Ellexus: The I/O Profiling Company

Dr Rosemary Francis

CEO

Good I/O evangelist

Tuning I/O and sizing storage for the cloud with the Sanger Institute

Originally presented at the AWS Life Sciences workshop, London 2018

Ellexus Ltd: The I/O Profiling Company

Products: We make tools to help you

- improve application performance,
- protect shared storage, and
- manage application dependencies.

Customers include:

TACC · TEXAS The University of Texas at Austin · diamond

wellcome sanger institute · DELL EMC

ANSYS · arm · Mentor A Siemens Business

Ellexus enterprise products

Take control of the way you access your data

Detailed I/O Profiling

- Dependency analysis
- Cloud migration made easy
- Debug devops and I/O issues

Live System Telemetry

- I/O profiling at scale
- Protect storage from rogue jobs
- Find bottlenecks in production

Tuning cancer pipelines at the Sanger Institute

Cancer, ageing and somatic mutation group have put in a lot of work to optimise their pipelines for the Pancancer project:

→ 2000 whole genomes (each sample can generate 250GB data)
→ Pipeline has to be portable (take the compute to the data)
→ Docker pipelines developed and made available
→ I/O tuned to make cloud viable

"With the Ellexus tools we were able to identify why we were hitting I/O bottlenecks when we expected full CPU utilisation"

Kerian Raine, Cancer Researcher, Sanger Institute.

Where you can find the CASM pipeline

Pipeline:

https://github.com/cancerit/dockstore-cgpwgs

How to run chromosome 21 (the smallest)

```
ds-cgpwxs.pl -r /data/step2/input/core_ref_GRCh37d5.tar.gz
-a /data/step2/input/VAGrENT_ref_GRCh37d5_ensembl_75.tar.gz
-si /data/step2/input/SNV_INDEL_ref_GRCh37d5-
fragment.tar.gz -t /data/step2/input/COLO-829.bam -tidx
/data/step2/input/COLO-829.bam.bai -n
/data/step2/input/COLO-829-BL.bam -nidx
/data/step2/input/COLO-829-BL.bam.bai -exclude
"1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,22,X,Y,
MT,NC_007605,hs37d5,GL%" -outdir /data/step2/output
```

Case Study: Sizing storage for AWS

Ubuntu 18.04 m5.xlarge or m5d.xlarge
(15 ECUs, 4 vCPUs, 2.5 GHz, Intel Xeon Platinum 8175, 16 GiB memory, EBS only)

HDD:

- Magnetic EBS

NVMe:

- 1 x 150GB (SSD) NVMe

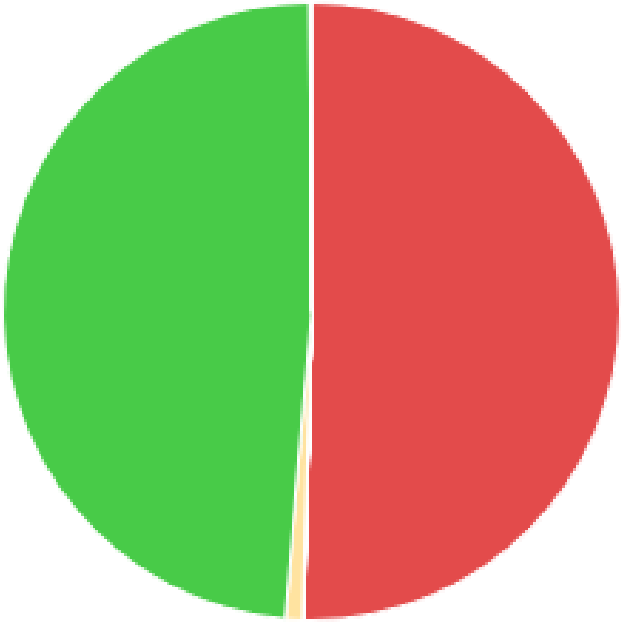
How long did it take?

1 x 150GB NVMe SSD	51m 27s	$191.79 /mth
Magnetic EBS HDD	1h 01m 44s*	$ 174.43 /mth

⇒ Initial conclusion is that SSD is worth it
20% faster for 10% cost increase.

*profiling overhead over head (on magnetic storage) is:
0.3% for Breeze and 0.08% Mistral of the total run time or
0.73% Breeze and 0.2% Mistral during intensive I/O

I/O patterns: Standard HDD vs NVMe SSD



Magnetic HDD

- 12m27s small reads
- 4s small writes
- 13s sync operations
- 12m3s good streaming I/O

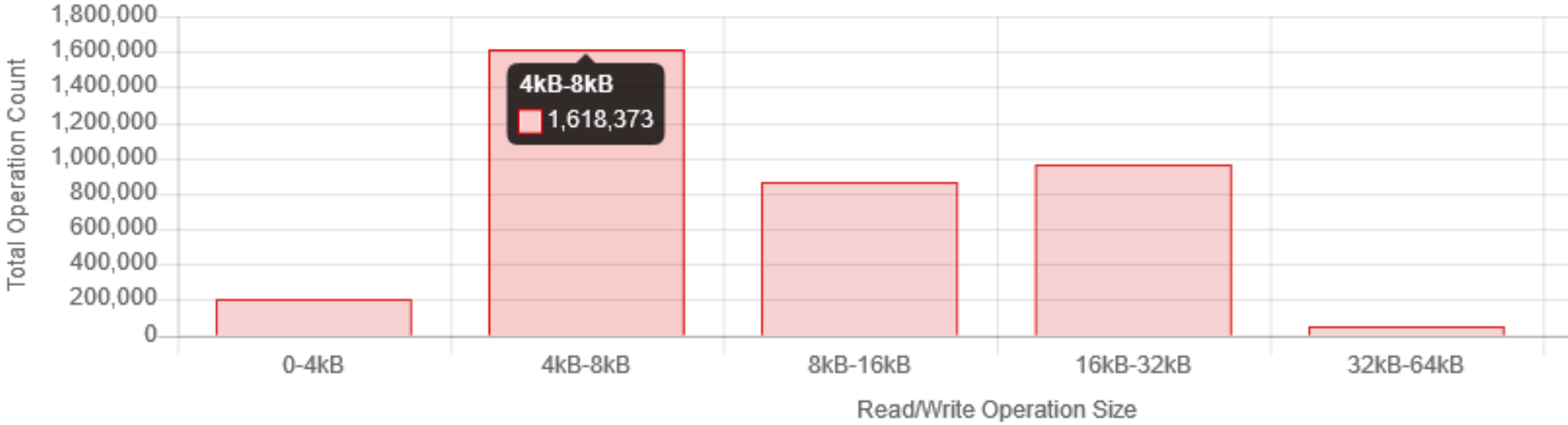
NVMe SSD

- 12s small reads
- 4s small writes
- 43s sync operations
- 53s good streaming I/O

A closer look at small reads

Small I/O is always a problem in life sciences

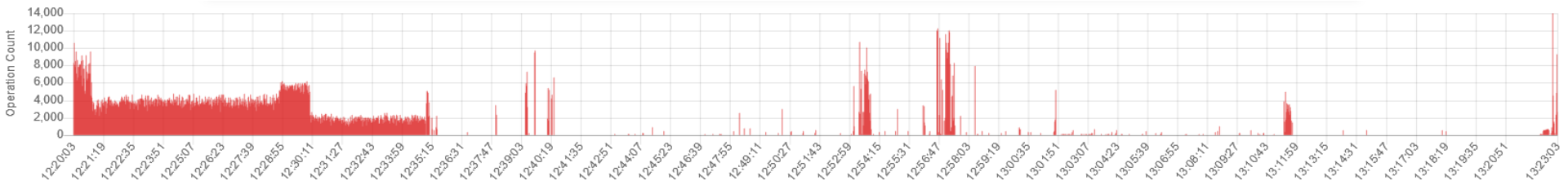
In this pipeline the I/O is not that small:
Reads in the same block have been aggregated



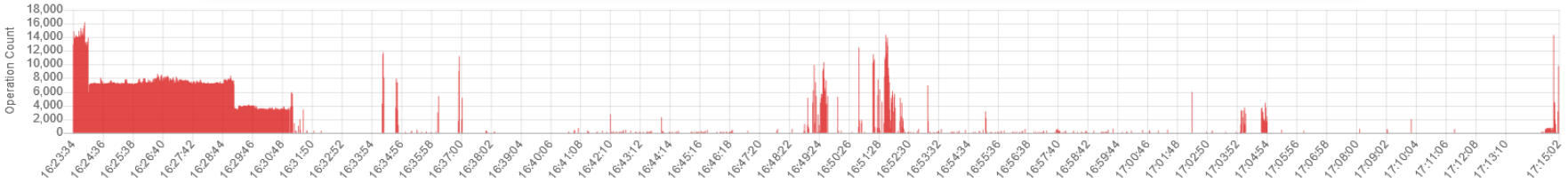
Number of read/write operations by I/O size

IOPs over time

Number of I/O operations over time on Magnetic HDD



Number of I/O operations over time on NVMe SSD



This shows why a big variation in I/O time didn't have a big impact on run time

A wider storage survey

Ubuntu 18.04 m5.xlarge or m5d.xlarge
(15 ECUs, 4 vCPUs, 2.5 GHz, Intel Xeon Platinum 8175, 16 GiB memory, EBS only)

SSD:

- GP2
- Provisioned 100 IOPS
- Provisioned 500 IOPS
- 1 x 150GB (SSD) NVMe

HDD:

- Magnetic EBS
- Throughput optimised HDD (500GB)

Storage comparison

	Time*		Cost per month	
GP2	52m 23s	100%	174.11	100%
Magnetic EBS	1h 01m 44s	118%	174.43	100%
Provisioned 100 IOPS	1h 42m 01s	195%	184.61	106%
Throughput optimised HDD	1h 19m 32s	152%	189.01	109%
150GB NVMe	51m 27s	98%	191.79	110%
Provisioned 500 IOPS	54m 22s	104%	215.01	123%

⇒ The throughput optimised HDD performed very badly
⇒ The Provisioned IOPS SSDs also weren't enough

⇒ AWS default option, GP2 is the best
(NVMe is only 2% faster for 10% price increase)

How long did this work take?

One day to work out how to run the pipeline

One day to run the experiments

$23 to test all the different configurations

We saved 10% of cloud costs for the project by not having to pick the fastest storage.

The experiment should be re-run with a whole genome as the trade offs are sensitive to the amount of data.

… and then rerun every few months to check performance –cost trade offs as AWS evolves their solution

Summary

Improving run time often doesn't require extensive rewrites.
Knowing where to look is key.

Keiran Raine
CASM, Sanger Institute

I/O profiling is important for performance and cost

Understanding dependencies and I/O patterns lets you take control of your data

Dr Rosemary Francis
CEO and director of technology
rosemary.francis@ellexus.com