Using server performance statistics to make consortium fees fairer
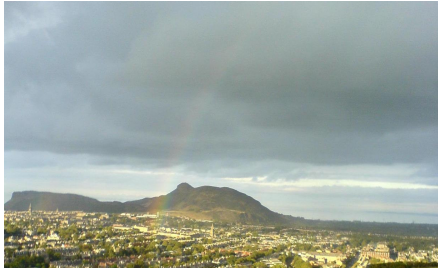
Clare Whittaker

IGeLU 2007 Brno, 3 September.

sdlc scottish digital library consortium

How we used server performance statistics to make Consortium membership fees fairer

A script which captures system load

Analysing the data

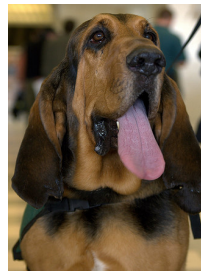**A new price structure**

Dire performance problems

- Circ hanging & system slowness coincided with a new site joining the Consortium
- more unplanned downtime in 6 months than in the previous 5 years

Attribution
http://www.flickr.com/photos/14516334@N00/337095799/

We tracked down a random memory segment error

- Was it ad-hoc SQL?
- Were our sizing metrics wrong?

There were a number of triggers…

Attribution:
http://www.flickr.com/photos/danakay/126738338/

Production Infrastructure autumn 2004

Mirrors

Database Server

Application Server Academic Hub

Application Server Special Libraries' Hub

---

sdlc scottish digital library consortium

nls National Library of Scotland

- Founded 1999 as Scottish Voyager Consortium
- Financial savings
- Operational benefits
- Cross-sectoral synergy
- A successful model
- Organic growth

Slide 1

sdlc scottish digital library consortium

members using Voyager

nls National Library of Scotland

THE UNIVERSITY OF EDINBURGH

eca edinburgh college of art

ROYAL OBSERVATORY EDINBURGH

HERIOT WATT UNIVERSITY

UNIVERSITY of ABERTAY DUNDEE

Slide 2

Analysing load in the SDLC environment

- newvs
- unix commands

e.g.

```
ps-ef grep
   [database name]
```

- Inhouse script

Attribution:
http://www.flickr.com/photos/graeme
newcomb/358329131/

Slide 3

A script to snapshot system load

Captures hourly a count of svr processes
  By service
  By database

Allows us to monitor peaks by database (institution)

Attribution:
http://www.flickr.com/photos/sandwick
/369211852/

Slide 4

Perl script (1)

```
open (PS, "/usr/ucb/ps -axuwww| grep '^voyager' |") || die ("ps failed\n");
$num_oracle_ps=`/usr/bin/ps -fu oracle | /usr/bin/grep oracleLIBR | /usr/bin/wc -l`;
$datestamp=`date +%d/%m/%y-%H%M`;
chomp $datestamp;
while ( $Line = <PS> )
{
  $argstring = substr($Line, 61);          # argument line
  @args = split(/\s+/, $argstring);        # split into components
  @cmd = split(/\//, $args[0]);                    # get the command
  $command = $cmd[$#cmd];                           # get the basename of the
command
  if ( $command eq "circsvr" || $command eq "z3950svr" ||
        $command eq "opacsvr" || $command eq "keysvr" ||
        $command eq "catsvr" || $command eq "acqsvr" )
```

Slide 5

Perl script (2)

```
# The last argument in $args will be the config file - the
 # third dir (eg /m1/voyager/dbname/foo) will tell you the database name
# and so the institution
  ($a, $b, $c, $dbname, $rest) = split(/\//, $args[$#$args]);
  $cmdusage{$command}++;
  $dbusage{$command}{$dbname}++;
  $alldb{$dbname}++;

#print "Voyager usage totals by database \n";
foreach $D (sort {$alldb{$b} <=> $alldb{$a} } keys %alldb)
{
#   print "$D = $alldb{$D} \n";
  printf "%s:total (by database):%s::%d\n",$datestamp,$D,$alldb{$D};
}
```

Slide 6

Perl script (3)

```
# Record some vague performance stats
$lav=`/usr/bin/uptime | /usr/bin/awk '{ print \$(NF-2) }' | sed 's/,\$//'`;
($freemem, $freeswap)=split(':', `/usr/bin/sar -r 1 2 | /usr/bin/tail -1 |  /usr/bin/awk '{
print \$2 ":"  \$3}'`);
$pgscan=`/usr/bin/sar -g  1 2 |  /usr/bin/tail -1 | /usr/bin/awk '{ print \$(NF-1) }'`;

chomp $lav;
chomp ($freeswap);
chomp $pgscan;

printf "%s:load - cpu lav:::%s\n",$datestamp,$lav;
printf "%s:load - mem free:::%s\n",$datestamp,$freemem;
printf "%s:load - swap free::%s\n",$datestamp,$freeswap;
printf "%s:load - page scan::%s\n",$datestamp,$pgscan;
```
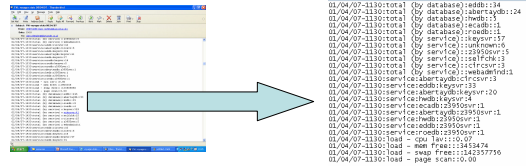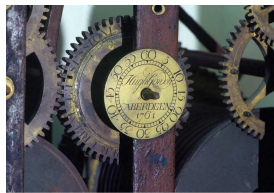
Analysing the data

Tools already in use for processing logs

- Standard unix tools
- Perl
- Logsurfer

Why we chose Excel as our analysis tool

- Needed to extract the information quickly
- Started by looking at a subset of the log data
- Existing familiarity with Excel

Reprise: the information we use

- We take an hourly snapshot of processes running against each institution's Voyager database
- We total the hourly count "total (by database)" for each institution
- This gives a weekly figure which we can use as a indicator of system use

What we expected to find (1)

nls
National Library of Scotland

=

THE UNIVERSITY OF EDINBURGH

Were the founder members still equal partners?

What we expected to find (2)

THE UNIVERSITY OF EDINBURGH

UNIVERSITY of ABERTAY DUNDEE

Were some universities hammering the system more than others?

Institutional breakdown of process use

Consortium process use

all NLS	28
signetsdb	1
cassdb	1
roedb	1
hwdb	9
eddb	47
ecadb	3
abertaydb	11

Series1

(x-axis: 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50)
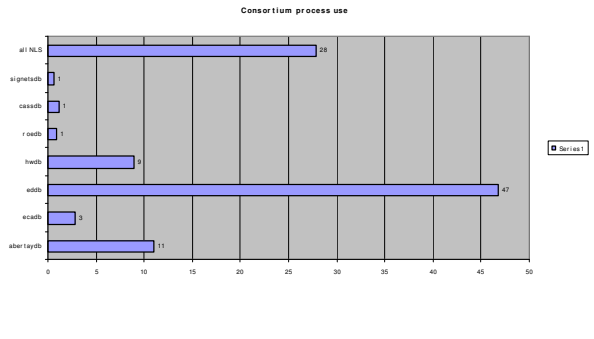
Bib count per institution

Bib count by institution

7,215
5,447
1,118,689
1
13,600
76,911
149,107
93,140
4,997,651

Legend: CASS, NLS *, Signet, EUL, ROE, ECA, HWU, UAD

(x-axis: 0, 1,000,000, 2,000,000, 3,000,000, 4,000,000, 5,000,000, 6,000,000)
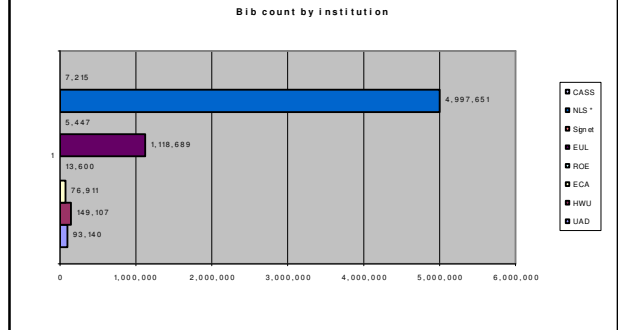
Academic and special libraries compared

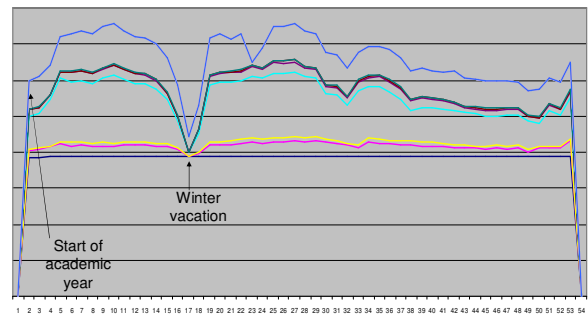Legend: UAD, HWU, ECA, ROE, EUL, Signet, NLS *, CASS
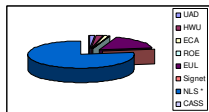
Bib count

National Library = 78%

Other libraries = 22%
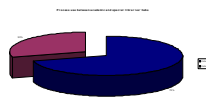
Process use

Academic libraries = 70%

Special libraries = 30%

Seasonal patterns

Winter vacation

Start of academic year

(x-axis: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54)

MEANWHILE...

Split hub architecture helped our performance problems

Production Infrastructure autumn 2005

Mirrors

Database Server, academic hub

Application Server Academic Hub
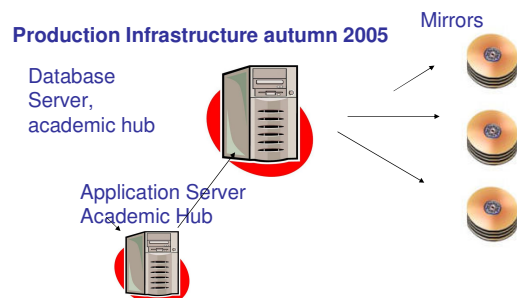
Patterns of server use by different library types
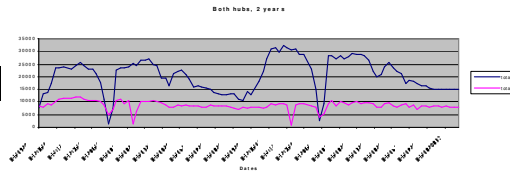
A new price structure

Our old charging model needed to be updated

- Was based on Endeavor's pricing model at the point when a library joined SDLC
- Took no account of system use or relative size of an institution in the Consortium

Indicators we use to compare Consortium members

Attribution: http://www.flickr.com/photos/rusty-projector/235037137/

- database size (number of bibs)
- number of staff user licenses
- peak number of connections to the OPAC
- number of Voyager server processes by database (i.e. library)

"% of Consortium"

expresses each library's relative size in the Consortium as a percentage which can be applied to the total annual running costs of the Consortium across both hubs.

"% of Consortium" is applied to

- Sun server costs
- Facilities management
- Consortium support costs
- Consortium management costs

New charging system in place for Financial Year 2005-6

- Consortium Board approval
- Member agreement
- Income remained the same
- Costs more fairly distributed

© National Museums Scotland.
Licensor www.scran.ac.uk

The future …

Refine the script

Better tools for analysing the data → **More refined analysis**

Costing model subject to further revision

Questions?

sdlc
scottish digital
library consortium

Thank you!

Clare Whittaker
clare.whittaker@ed.ac.uk
http://www.nls.uk/sdlc