



The Life of a SysAdmin

in a Research Environment

Eric G. Burgueño | <http://ebur.co>



\$ whoami

I am Eric

a SysAdmin @



(...because being a lawyer was too hard)

Reproducible Research

The Role

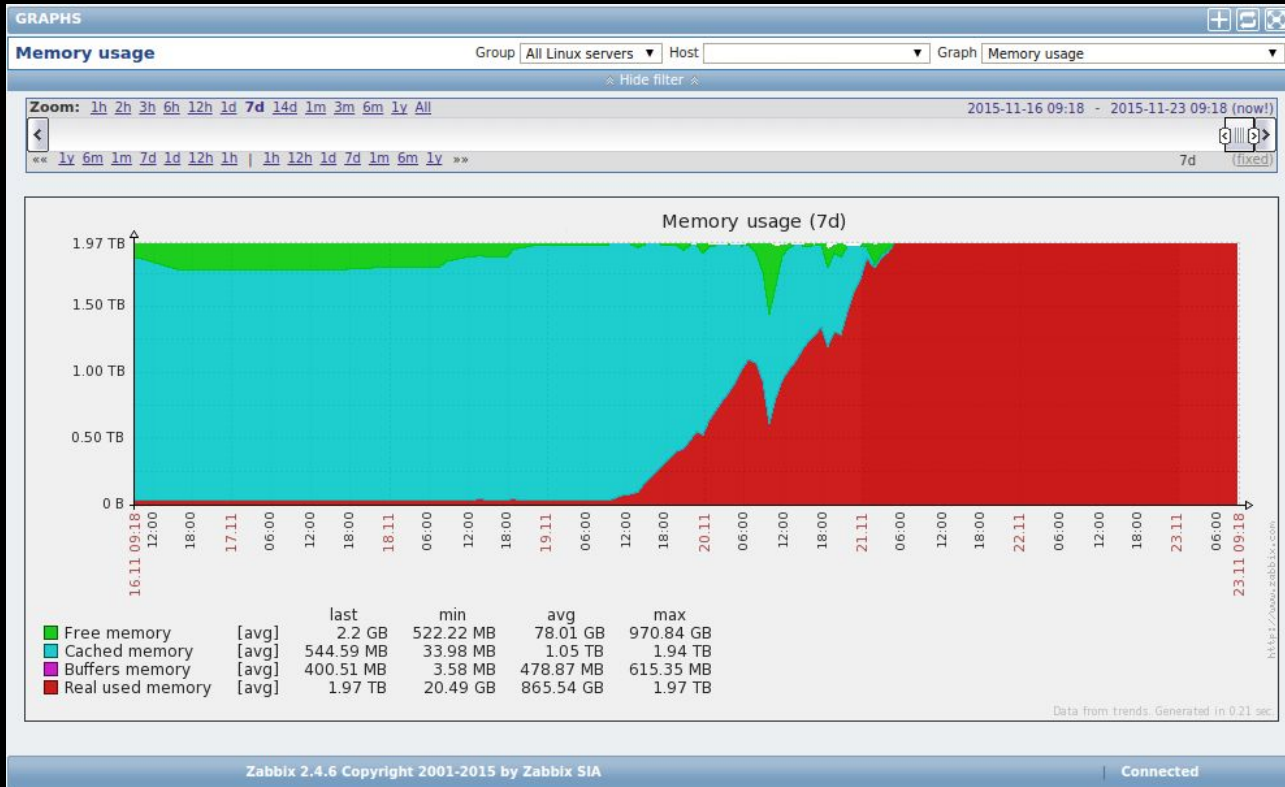
- Science > Money
- Makefiles <3 Visio
- Unus pro omnibus...



The Environment

- Some very big servers
- Running specialised software
- Heavily customised

Some very BIG servers...



...and some small ones

```
$ lshosts
```

HOST_NAME	ncpus	maxmem	maxswp
host01	4	15943M	8015M
host02	8	32102M	8015M
host03	4	15943M	8015M
host04	16	84617M	32767M
host05	8	32102M	8015M
host06	80	2068067M	10239M
host07	80	2068067M	10239M
host08	16	84617M	32767M
host09	8	258326M	8015M
host10	40	258269M	10239M
host11	32	129013M	2047M
host12	32	129013M	2047M
hostnn	32	129013M	2047M

Specialised Software

Genome/Transcriptome assembly
Phylogenetic analysis
RNA analysis tools
Protein annotation
Search pattern, motif, profiles domains, families
Data visualisation
Other tools
Find transcriptional elements
Organism classification
Align sequences
DNA Sequence analysis tools
Detection of repeats, restriction sites, and patterns
Gene finding
Sequence alignment
Find tRNA
Expression analysis
Protein function assignment
Search, computation & analysis of pathways
Protein-protein interactions
Protein sequence analysis
Multiple sequence alignment

Customised Environment

```
[eric@host01]$ lsd  
openlava project 2.2, Oct 19 2015
```

```
My cluster name is genomecluster  
My master name is host01
```

```
[eric@host01 myapp-5.15-54.0]$ grep temp launcher.sh  
TMP="$(dirname $0)/temp"
```

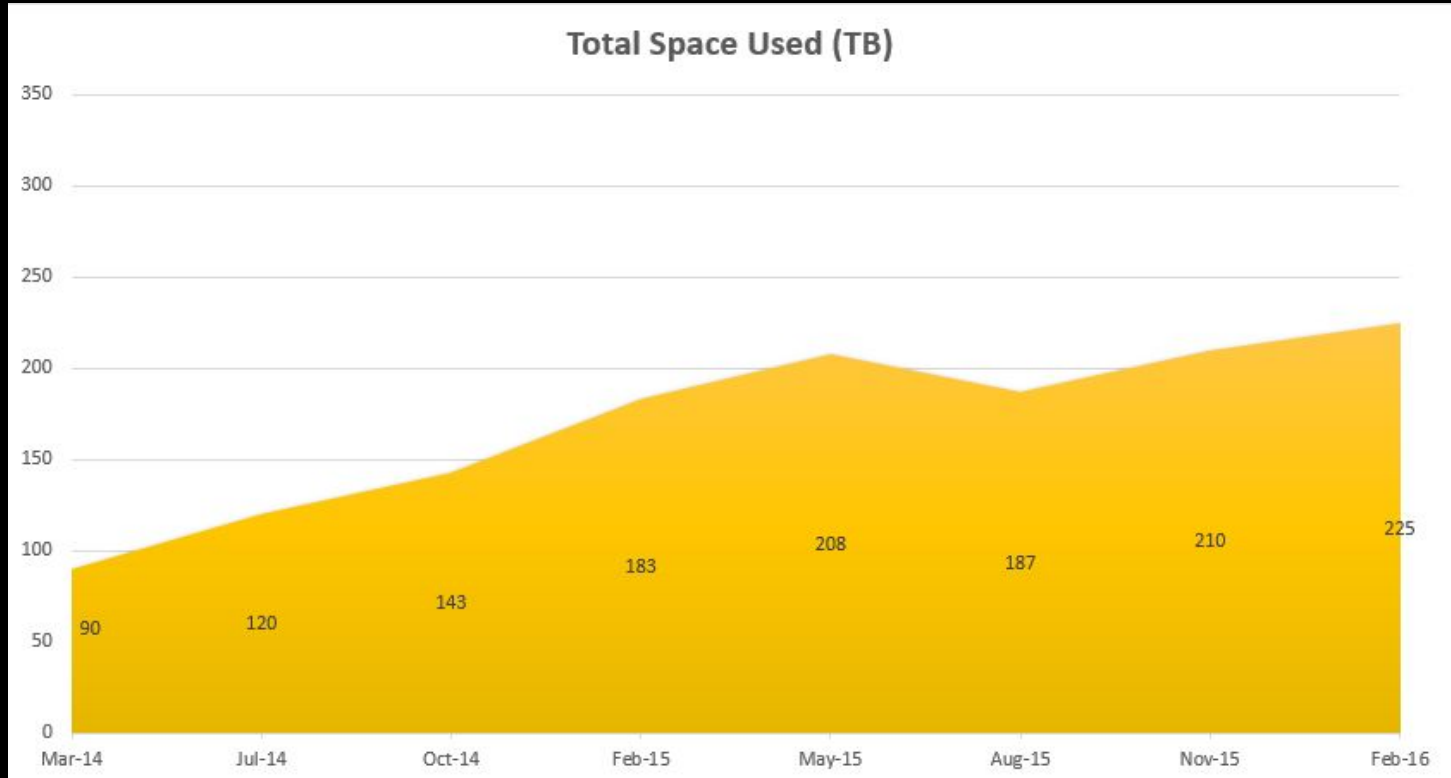
```
[eric@host01 myapp-5.15-54.0]$ pwd  
/apps/myapp-5.15-54.0
```

```
[eric@host01 myapp-5.15-54.0]$ ll temp  
lrwxrwxrwx. 1 nonpriv nonpriv 35 Jan  5 12:38 temp ->  
/nfs/appscratch/myapp/temp
```

The Struggles

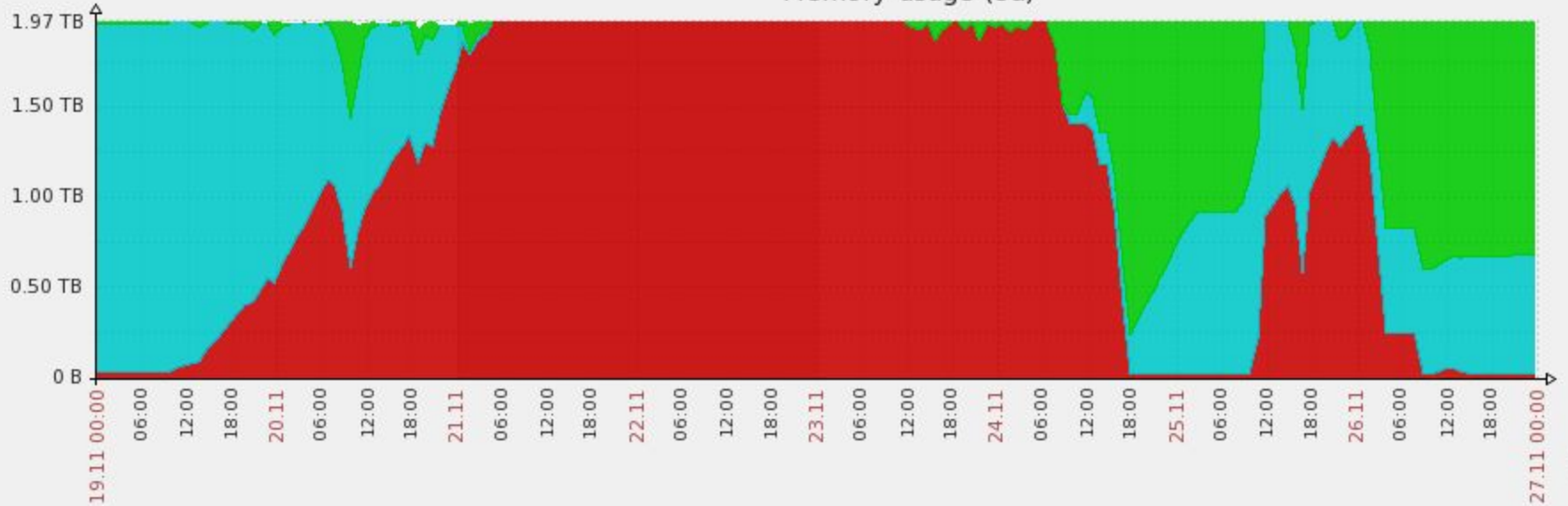
- Some very big servers
- Running specialised software
- Heavily customised

Some very BIG servers...



Some very BIG servers...

Memory usage (8d)



		last	min	avg	max
Free memory	[avg]	1.3 TB	486.14 MB	310.35 GB	1.75 TB
Cached memory	[avg]	671.52 GB	15.25 MB	545.23 GB	1.94 TB
Buffers memory	[avg]	404.71 MB	1.93 MB	319.92 MB	615.35 MB
Real used memory	[avg]	17.86 GB	13.33 GB	1.14 TB	1.97 TB

Some very BIG servers...

“With the expected growth in the size of data sets, moving to distributed-memory parallel computing seems inevitable.”

Bill - July 12, 2015 at 3:33 pm

<http://bit.ly/1WCqSIQ>

Specialised Software

```
$ ls /software/
```

```
454
```

```
AATpackage
```

```
AATpackage-r03052011
```

```
AdapterRemoval-1.4
```

```
AdapterRemoval-1.5.4
```

```
allpaths1g
```

```
allpaths1g-44837
```

```
allpaths1g-44913
```

```
allpaths1g-49722
```

```
allpaths1g-50191
```

```
amos
```

```
AMOS
```

```
amos-3.0.1
```

```
amos-3.1.0
```

```
AMOS-Short
```

```
Anchor-0.3.1
```

```
antismash
```

```
anytag-2.5.2
```

```
arachne
```

```
arachne-39099
```

```
aragorn
```

```
cgal-0.9.6-beta
```

```
CGAT-0.2.4
```

```
circos
```

```
circos-0.62-1
```

```
clustal-omega-1.1.0
```

```
clustal-omega-1.2.1
```

```
clustalw-2.1
```

```
cndsrc
```

```
cndsrc
```

```
cndsrc
```

```
cndsrc
```

```
cndsrc
```

```
correction
```

```
correction
```

```
cortex_con
```

```
cortex_var
```

```
crac-1.2.3
```

```
cufflinks
```

```
cufflinks-1.3.0.Linux_x86_64
```

```
cufflinks-2.0.2.Linux_x86_64
```

```
cufflinks-2.1.1.Linux_x86_64
```

```
geneid_v1.4.4.Jan_13_2011
```

```
genomemapper0.4.3
```

```
genometools
```

```
genometools-1.3.3
```

```
genometools-1.5.6
```

```
genometools
```

```
genometools
```

```
genometools
```

```
genometools
```

```
gmap
```

```
gmap-2014-07-04
```

```
gmap-2014-10-22
```

```
gmap-2014-12-29
```

```
gmap-2015-09-29
```

```
gmato
```

```
gmatoV1.2Build20130106
```

```
gmes
```

```
gmorse
```

```
gmorse_02mar2011
```

```
gmp-6.0.0
```

```
hagfish
```

Dependency Hell

Customised Environment

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.



YEAH!

SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.

Customised Environment

```
[eric@host01]$ aws-shell || gcloud  
-bash: aws-shell: command not found  
-bash: gcloud: command not found
```

```
[eric@host01]$ ./corkscrew ssh login01@eu.collaborator.edu  
'bhosts || qstat -a || qhost'  
-bash: bhosts: command not found  
-bash: qstat: command not found  
-bash: qhost: command not found
```

```
[eric@host01]$ :(
```

Reproducible? Research

The Solution?

“

Put it in a container!

Dear Google: Please give us sparkles in Google Slides.



Thank you

...any questions?

Eric G. Burgueño | <http://ebur.co>