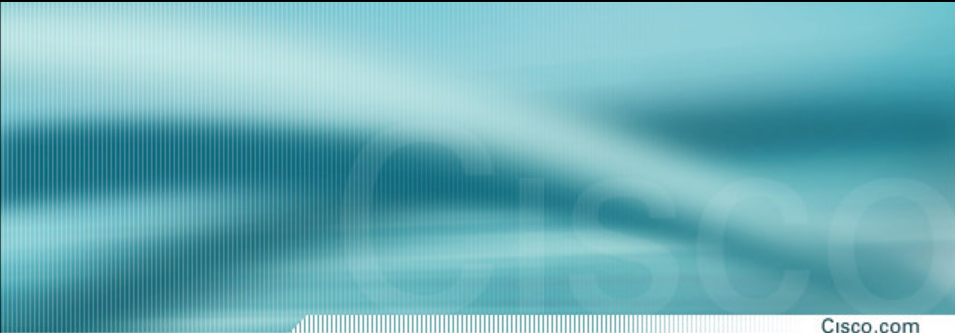




Cisco.com

Yet Another Data Aggregator An Oracle free Oracle

Ross Cartlidge
Network Consulting Engineer
Advanced Services, Cisco Systems Australia

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

2

What's the problem?

Cisco.com

- **Data is everywhere**
- **Good analysis can tell a lot about your networks past, present and future**
- **Un-analysed data can't**
- **Databases imprison data**
- **What about data without end**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

3

What do people usually do?

Cisco.com

- **Nothing**
- **Commercial data analysis systems**
- **Collections of ad-hoc scripts**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

4

Whats wrong with these approaches

- Nothing

Cisco.com

- **Collected data just takes up disk space**
- **You have no idea of trends**
- **You can't analyse past incidents**
- **No pretty graphs for management**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

5

Whats wrong with these approaches

- Commercial Systems

Cisco.com

- **Expensive**
- **Can be hard to customize**
- **Need a lot of setup**
- **Slow**
- **Do produce nice pictures**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

6

Whats wrong with these approaches - Ad Hoc Scripts

Cisco.com

- **Began because of a need (Good)**
- **Often do lots of similar things**
- **Hard to re-use code**
- **No defined interfaces**
- **Can be error prone**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

7

A third way?

Cisco.com

- **I was analysing Modem CallRecords for a large SP**
- **300,000 per day**
- **120 Mbytes**
- **Wrote some grep/awk/sh/sed scripts give**
- **Calls/min, average call length, error types...**
- **My service manager always wanted more...**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

8

Yada version 1

Cisco.com

- **Simple but general HTML/CGI front end**
- **Treats all fields equal**
- **Written in TCL but data processed in awk
– flexible and fast**
- **Simple graphical representation and
analysis with csv for export and further
analysis**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

9

It did a whole lot more than CallRecords

Cisco.com

- **Syslogs – 1million/day over years**
- **CDRs – 400 gateways, 20million/day,
1billion records in total**
- **Show Version data**
- **Sniffer traces - pdml**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

10

YADA problems

Cisco.com

- **Code was Monolithic**
- **Became hard to add new features**
- **Hard to script for auto report generation**
- **Hard to interface with other systems**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

11

YADA v2 - csvtools

Cisco.com

- **Unix tools pipeline approach**
- **Unix tools great idea that has been neglected in favour of monolithic perl/tcl/python scripts**
- **Uses “csv” style flat files**
- **Used Postgres functionality as a base**
- **Tightly integrated with bash**
- **Option of XML in and out for interworking**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

12

The Csvtools and their SQL equivalents

Cisco.com

- | | |
|--------------------|-------------------------------------|
| • Csvselect | • Select |
| • Csvproc | • Update |
| • Csvjoin | • Join |
| • Csvsort | • Sort |
| • Csvyada | • Groupby (but hierarchical) |
| • Csvunion | • Union |
| • Csv2tbl | • Tabular output |
| • Csv2xml | • Table as XML |
| • ... | • ... |

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

13

A real world example

Cisco.com

```
#!/bin/sh
ALL_DDTS=~/csv/ddts_10_11_12.csv
ddts_list=$1
train=$2
csvjoin -left ddts ddts_10_11_12.csv $ddts_list |
csvproc -call tel iosttrain -in release -out train |
csvselect -match train "^$train$" -all +print train |
csvsort ddts
```

Take a list of ddts's "\$1" and find the trains it is integrated in
And then look at the releases in a specified train.

Lets do it for real...

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

14

A really cool example :-)

Cisco.com

```
csvproc\  
< ~/csv/connect-0504290000.csv\  
-in time -out time_fibonacci -call awk fibonacci |  
csvsort -numeric time_fibonacci |  
csvyada +skip -yada -aggregate host - -newlevel -aggregate time_fibonacci |  
yada_scatter connect_fibonacci_scatter
```

Produce a Data-Xray of a days worth of call records highlighting call duration Categories

Lets do this for real...

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

15

Cool things is has ☺

Cisco.com

- **Auto generation of bash completion on options and fields (dynamic)**
- **Auto usage and man page generation**
- **Up to 300000 records/sec on this lap top (1.5G Pentium M)**
- **Easy to add new tools**
- **Language independent**
- **Uses Tcl, Awk, Perl and Bash**
- **Tcl control plane, Awk data plane**
- **Easy to cluster**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

16

Real soon now

Cisco.com

- **Tutorial for users and developers**
- **Modular tool generated Web front-end**
- **Integration with RDBMS (easy)**
- **Output as fancy (not csv) MS Excel to exploit cool presentation features and ubiquity**
- **More flexible graphical output**
- **Transparent clustering**

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

17

CISCO SYSTEMS



EMPOWERING THE
INTERNET GENERATION

Cisco.com

Questnet 2005

© 2005, Cisco Systems, Inc. All rights reserved.

18