %in% knownFields]

eventTable<-eventTable[c(knownFields,unknownFields)]

} # end if(doEvents)

# to browse logTable, try str(logTable,3,vec.len=3,list.len=3)

#all file writing here

if(saveToFile){

#create folder tree in target path

folderList<-list.dirs(rawFolder, recursive=FALSE)

folders<-character(0)

for(f in folderList)

{folders<- unique(c(folders,list.dirs(f,recursive=TRUE,full.names=FALSE)))}

folders<-file.path(aggrDateFile,folders)

tmp<-lapply(folders,dir.create,showWarnings=FALSE)

#write data to files

for (node in names(rawDataList))

{

cpuFile<-file.path(aggrDateFile,paste(node,"\_cpu.csv",sep=""))

memFile<-file.path(aggrDateFile,paste(node,"\_mem.csv",sep=""))

hdFile<-file.path(aggrDateFile,paste(node,"\_hd.csv",sep=""))

intFile<-file.path(aggrDateFile,paste(node,"\_int.csv",sep=""))

evFile<-file.path(aggrDateFile,paste(node,"\_ev.csv",sep=""))

cpu<-logTable[[node]]$cpu

mem<-logTable[[node]]$mem

hd <-logTable[[node]]$hd

int<-logTable[[node]]$int

ev <-eventTable[[node]]$ev

#debugTable collects write results

debugTable[[node]]<-list()

if(!is.null(cpu)&&dim(cpu)[2]>0) {write.table(cpu,cpuFile)} else

{debugTable[[node]]$cpu<-NA}

if(!is.null(mem)&&dim(mem)[2]>0) {write.table(mem,memFile)} else

{debugTable[[node]]$mem<-NA}

if(!is.null(hd)&&dim(hd)[2]>0) {write.table(hd,hdFile)} else

{debugTable[[node]]$hd<-NA}

if(!is.null(int)&&dim(int)[2]>0) {write.table(int,intFile)} else

{debugTable[[node]]$int<-NA}

}

write.table(eventTable,eventTableFile)

writeLines(eventRawTable,eventRawFile)

}

#debugTable$wrongNameFiles<-wrongNameFiles

#debugTable$parseError<-parseErrorTable

resultList<-list(logTable=logTable,eventTable=eventTable,debugTable=debugTable,emptyLog=emptyLogTable,wrongNameFiles=wrongNameFiles,parseError=parseErrorTable)

} # end of function aggregateDataFiles

#resultList<-aggregateDataFiles(FALSE,TRUE,FALSE)

## [Comments]

# Dataset 4: SCF-1-4-visualization-tables

## Context

Derived from dataset SCF-1-3-aggrDateFiles. Generates csv files to feed a prototype dashboard. We have the following requirements:

* SCF-1-4a: resources-[node].csv: Multiple line chart for a specific node resources:
  + date,cpu,mem,hd,int
* SCF-1-4b: event-[node]: Scatterplot for a specific node events:
  + date,event-type
* SCF-1-4c: resources-allnodes.csv
  + date, node,resource-type,value
* SCF-1-4d: event-allnodes.csv
  + date, node,event-type

## Overview

We need to perform the following transformations:

4

eventTable -> event-allnodes

* Original fields: date, severity, type, node, createdBy, clearedOn, clearedBy, msg
* Target fields: date, node, event-type
  + select(date,node,type)
  + rename(type="event-type")
* Other transformations:
  + filter(severity != "BLANK")

2

event-allnodes -> event-[node]

* Operation: split data frame into N data frames by value of node
* split(event\_allnodes\_table, event\_allnodes\_table$node)

1

resources\_rawList<-resultList$logTable

resources\_bynode<-llply(resources\_rawList,

function(node)

{nodeResources<-

llply(names(node), function(resource,node)

{res2<-node[[resource]][,2:3];

names(res2)<-c("date",resource); res2},

node

)

Reduce(function(d1,d2){join(d1,d2,by="date")},nodeResources)

})

## Content

## Format

## [Generation]

Full processing script

# This script takes resources ('resultList$logTable), and event data ('eventTable') from SCF,

# as generated from script [SCF]/...

# and generates the following output:

#

require(reshape2)

require(plyr)

require(dplyr)

# folder to load utility functions

utilsPath<- "C:/Users/capelastegui/workspace/OFP/utils"

# folder to load data

baseDataPath <- "C:/Users/capelastegui/workspace/OFP/SCF/SCF-1/1-Data"

# Folder to store output

savePath <- file.path(baseDataPath,"4-VisualizationTables")

# load utility functions

source(file.path(utilsPath,"refactor.R"))

source(file.path(utilsPath,"removeNAs.R"))

# events\_allnodes: table with events(date,node,type) for all system nodes

events\_allnodes<-eventTable %.% filter(severity!="BLANK") %.% select (date,node,type) %.% refactor()

# events\_allnodes.filtered: table with a subset of events\_allnodes (from 1/2/14 to 16/2/14)

events\_allnodes.filtered <- events\_allnodes %.% filter(date > dmy(01022014) & date < dmy(16022014)) %.% refactor()

# events\_bynode - list of event tables for each node

events\_bynode<-split(event\_allnodes\_table, event\_allnodes\_table$node)

# resources\_rawList - not fully parsed resources data, for all nodes

resources\_rawList<-resultList$logTable

#resources\_bynode - list of resource tables for each node

resources\_bynode<-llply(resources\_rawList,

function(node)

{nodeResources<-

llply(names(node), function(resource,node)

{res2<-node[[resource]][,2:3];

names(res2)<-c("date",resource); res2},

node

)

result<-Reduce(function(d1,d2){join(d1,d2,by="date")},nodeResources)

result<-result[complete.cases(result),]

result<-result[!duplicated(result$date),]

result %.% arrange (date)

})

#resources\_allnodes - list of resources for all nodes, long format

resources\_allnodes<-ldply(resources\_bynode,.id="node") %.% select(date,node,cpu,int,mem,hd) %.% arrange(date)

# remove path from node names

levels(resources\_allnodes$node)<-laply(levels(resources\_allnodes$node),function(s){tail(strsplit(s,"/")[[1]],1)})

# for visualization performance, choose a random subset with only a few nodes

set.seed(1)

resources\_allnodes.filtered <- resources\_allnodes %.%

filter(node %in% c("lbpeuin01.sceu.corp",

"lbpeuin02.sceu.corp",

"lbpeuid01.sceu.corp",

"lbpeuid02.sceu.corp")) %.%

filter(date > dmy(01022014) & date < dmy(16022014)) %.% refactor()

resources\_allnodes.filtered <- resources\_allnodes.filtered [sample(nrow(resources\_allnodes.filtered),4000),] %.% tbl\_df()

#res.events\_allnodes - combined table with events, resources for all nodes

res.events\_allnodes<-rbind.fill(resources\_allnodes,events\_allnodes)

# need to filter out rows with NAs,

res.events\_allnodes<-removeNAs.df(res.events\_allnodes, replace.fac="not.event") %.%

mutate(event=type!="not.event") %.% # and indicate which rows are events/resources

melt(id=c("date", "node", "type")) %.% # and melt it into a tall format

filter(!(variable=="event"&type=="not.event")& !(type!="not.event" & value==0) )

# and filter out inconsistencies between resource/event rows

# for visualization performance, choose a random subset with only a few nodes, 'res.events\_allnodes.sample'

set.seed(1)

res.events\_allnodes.sample<- res.events\_allnodes[sample(nrow(res.events\_allnodes),20000),] %.% tbl\_df()

#hack to fix visualization: change event type to numeric variable, with "not\_event" = 0.

res.events\_allnodes.sample$type <- as.numeric(res.events\_allnodes.sample$type)

res.events\_allnodes.sample$type[res.events\_allnodes.sample$type==2]<-0

# save output to files

write.table(resources\_bynode[[1]],

file.path(savePath,"resources-node1.csv") ,row.names=FALSE, quote=FALSE, sep= ",")

write.table(resources\_allnodes,

file.path(savePath,"resources-allnodes.csv"),row.names=FALSE, quote=FALSE, sep= ",")

write.table(resources\_allnodes.filtered,

file.path(savePath,"resources-allnodes.filtered.csv"),row.names=FALSE, quote=FALSE, sep= ",")

write.table(events\_bynode$lbpeuid01.sceu,

file.path(savePath,"events-node1.csv"), row.names=FALSE, quote=FALSE, sep= ",")

write.table(events\_allnodes,

file.path(savePath,"events-allnodes.csv"),row.names=FALSE, quote=FALSE, sep= ",")

write.table(events\_allnodes.filtered,

file.path(savePath,"events-allnodes.filtered.csv"),row.names=FALSE, quote=FALSE, sep= ",")

write.table(res.events\_allnodes,

file.path(savePath,"res.events-allnodes.csv"),row.names=FALSE, quote=FALSE, sep= ",")

write.table(res.events\_allnodes.sample,

file.path(savePath,"res.events-allnodes.sample.csv"),row.names=FALSE, quote=FALSE, sep= ",")

## [Comments]

### Parsing trouble

I'm having trouble parsing certain dates:

head(eventDataList[[2]][[1]][[6]])

[1] "\"Created On\",\"Severity\",\"Name\",\"Event\",\"Created By\",\"Cleared On\",\"Event Type\",\"Cleared By\",\"Model Type Name\",\"Event Precedence\""

[2] "\"03-mar-2014 19:52:43 CET\",\"\",\"aswpeuib02.sceu.corp\",\"A vlanTrunkPortDynamicStatusChange trap has been received from model aswpeuib02.sceu.corp of type Rtr\_Cisco. vlanTrunkPortDynamicStatus - 1\",\"System\",\"\",\"0x2104c7\",\"\",\"Rtr\_Cisco\",\"10\""

[3] "\"03-mar-2014 19:52:36 CET\",\"\",\"aswpeuib02.sceu.corp\",\"A vlanTrunkPortDynamicStatusChange trap has been received from model aswpeuib02.sceu.corp of type Rtr\_Cisco. vlanTrunkPortDynamicStatus - 2\",\"System\",\"\",\"0x2104c7\",\"\",\"Rtr\_Cisco\",\"10\""

[4] "\"03-mar-2014 19:43:28 CET\",\"\",\"aswpeuib02.sceu.corp\",\"A vlanTrunkPortDynamicStatusChange trap has been received from model aswpeuib02.sceu.corp of type Rtr\_Cisco. vlanTrunkPortDynamicStatus - 1\",\"System\",\"\",\"0x2104c7\",\"\",\"Rtr\_Cisco\",\"10\""

[5] "\"03-mar-2014 19:43:22 CET\",\"\",\"aswpeuib02.sceu.corp\",\"A vlanTrunkPortDynamicStatusChange trap has been received from model aswpeuib02.sceu.corp of type Rtr\_Cisco. vlanTrunkPortDynamicStatus - 2\",\"System\",\"\",\"0x2104c7\",\"\",\"Rtr\_Cisco\",\"10\""

[6] "\"03-mar-2014 19:42:46 CET\",\"\",\"aswpeuib02.sceu.corp\",\"A vlanTrunkPortDynamicStatusChange trap has been received from model aswpeuib02.sceu.corp of type Rtr\_Cisco. vlanTrunkPortDynamicStatus - 1\",\"System\",\"\",\"0x2104c7\",\"\",\"Rtr\_Cisco\",\"10\""

quickStr(eventTableList[[2]])

'data.frame': 34 obs. of 11 variables:

$ .id : chr "6" "6" "6" ...

$ Created.On : POSIXct, format: NA NA NA ...

$ Severity : chr "" "" "" ..

Traced back to lubridate bug?

> dmy("03-jan-2014",locale="English")

[1] "2014-01-03 UTC"

> dmy("03-feb-2014",locale="English")

[1] "2014-02-03 UTC"

> dmy("03-mar-2014",locale="English")

[1] "2014-03-03 UTC"

> dmy("03-apr-2014",locale="English")

[1] "2014-04-03 UTC"

> dmy("03-ene-2014",locale="Spanish")

[1] "2014-01-03 UTC"

> dmy("03-feb-2014",locale="Spanish")

[1] "2014-02-03 UTC"

> dmy("03-mar-2014",locale="Spanish")

[1] NA

Warning message:

All formats failed to parse. No formats found.

> dmy("03-may-2014",locale="Spanish")

[1] "2014-05-03 UTC"

> parse\_date\_time("03-feb-2014",order="dmy",locale="Spanish")

[1] "2014-02-03 UTC"

> parse\_date\_time("03-mar-2014",order="dmy",locale="Spanish")

[1] NA

Warning message:

All formats failed to parse. No formats found.

> parse\_date\_time("03-abr-2014",order="dmy",locale="Spanish")

[1] "2014-04-03 UTC"

> Sys.getlocale("LC\_TIME")

[1] "Spanish\_Spain.1252"

> format(now()-2\*30\*24\*3600,"%Y-%b-%d")

[1] "2014-mar-06"

> strptime("03-abr-2014","%d-%b-%Y")

[1] "2014-04-03"

> dmy(c("03-mar-2014","03-mar-2014"),locale="Spanish")

[1] NA NA

Warning message:

All formats failed to parse. No formats found.

> dmy(c("03-mar-2014","03-feb-2014"),locale="Spanish")

[1] "2014-03-03 UTC" "2014-02-03 UTC"

> dmy(c("03-mar-2014","03-03-2014"),locale="Spanish")

[1] NA "2014-03-03 UTC"

Warning message:

1 failed to parse.

> dmy(c("03-mar-2014","03-marzo-2014"),locale="Spanish")

[1] "2014-03-03 UTC" "2014-03-03 UTC"

The failure only comes up if:

* locale Spanish
* date range includes dates for march, abreviated as "mar"
* date range includes no other months

Hypothesis

* overlap
  + mar - marzo
  + mar - martes
  + weekday -month

solution: revert to strptime

TODO: fix or report bug

### Issue: strange resources?

Need to identify origin

> llply(resources\_bynode,names)

$`/internet/aswpeuib01.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/aswpeuib02.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/aswpeuib03.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/aswpeuib04.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/aswpeuid01.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/aswpeuid02.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/aswpeuid03.sceu.corp`

[1] "date" "cpu" "int" "mem" "memu"

$`/internet/aswpeuid04.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/dnsautpeuid.sceu.corp`

[1] "date" "mem"

$`/internet/dnsautpeuid01.sceu.corp`

[1] "date" "cpu" "hd" "int" "mem"

$`/internet/dnsautpeuid02.sceu.corp`

[1] "date" "cpu" "hd" "int" "mem"

$`/internet/dswpeuwt01.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/dswpeuwt02.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/dswpeuwt03.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/dswpeuwt04.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/fwpeuib01`

[1] "date" "cpu" "hd" "int" "mem"

$`/internet/fwpeuib01.sceu.corp`

[1] "date" "cpu" "hd" "int" "mem"

$`/internet/fwpeuib02`

[1] "date" "cpu" "hd" "int" "mem"

$`/internet/fwpeuib02.`

[1] "date" "hd"

$`/internet/fwpeuib02.sceu.corp`

[1] "date" "cpu" "hd" "int" "mem"

$`/internet/fwpeuid01`

[1] "date" "cpu" "hd" "int" "mem"

$`/internet/fwpeuid01.sceu.corp`

[1] "date" "cpu" "hd" "int" "mem"

$`/internet/fwpeuid02`

[1] "date" "cpu" "hd" "int" "mem"

$`/internet/fwpeuid02.sceu.corp`

[1] "date" "cpu" "hd" "int" "mem"

$`/internet/lbpeuib01.sceu.corp`

[1] "date" "cpu" "int" "intu" "mem"

$`/internet/lbpeuib02.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/lbpeuid01.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/internet/lbpeuid02.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/aswpeuin01.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/aswpeuin02.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/aswpeuin03.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/aswpeuin04.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/aswpeuin05.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/aswpeuin06.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/aswpeuin07.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/aswpeuin08.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/dnsrespeuin01.sceu.corp`

[1] "date" "cpu" "hd" "int" "mem"

$`/intranet/dnsrespeuin02.sceu.corp`

[1] "date" "cpu" "hd" "int" "mem"

$`/intranet/dswpeuin01.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/dswpeuin02.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/fwpeuin01`

[1] "date" "cpu" "int" "mem"

$`/intranet/fwpeuin02`

[1] "date" "cpu" "int" "mem"

$`/intranet/lbpeuin01.sceu.corp`

[1] "date" "cpu" "int" "mem"

$`/intranet/lbpeuin02.sceu.corp`

[1] "date" "cpu" "int" "mem"

### Future code cleanup

#### Function separation

For aggr-dateFiles, we should split processing in three steps, each with a separate function

* Load files from disk to memory
* Parse files
* save to disk

#### Parser operation

For historic reasons, eventParser works with a disk file, even though we have already loaded that data to memory.

eventDataList <- ... # table with events, by file

eventTableList<-llply (eventDataList ,.fun=applyEventParser) # we parse by loading from disk again!

eventRawTable<-llply (eventDataList ,.fun=applyGetRawEvents) # yet we get raw event table from memory!

Discussion: we did this to use read.table, which uses a file as input. As a solution, we can use a textConnection()

<http://stackoverflow.com/questions/3261066/in-r-how-can-i-parse-csv-data-from-a-character-vector-to-extract-a-data-frame>

You can use textConnection() to pass the character vector to read.table(). An example:

x <- "first,second\nthird,fourth\n"

x1 <- read.table(textConnection(x), sep = ",")

# Dataset N

## Context

## Overview

## Content

## Format

## [Generation]

## [Comments]

# Dataset analysis

## List of failure events

As a reminder from section 2.3.4, log events have the following data fields:

* date : POSIXct date object
* severity : Factor with 4 levels: BLANK, Critical, Major, Minor
* type : Factor
* node : Factor
* createdBy : Factor
* clearedOn : POSIXct date object
* clearedBy : Factor
* msg :Factor

### Subset 1: 27/11 to 13/12

We start by examining the relationship between event level, type, and node.

library(ggplot2)

#Copy event table to work with it.

miniTable<-eventTable

#Number of events

nrow(miniTable)

#base ggplot

lastDay<-miniTable$date[nrow(miniTable)]

#use last 3 days from minitable

miniTable<-miniTable[miniTable$date>lastDay-(3600\*24\*3),]

failTable<-miniTable[miniTable$severity!="BLANK",]

debugTable<-miniTable[miniTable$severity=="BLANK",]

p<-ggplot(miniTable)

pf<-ggplot(failTable)

pd<-ggplot(debugTable)

#Plot events per hour

p.date.hist<-p+aes(date)+geom\_bar(fill="white",colour="black",binwidth=3600\*1)

p.date.hist+facet\_grid(severity~.)

p.date.hist+facet\_grid(severity~node)

#Plot failure events per hour

pf.date.hist<-pf+aes(date,fill=type)+geom\_bar(binwidth=3600\*1)

pf.date.hist+facet\_grid(node~severity,margins=TRUE)

pf.date.hist+facet\_grid(node~severity,scales="free\_y",margins=TRUE)

#Plot debug events per hour

pd.date.hist<-pd+aes(date,fill=type)+geom\_bar(binwidth=3600\*1)

pd.date.hist+facet\_grid(node~severity,margins=TRUE)

pd.date.hist+facet\_grid(node~severity,scales="free\_y",margins=TRUE)



Figure 1 - Failure Events per date, by severity,node,type

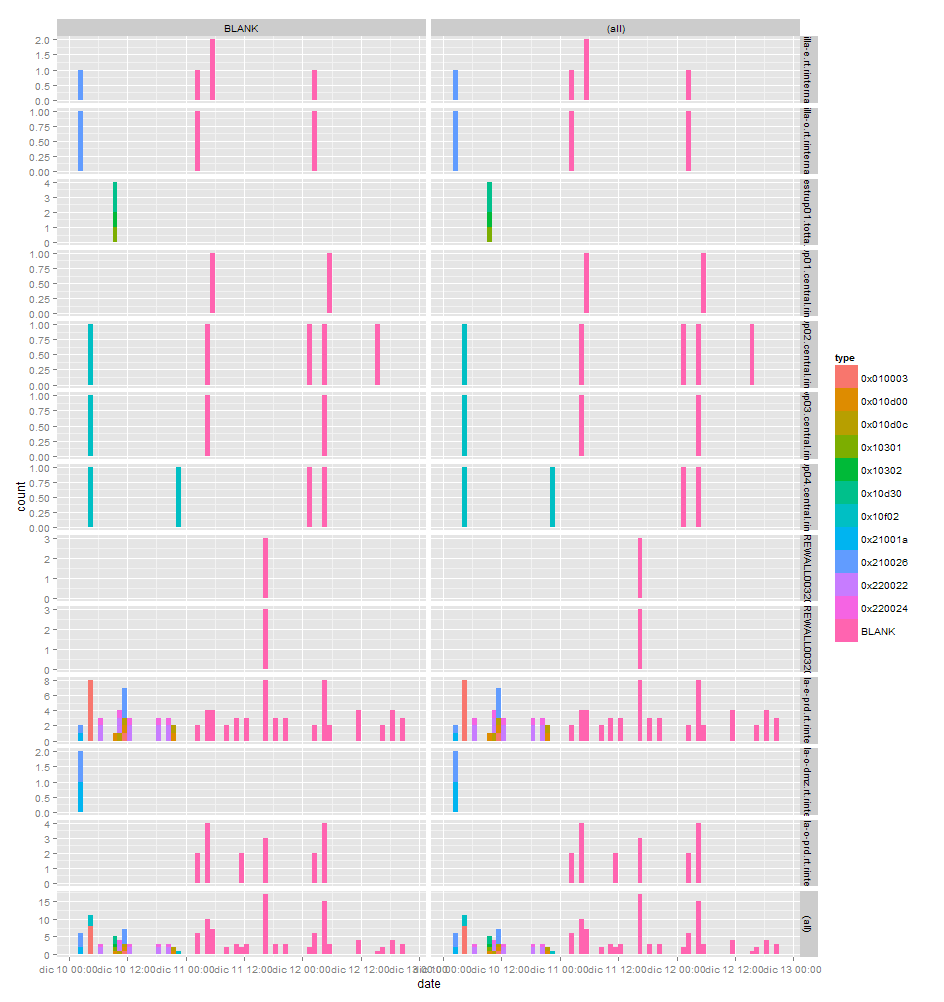


Figure 2 - non-failure Events per date, by severity,node,type

## Failure analysis

### Analysis exploration 1

The first observed system failure is described as follows:

* Source File: bvtowpocwp01.central.rinterna.local\_EV\_151113.csv
* Node: bvtowpocwp01.central.rinterna.local
* Date: 15-nov-2013 5:26:32 CET
* "Major","15-nov-2013 5:26:32 CET","bvtowpocwp01.central.rinterna.local","High CPU Utilization. The utilization of 98% for CPU instance 2 named 'CPU: 2' has exceeded the 85% threshold on model bvtowpocwp01.central.rinterna.local for more than the acceptable time period.","System","15-nov-2013 5:26:32 CET","System"

We explore the failure using the following script:

Code 4: totta-1-0-raw-1-exploreFailure1.R

baseDataPath <- "C:/Users/capelastegui/workspace/OFP/Totta/Totta-1/1-Data"

testFolder <- file.path(baseDataPath,"0-test")

dataPath <-testFolder

fileIn <- file(file.path(dataPath, "totta-1-tmp-cpu.csv"))

cpu<-read.csv(fileIn)

names(cpu)<-c("date","cpu")

cpu$date<-strptime (cpu$date,"%m/%d/%Y %H:%M")

cpu$cpu<-as.numeric(cpu$cpu)

library(xts)

cpuXts<-xts(cpu$cpu,order.by=cpu$date)

cpuTs<-ts(cpu$cpu)

plot(cpu$date[55:75],cpu$cpu[55:75],type="l",xaxt="n",xlab=NA)

points(cpu$date[55:75],cpu$cpu[55:75])

axis.POSIXct(side=1, las=2,at=as.character(cpu$date))

plot(cpuXts[55:75],xaxt="n",main="")

points(cpu$date[55:75],cpu$cpu[55:75])

axis.POSIXct(side=1, las=2,at=as.character(cpu$date))

fileIn <- file(file.path(dataPath, "totta-1-tmp-mem.csv"))

mem<-read.csv(fileIn)

names(mem)<-c("date","mem")

mem$date<-strptime (mem$date,"%m/%d/%Y %H:%M")

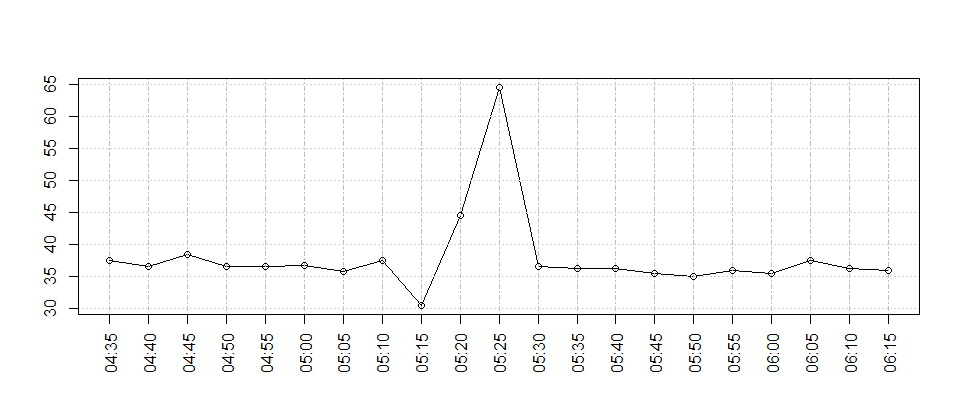
mem$mem<-as.numeric(mem$mem)

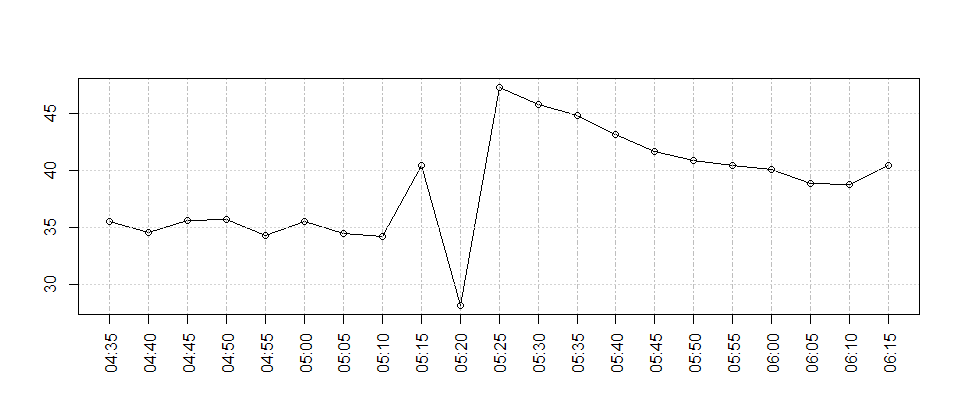
memXts<-xts(mem$mem,order.by=mem$date)

plot(memXts[55:75],xaxt="n",main="")

points(mem$date[55:75],mem$mem[55:75])

axis.POSIXct(side=1, las=2,at=as.character(mem$date))





### Analysis Exploration 2 - Failure list, 15 days

We study the failures in a 15-day capture period

#Copy event table to work with it.

miniTable<-eventTable[,c("date","Severity","Event.Type","Event")]

#Number of events

dim(eventTable)[1]

[1] 187

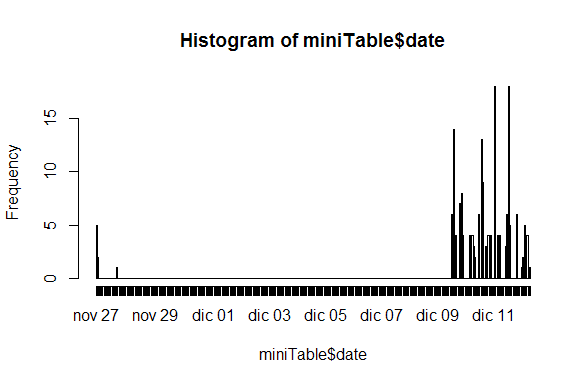
#Renaming null, empty severity levels

miniTable$Severity[is.na(miniTable$Severity)] <- ""

levels(miniTable$Severity)[levels(miniTable$Severity)==""]<-"0"

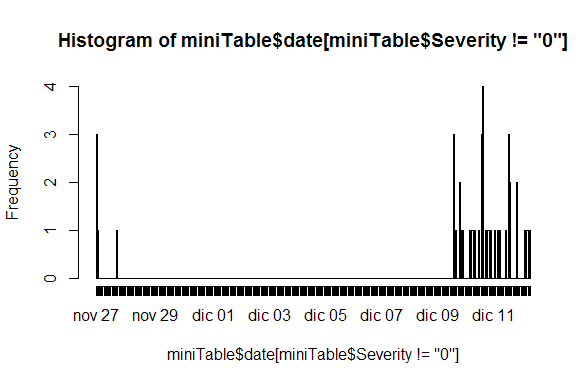
#Plot events per hour

hist(miniTable$date,breaks="hours", freq=TRUE)



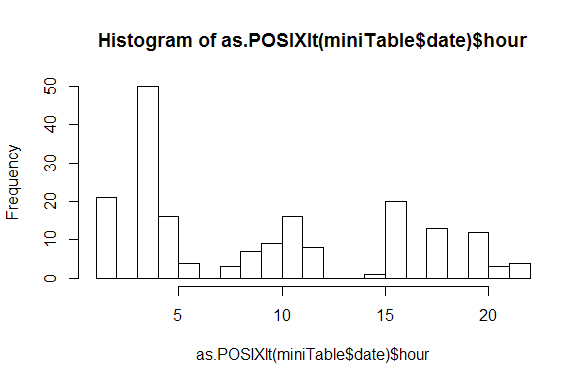
#Plot failure events per hour

hist(miniTable$date[miniTable$Severity!="0"],breaks="hours", freq=TRUE)



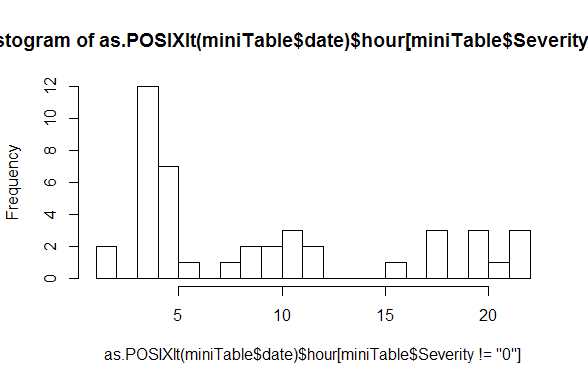
#Plot events per day hour

hist(as.POSIXlt(miniTable$date)$hour,breaks=24, freq=TRUE)



#Plot failure events per day hour hour

hist(as.POSIXlt(miniTable$date)$hour[miniTable$Severity!="0"],breaks=24, freq=TRUE)



> sort(table(miniTable$Event.Type))

0x10301 0x10302 0x10d35 0x21001a 0x10d30 0x010d00 0x010d0c 0x010daa 0x220024 0x220031 0x10f02

1 1 1 2 2 3 3 3 5 5 7

0x210026 0x010003 0x10f03 0x220022 0

8 9 9 10 118

#Divide table by value of Event Type

miniTable<-dlply(miniTable,.(Event.Type))

#For each subtable, show levels of severity, event message

summarizeMiniTable<-function(mTable)

{

tmp<-mTable[,c("Severity","Event")]

tmp<-llply(tmp,factor)

tmp<-llply(tmp,table)

}

miniTableSummary<-llply(miniTable,summarizeMiniTable)

miniTableSummary

> miniTableSummary

$`0x210026`

$`0x210026`$Severity

0

8

$`0x210026`$Event

A configuration management event has been recorded in the ccmHistoryEventTable. The source of the command that instigated the event is snmp. The configuration data source for the event is running. The configuration data destination for the event is networkTftp.

4

A configuration management event has been recorded in the ccmHistoryEventTable. The source of the command that instigated the event is commandLine. The configuration data source for the event is commandSource. The configuration data destination for the event is running.

2

A configuration management event has been recorded in the ccmHistoryEventTable. The source of the command that instigated the event is commandLine. The configuration data source for the event is running. The configuration data destination for the event is commandSource.

1

A configuration management event has been recorded in the ccmHistoryEventTable. The source of the command that instigated the event is commandLine. The configuration data source for the event is running. The configuration data destination for the event is startup.

1

$`0x010003`

$`0x010003`$Severity

0

9

$`0x010003`$Event

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 0. loctcpConnOutBytes = 0. ElapsedTime = 2 time ticks. (Trap type : 0x10003)

(...)

1

$`0x010d00`

$`0x010d00`$Severity

0

3

$`0x010d00`$Event

Device rs-boadilla-e-prd.rt.rinterna.local of type SwCiscoIOS is no longer responding to primary management requests (e.g. SNMP), but appears to be responsive to other communication protocol (e.g. ICMP).

3

$`0x010d0c`

$`0x010d0c`$Severity

0

3

$`0x010d0c`$Event

Device rs-boadilla-e-prd.rt.rinterna.local of type SwCiscoIOS is again responding to primary management requests.

3

$`0x010daa`

$`0x010daa`$Severity

Major

3

$`0x010daa`$Event

Device rs-boadilla-e-prd.rt.rinterna.local of type SwCiscoIOS is no longer responding to primary management requests (e.g. SNMP), but appears to be responsive to other communication protocol (e.g. ICMP). This condition has persisted for an extended amount of time. An alarm will be generated.

3

$`0x21001a`

$`0x21001a`$Severity

0

2

$`0x21001a`$Event

%SYS-4-SNMP\_WRITENET: .

2

$`0x220022`

$`0x220022`$Severity

0

10

$`0x220022`$Event

An ospfNbrStateChange trap has been received for router 172.26.64.7. The neighbor router is 172.18.187.3, and the new state is DOWN.

5

An ospfNbrStateChange trap has been received for router 172.26.64.7. The neighbor router is 172.18.187.3, and the new state is FULL.

5

$`0x220024`

$`0x220024`$Severity

0

5

$`0x220024`$Event

An ospfNbrStateChange trap has been received for router 172.26.64.7. The neighbor router is 172.18.187.3, and the new state is FULL. Previous alarms will be cleared.

5

$`0x220031`

$`0x220031`$Severity

Minor

5

$`0x220031`$Event

An ospfNbrStateChange trap has been received for router 172.26.64.7. The neighbor router is 172.18.187.3, and the new state is DOWN.

5

$`0x10301`

$`0x10301`$Severity

0

1

$`0x10301`$Event

Model bvtolestrup01.totta.corp of type Host\_Device has been contacted.

1

$`0x10302`

$`0x10302`$Severity

0

1

$`0x10302`$Event

Contact has been lost with model bvtolestrup01.totta.corp of type Host\_Device.

1

$`0x10d30`

$`0x10d30`$Severity

0

2

$`0x10d30`$Event

The condition causing the loss of contact on the device model has cleared ( name - bvtolestrup01.totta.corp, type - Host\_Device ).

2

$`0x10d35`

$`0x10d35`$Severity

Critical

1

$`0x10d35`$Event

Device bvtolestrup01.totta.corp of type Host\_Device has stopped responding to polls and/or external requests. An alarm will be generated.

1

$`0x10f02`

$`0x10f02`$Severity

0

7

$`0x10f02`$Event

Normal CPU Utilization. The CPU Utilization for all CPU instances is now below the 70% reset threshold for model bvtowpocwp01.central.rinterna.local.

1

Normal CPU Utilization. The CPU Utilization for all CPU instances is now below the 70% reset threshold for model bvtowpocwp02.central.rinterna.local.

1

Normal CPU Utilization. The CPU Utilization for all CPU instances is now below the 70% reset threshold for model bvtowpocwp03.central.rinterna.local.

2

Normal CPU Utilization. The CPU Utilization for all CPU instances is now below the 70% reset threshold for model bvtowpocwp04.central.rinterna.local.

3

$`0x10f03`

$`0x10f03`$Severity

Major

9

$`0x10f03`$Event

High CPU Utilization. The utilization of 93% for CPU instance 3 named 'CPU: 3' has exceeded the 85% threshold on model bvtowpocwp01.central.rinterna.local for more than the acceptable time period.

1

High CPU Utilization. The utilization of 99% for CPU instance 1 named '' has exceeded the 85% threshold on model bvtowpocwp02.central.rinterna.local for more than the acceptable time period.

2

High CPU Utilization. The utilization of 94% for CPU instance 1 named '' has exceeded the 85% threshold on model bvtowpocwp03.central.rinterna.local for more than the acceptable time period.

1

High CPU Utilization. The utilization of 85% for CPU instance 3 named 'CPU: 3' has exceeded the 85% threshold on model bvtowpocwp03.central.rinterna.local for more than the acceptable time period.

1

High CPU Utilization. The utilization of 95% for CPU instance 1 named '' has exceeded the 85% threshold on model bvtowpocwp03.central.rinterna.local for more than the acceptable time period.

1

High CPU Utilization. The utilization of 85% for CPU instance 3 named 'CPU: 3' has exceeded the 85% threshold on model bvtowpocwp04.central.rinterna.local for more than the acceptable time period.

1

High CPU Utilization. The utilization of 96% for CPU instance 2 named 'CPU: 2' has exceeded the 85% threshold on model bvtowpocwp04.central.rinterna.local for more than the acceptable time period.

1

High CPU Utilization. The utilization of 94% for CPU instance 1 named '' has exceeded the 85% threshold on model bvtowpocwp04.central.rinterna.local for more than the acceptable time period.

1

$`0`

$`0`$Severity

0 1 2 Major Minor

93 3 1 14 7

$`0`$Event

A configuration management event has been recorded in the ccmHistoryEventTable. The source of the command that instigated the event is snmp. The configuration data source for the event is running. The configuration data destination for the event is networkTftp.

8

Device boadilla-e.rt.rinterna.local of type Rtr\_Cisco is again responding to primary management requests.

1

Device boadilla-e.rt.rinterna.local of type Rtr\_Cisco is no longer responding to primary management requests (e.g. SNMP), but appears to be responsive to other communication protocol (e.g. ICMP).

1

Device boadilla-e.rt.rinterna.local of type Rtr\_Cisco is no longer responding to primary management requests (e.g. SNMP), but appears to be responsive to other communication protocol (e.g. ICMP). This condition has persisted for an extended amount of time. An alarm will be generated.

1

A device FIREWALL003201 of type NokiaFW has reported a permanent configuration file save change for file /config/db/initial. -

1

A device FIREWALL003201 of type NokiaFW has reported a temporary configuration change. -

2

A device FIREWALL003202 of type NokiaFW has reported a permanent configuration file save change for file /config/db/initial. -

1

A device FIREWALL003202 of type NokiaFW has reported a temporary configuration change. -

2

%SYS-4-SNMP\_WRITENET: .

4

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 0. loctcpConnOutBytes = 0. ElapsedTime = 2 time ticks. (Trap type : 0x10003)

4

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1011. loctcpConnOutBytes = 951. ElapsedTime = 1183 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = synReceived. sessionType = tcp. loctcpConnInBytes = 911. loctcpConnOutBytes = 883. ElapsedTime = 965 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = synReceived. sessionType = tcp. loctcpConnInBytes = 911. loctcpConnOutBytes = 883. ElapsedTime = 969 time ticks. (Trap type : 0x10003)

2

A configuration management event has been recorded in the ccmHistoryEventTable. The source of the command that instigated the event is commandLine. The configuration data source for the event is commandSource. The configuration data destination for the event is running.

3

A configuration management event has been recorded in the ccmHistoryEventTable. The source of the command that instigated the event is commandLine. The configuration data source for the event is running. The configuration data destination for the event is commandSource.

1

A configuration management event has been recorded in the ccmHistoryEventTable. The source of the command that instigated the event is commandLine. The configuration data source for the event is running. The configuration data destination for the event is startup.

2

An ospfNbrStateChange trap has been received for router 172.26.64.7. The neighbor router is 172.18.187.3, and the new state is DOWN.

14

An ospfNbrStateChange trap has been received for router 172.26.64.7. The neighbor router is 172.18.187.3, and the new state is FULL.

8

An ospfNbrStateChange trap has been received for router 172.26.64.7. The neighbor router is 172.18.187.3, and the new state is FULL. Previous alarms will be cleared.

8

Device rs-boadilla-e-prd.rt.rinterna.local of type SwCiscoIOS is again responding to primary management requests.

4

Device rs-boadilla-e-prd.rt.rinterna.local of type SwCiscoIOS is no longer responding to primary management requests (e.g. SNMP), but appears to be responsive to other communication protocol (e.g. ICMP).

4

Device rs-boadilla-e-prd.rt.rinterna.local of type SwCiscoIOS is no longer responding to primary management requests (e.g. SNMP), but appears to be responsive to other communication protocol (e.g. ICMP). This condition has persisted for an extended amount of time. An alarm will be generated.

4

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1011. loctcpConnOutBytes = 951. ElapsedTime = 1181 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1011. loctcpConnOutBytes = 951. ElapsedTime = 1186 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1027. loctcpConnOutBytes = 951. ElapsedTime = 1190 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1027. loctcpConnOutBytes = 951. ElapsedTime = 1221 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = synReceived. sessionType = tcp. loctcpConnInBytes = 17303. loctcpConnOutBytes = 92527. ElapsedTime = 27838 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1011. loctcpConnOutBytes = 951. ElapsedTime = 1182 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1027. loctcpConnOutBytes = 951. ElapsedTime = 1202 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1027. loctcpConnOutBytes = 951. ElapsedTime = 1224 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = synReceived. sessionType = tcp. loctcpConnInBytes = 2567. loctcpConnOutBytes = 3419. ElapsedTime = 88460 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-e-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = synReceived. sessionType = tcp. loctcpConnInBytes = 911. loctcpConnOutBytes = 883. ElapsedTime = 966 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-o-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 0. loctcpConnOutBytes = 0. ElapsedTime = 2 time ticks. (Trap type : 0x10003)

2

A cisco router, SwCiscoIOS (name - rs-boadilla-o-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1011. loctcpConnOutBytes = 951. ElapsedTime = 1180 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-o-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1027. loctcpConnOutBytes = 951. ElapsedTime = 1201 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-o-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1276. loctcpConnOutBytes = 1067. ElapsedTime = 1144 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-o-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1292. loctcpConnOutBytes = 1067. ElapsedTime = 849 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-o-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = synReceived. sessionType = tcp. loctcpConnInBytes = 3519. loctcpConnOutBytes = 5975. ElapsedTime = 21902 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-o-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = synReceived. sessionType = tcp. loctcpConnInBytes = 911. loctcpConnOutBytes = 883. ElapsedTime = 967 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-o-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1011. loctcpConnOutBytes = 951. ElapsedTime = 1185 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-o-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = finWait2. sessionType = tcp. loctcpConnInBytes = 1027. loctcpConnOutBytes = 951. ElapsedTime = 1204 time ticks. (Trap type : 0x10003)

1

A cisco router, SwCiscoIOS (name - rs-boadilla-o-prd.rt.rinterna.local) has reported the end of a TCP session. tcpConnEntry = synReceived. sessionType = tcp. loctcpConnInBytes = 911. loctcpConnOutBytes = 883. ElapsedTime = 969 time ticks. (Trap type : 0x10003)

1

High CPU Utilization. The utilization of 97% for CPU instance 2 named 'CPU: 2' has exceeded the 85% threshold on model bvtowpocwp01.central.rinterna.local for more than the acceptable time period.

1

Normal CPU Utilization. The CPU Utilization for all CPU instances is now below the 70% reset threshold for model bvtowpocwp01.central.rinterna.local.

2

High CPU Utilization. The utilization of 86% for CPU instance 4 named 'CPU: 4' has exceeded the 85% threshold on model bvtowpocwp01.central.rinterna.local for more than the acceptable time period.

1

High CPU Utilization. The utilization of 99% for CPU instance 1 named '' has exceeded the 85% threshold on model bvtowpocwp02.central.rinterna.local for more than the acceptable time period.

1

Normal CPU Utilization. The CPU Utilization for all CPU instances is now below the 70% reset threshold for model bvtowpocwp02.central.rinterna.local.

4

High CPU Utilization. The utilization of 86% for CPU instance 4 named 'CPU: 4' has exceeded the 85% threshold on model bvtowpocwp02.central.rinterna.local for more than the acceptable time period.

1

High CPU Utilization. The utilization of 87% for CPU instance 1 named '' has exceeded the 85% threshold on model bvtowpocwp02.central.rinterna.local for more than the acceptable time period.

1

High CPU Utilization. The utilization of 96% for CPU instance 2 named 'CPU: 2' has exceeded the 85% threshold on model bvtowpocwp02.central.rinterna.local for more than the acceptable time period.

1

High CPU Utilization. The utilization of 94% for CPU instance 1 named '' has exceeded the 85% threshold on model bvtowpocwp03.central.rinterna.local for more than the acceptable time period.

1

Normal CPU Utilization. The CPU Utilization for all CPU instances is now below the 70% reset threshold for model bvtowpocwp03.central.rinterna.local.

2

High CPU Utilization. The utilization of 97% for CPU instance 1 named '' has exceeded the 85% threshold on model bvtowpocwp03.central.rinterna.local for more than the acceptable time period.

1

Normal CPU Utilization. The CPU Utilization for all CPU instances is now below the 70% reset threshold for model bvtowpocwp04.central.rinterna.local.

2

High CPU Utilization. The utilization of 96% for CPU instance 1 named '' has exceeded the 85% threshold on model bvtowpocwp04.central.rinterna.local for more than the acceptable time period.

1

High CPU Utilization. The utilization of 92% for CPU instance 1 named '' has exceeded the 85% threshold on model bvtowpocwp04.central.rinterna.local for more than the acceptable time period.

1

## Issues

Duplicate dates in tables?

cpuFailTable$date[which(duplicated(as.POSIXlt(cpuFailTable$date)))]

[1] "2013-11-28 CET" "2013-11-28 CET" "2013-11-28 CET" "2013-11-28 CET" "2013-11-28 CET" "2013-11-28 CET"

[7] "2013-11-28 CET" "2013-11-29 CET" "2013-11-29 CET" "2013-11-29 CET" "2013-11-29 CET" "2013-11-29 CET"

[13] "2013-11-29 CET" "2013-11-29 CET" "2013-11-29 CET" "2013-11-29 CET" "2013-11-29 CET" "2013-11-29 CET"

[19] "2013-11-29 CET" "2013-11-29 CET" "2013-11-29 CET" "2013-11-29 CET"

> llply(subTableCpu,function(x){which(duplicated(x))})

$bvtowpocwp01.totta.corp

[1] 290 579

$bvtowpocwp02.totta.corp

[1] 290 579

$bvtowpocwp03.totta.corp

[1] 290

$bvtowpocwp04.totta.corp

[1] 290 579

> which(duplicated(cpuFailTable$date))

[1] 290 291 292 293 294 295 296 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599

> llply(subTableCpu,dim)

$bvtowpocwp01.totta.corp

[1] 867 2

$bvtowpocwp02.totta.corp

[1] 867 2

$bvtowpocwp03.totta.corp

[1] 578 2

$bvtowpocwp04.totta.corp

[1] 867 2

> dim(cpuFailTable)

[1] 887 5

### Hypothesis

Time 0:00 is present twice for each daily table?

Inspecting aggregated table file

"289" 2013-11-28 00:00:00 35.75 39.9 42.98 36.83 41.44 38.37 44.82 34.99 44.09 35.72 41.76 38.05 41.46 38.35 56.31 23.5 41.59 38.22 41.45 38.35 41.42 38.39 42.37 37.44 41.6 38.2 41.57 38.23 41.47 38.34 41.63 38.18 41.32 38.49 41.41 38.4 41.53 38.28 41.8 38 41.43 38.38 41.34 38.46 41.66 38.14 41.45 38.36 43.95 35.86 42.75 37.05

"290" 2013-11-28 00:00:00 35.75 39.9 42.75 37.05 41.49 38.32 47.45 32.36 43.02 36.79 41.96 37.85 41.46 38.35 56.95 22.85 41.47 38.34 41.3 38.5 41.56 38.25 41.9 37.91 41.64 38.17 41.39 38.42 41.43 38.38 41.48 38.33 41.34 38.47 41.55 38.26 41.41 38.4 41.81 38 41.48 38.33 41.29 38.52 41.38 38.42 41.41 38.39 44.6 35.21 42.94 36.87

Same date, different values!