

Catching rogue jobs before they overload the storage: the importance of I/O profiling

Dr Rosemary Francis, CEO and Co-founder



The I/O Profiling Company - Protect. Balance. Optimise.

www.ellexus.com

Who we are

Ellexus is the I/O profiling company.

We work in HPC and wide-scale computing.
Our latest tool Mistral has been developed
with ARM.

This case study is taken from a real customer
deployment using LSF and Spectrum Scale.

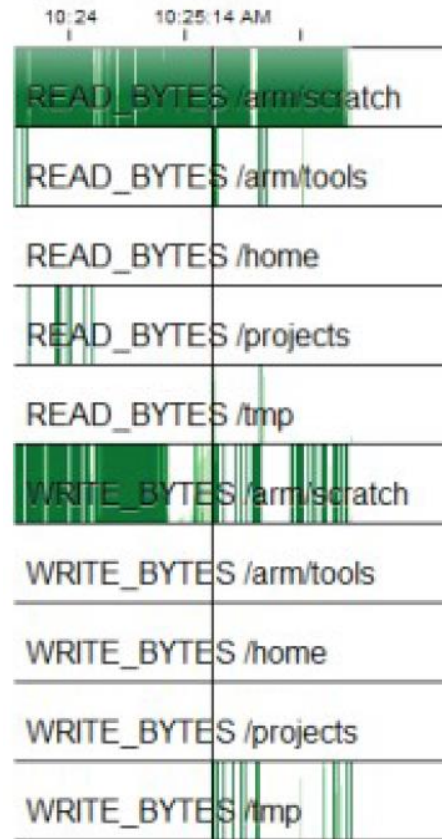
Customers include:



Mistral: Solving the noisy neighbour problem

In a compute cluster it is possible for a small number of jobs to overload the shared file system. **Example of a rogue job from ARM:**

Lots of data is written to remote storage (/scratch).



But almost nothing is written to local storage (/tmp).



I/O Profiling goals

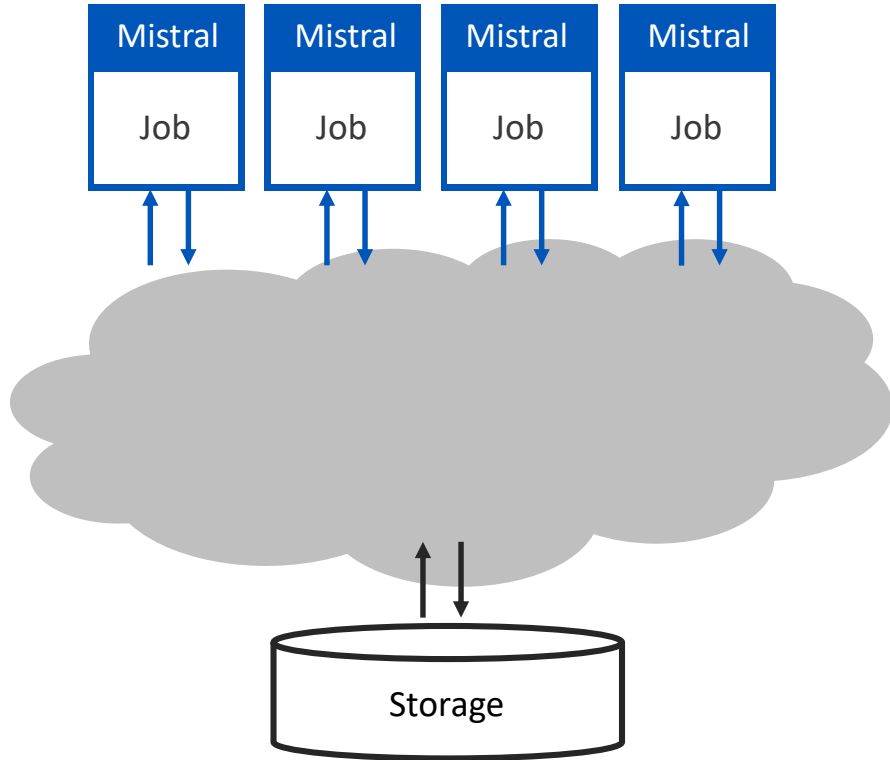
Protect your storage from bad I/O patterns and rogue jobs.

Load balance your storage for good quality of service.

Optimise your applications for the storage and your infrastructure for your jobs.



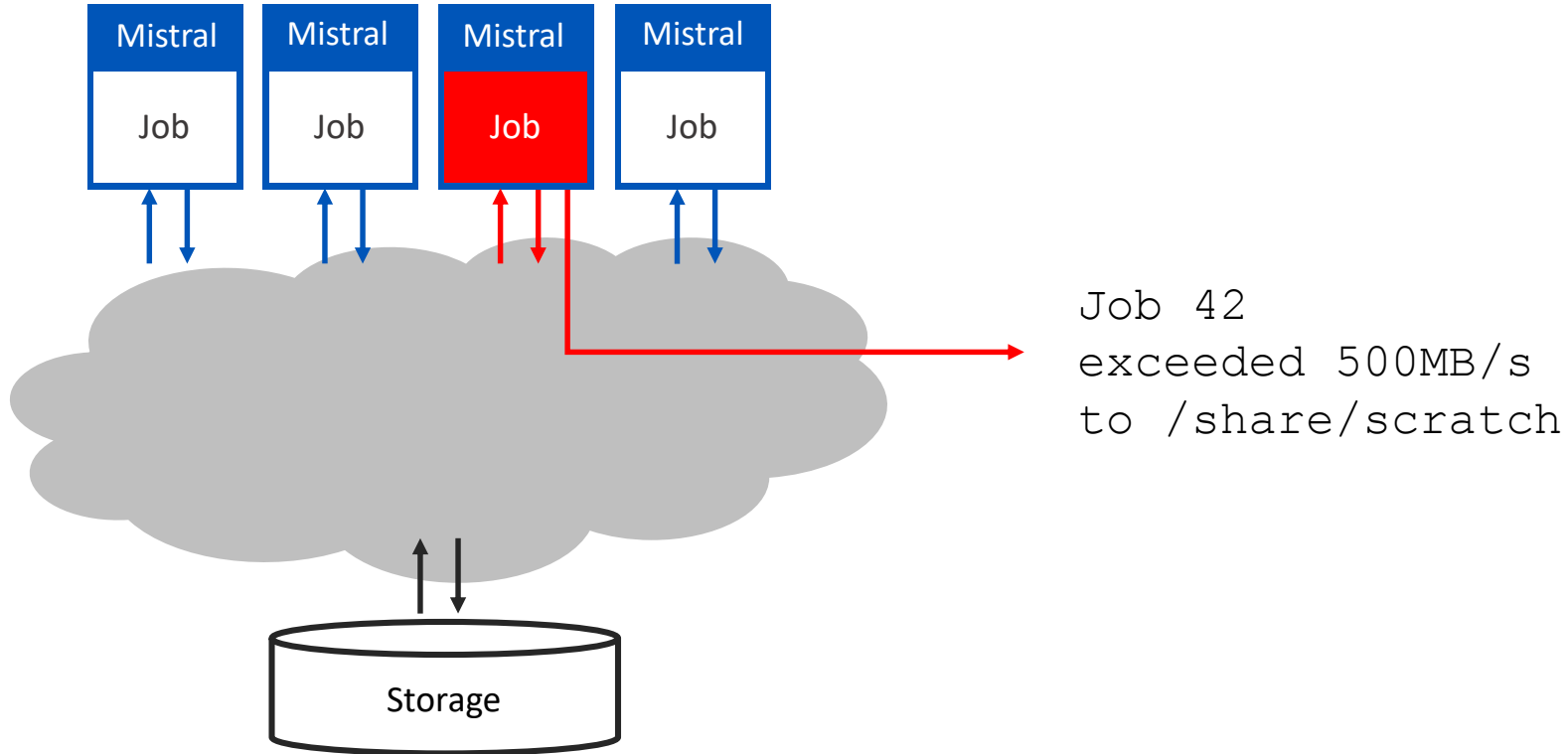
Mistral: How does it work?



Mistral sits at the application level.



Mistral: How does it work?

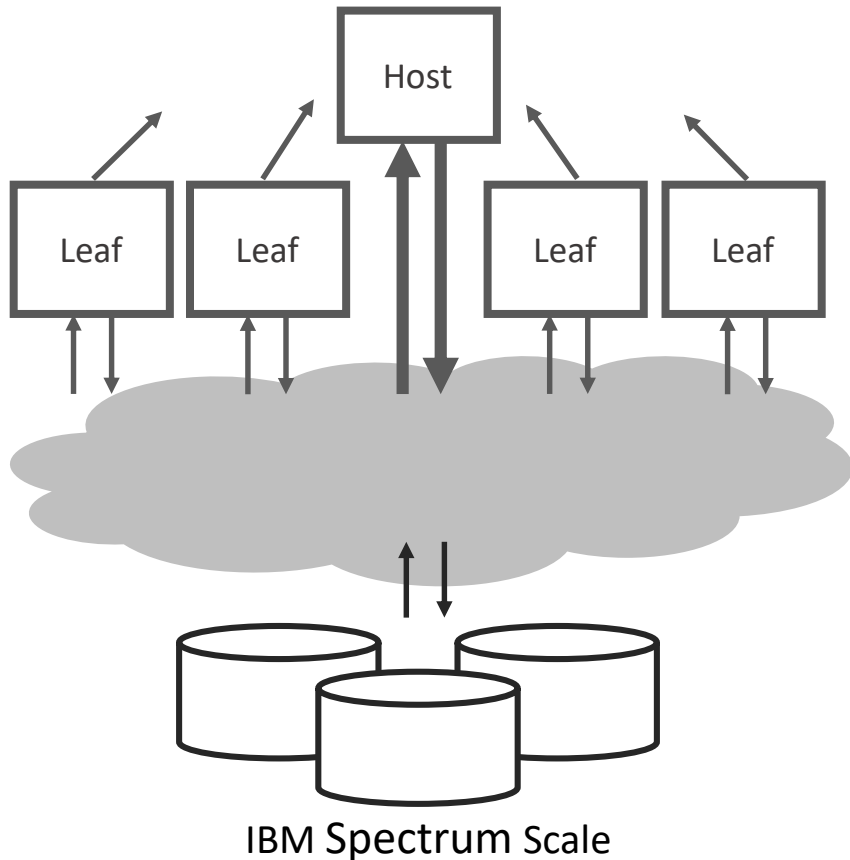


Customer case study: Catching rogue jobs

... stories from a Silicon Valley software 'startup'



Customer case study: system architecture



Host-Leaf architecture

Host compute nodes:

- Native Spectrum Scale
- 4 Gigabit/s Infiniband
- 32 Slots

Leaf compute nodes:

- Send data over RCP to host nodes
- Clustered NFS

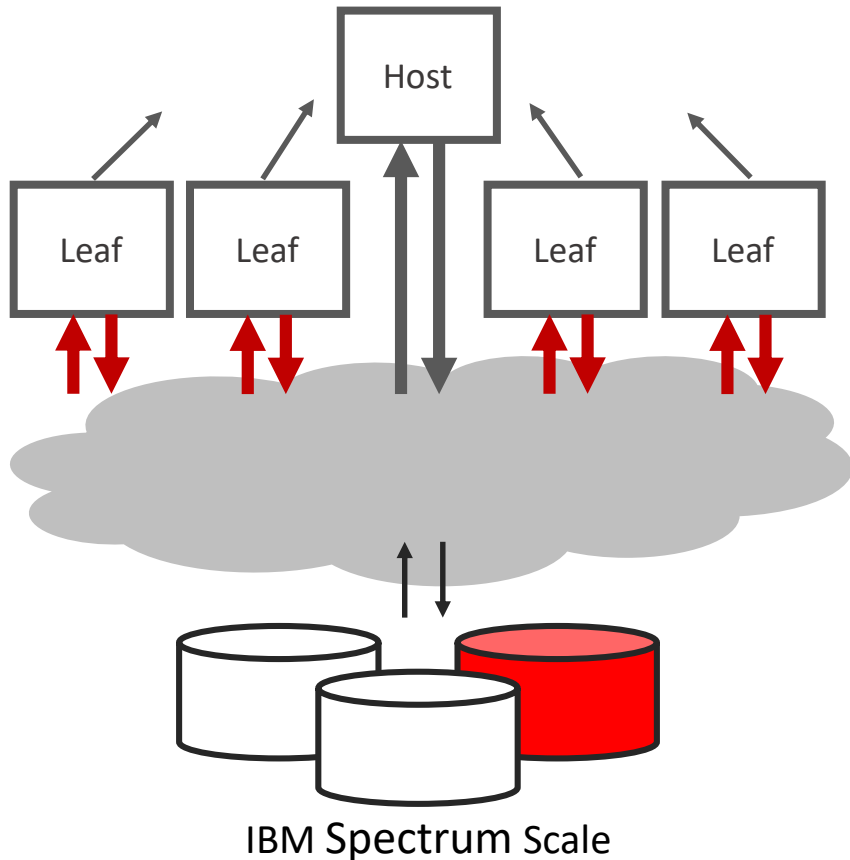
Storage:

- 8 Spectrum Scale servers



Customer case study

Problem 1: Bad working practices



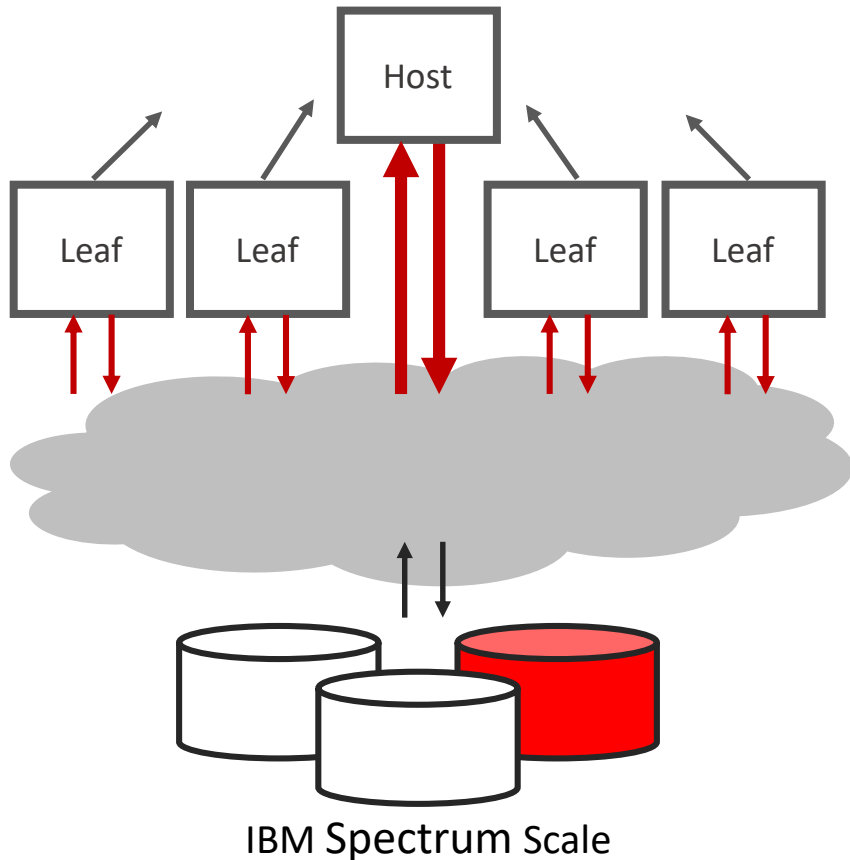
Jobs sometimes write from the leaf nodes instead of from the host.

If they hammer a particular mount point that can overload one of the filers.



Customer case study

Problem 2: Debug flag left on

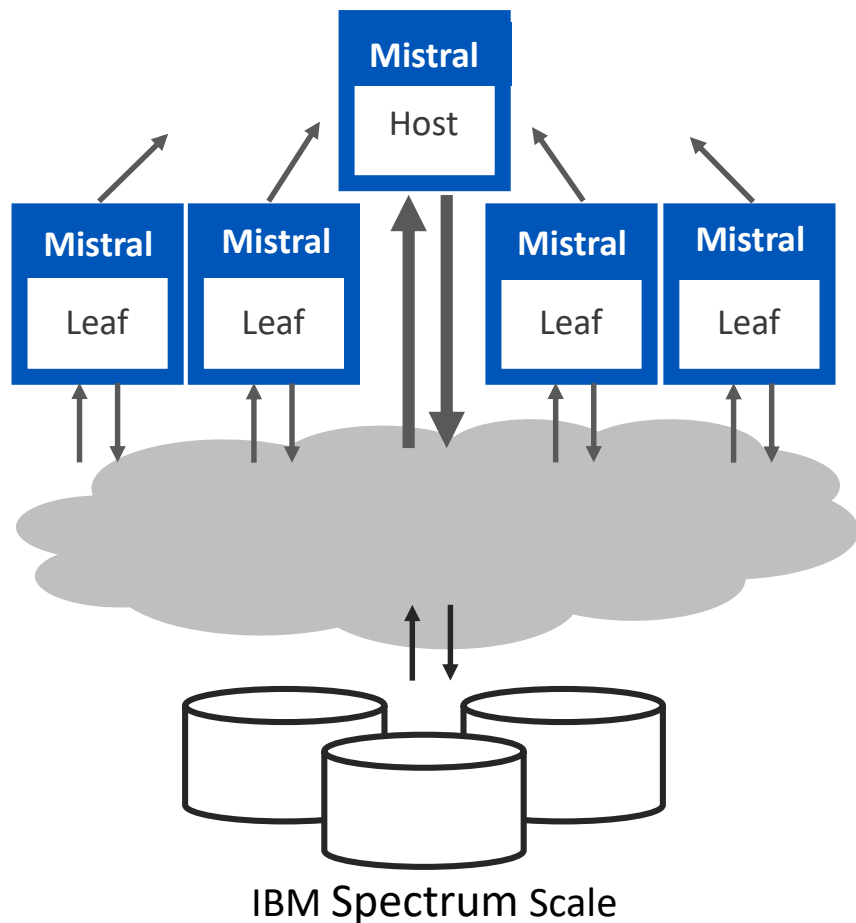


Job is debugged on one core then handed over to run at scale – debug flag left on accidentally.

This happens weekly.



Phase 1: Mistral deployment



Production jobs are wrapped in Mistral via an LSF job starter.

Mistral generates an alert if:

- Leaf nodes write to the file system
- Host nodes write too much data

Two use cases:

- My house is on fire
- My house was on fire, but I missed it

Integration with RTM makes this possible.



Example contracts

Example contract:

monitoringtimeframe, 60s

NAME	PATH/MOUNT POINT	IO TYPE	LIMIT
M1,	/gpfs/,	read,	500MB
M2,	/gpfs/,	write	500MB
M3,	/gpfs/,	create+delete,	2000

Example alert:

TIME	RULE	VIOLATION	PID	PROGRAM	FILE	JOB ID
17:30:05	M1	643MB/60s,	2784,	/tools/bin,	/gpfs/tools/config,	12
17:30:05	M2	505MB/60s,	2784,	/tools/bin,	/gpfs/data/file01,	12
17:30:06	M3	7405/60s,	2784,	/tools/bin,	/gpfs/data/set05,	12



Mistral RTM integration: Who is using the storage?

Current Mistral Events

Cluster: All Queue: All Project: All Rollup Data Go Clear

Search: Rule: All Records: 30 Refresh: 1 Minute

Presets: Last Day From: 2017-03-05 14:36 To: 2017-03-06 14:36 2 Hours

Showing Rows 1 to 30 of 24763 [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21...]

Cluster Name	Job ID	Host	Queue	Project	User	Rule	PID	Observed	Time	Logged Time
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31873	65.00 MB	1.00 Sec	2017-03-06 12:07:57
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31873	72.00 MB	1.00 Sec	2017-03-06 12:07:56
Motown	155471[995]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	123.64 MB	1.00 Sec	2017-03-06 12:07:56
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31873	159.00 MB	1.00 Sec	2017-03-06 12:07:55
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31873	65.00 MB	1.00 Sec	2017-03-06 12:07:54
Motown	155471[995]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	137.00 MB	1.00 Sec	2017-03-06 12:07:54
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31873	147.00 MB	1.00 Sec	2017-03-06 12:07:53
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:53
Motown	155471[995]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:52
Motown	155471[995]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:52
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:51
Motown	155471[995]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:50
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:50
Motown	155471[995]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:49
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:49
Motown	155471[995]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:48
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:48
Motown	155471[995]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:47
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:46
Motown	155471[995]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:46
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:45
Motown	155471[995]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:45
Motown	155471[998]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:44
Motown	155471[995]	platidpxn04	default	default		Higher_reads_gpfsfpo	31983	42.00 MB	1.00 Sec	2017-03-06 12:07:44
Motown	155471[998]	platidpxn04	default	default		Higher_writes_gpfsfpo	2	42.00 MB	1.00 Sec	2017-03-06 12:07:43
Motown	155471[995]	platidpxn04	default	default		Higher_writes_gpfsfpo	2	42.00 MB	1.00 Sec	2017-03-06 12:07:42
Motown	155471[998]	platidpxn04	default	default		Higher_writes_gpfsfpo	2	42.00 MB	1.00 Sec	2017-03-06 12:07:42
Motown	155471[995]	platidpxn04	default	default		Higher_writes_gpfsfpo	2	42.00 MB	1.00 Sec	2017-03-06 12:07:42

Command: `dd bs=1MB count=998 if=/platgpfs/gpfsfpo/user/ladams/155471/998/155471.998.ddout of=/dev/null`
File: `/platgpfs/gpfsfpo/user/ladams/155471/998/155471.998.ddout`

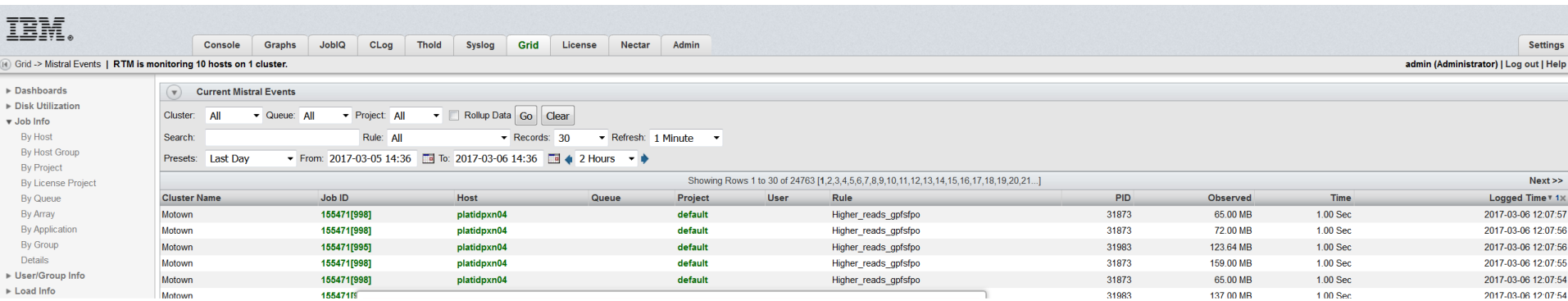
Showing Rows 1 to 30 of 24763 [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21...]

Table showing:

- Cluster name
- Job ID
- Host Queue
- Project
- User
- Mistral I/O rule
- PID
- I/O observed (MB)
- Time frame
- Log time



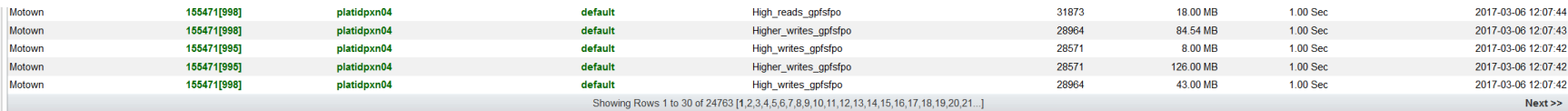
Mistral RTM integration: Who is using the storage?



The screenshot shows the Mistral RTM interface with the 'Grid' tab selected. The main content area displays a table titled 'Current Mistral Events'. The table has columns for Cluster Name, Job ID, Host, Queue, Project, User, Rule, PID, Observed, Time, and Logged Time. The data shows several jobs running on the 'Motown' cluster, all using the 'plattidpxn04' host and the 'default' project. The rules are 'Higher_reads_gpfsfpo' and 'Higher_writes_gpfsfpo'. The observed storage usage ranges from 65.00 MB to 137.00 MB.

Cluster Name	Job ID	Host	Queue	Project	User	Rule	PID	Observed	Time	Logged Time
Motown	155471[998]	plattidpxn04		default		Higher_reads_gpfsfpo	31873	65.00 MB	1.00 Sec	2017-03-06 12:07:57
Motown	155471[998]	plattidpxn04		default		Higher_reads_gpfsfpo	31873	72.00 MB	1.00 Sec	2017-03-06 12:07:56
Motown	155471[995]	plattidpxn04		default		Higher_reads_gpfsfpo	31983	123.64 MB	1.00 Sec	2017-03-06 12:07:56
Motown	155471[998]	plattidpxn04		default		Higher_reads_gpfsfpo	31873	159.00 MB	1.00 Sec	2017-03-06 12:07:55
Motown	155471[998]	plattidpxn04		default		Higher_reads_gpfsfpo	31873	65.00 MB	1.00 Sec	2017-03-06 12:07:54
Motown	155471[998]	plattidpxn04		default		Higher_reads_gpfsfpo	31983	137.00 MB	1.00 Sec	2017-03-06 12:07:54

After 12 months
(Almost) No more rogue jobs!



The screenshot shows a table of mistral events with columns for Cluster Name, Job ID, Host, Queue, Project, User, Rule, PID, Observed, Time, and Logged Time. The data shows several jobs running on the 'Motown' cluster, all using the 'plattidpxn04' host and the 'default' project. The rules are 'High_reads_gpfsfpo' and 'High_writes_gpfsfpo'. The observed storage usage ranges from 8.00 MB to 126.00 MB.

Cluster Name	Job ID	Host	Queue	Project	User	Rule	PID	Observed	Time	Logged Time
Motown	155471[998]	plattidpxn04		default		High_reads_gpfsfpo	31873	18.00 MB	1.00 Sec	2017-03-06 12:07:44
Motown	155471[998]	plattidpxn04		default		High_writes_gpfsfpo	28964	84.54 MB	1.00 Sec	2017-03-06 12:07:43
Motown	155471[995]	plattidpxn04		default		High_writes_gpfsfpo	28571	8.00 MB	1.00 Sec	2017-03-06 12:07:42
Motown	155471[995]	plattidpxn04		default		High_writes_gpfsfpo	28571	126.00 MB	1.00 Sec	2017-03-06 12:07:42
Motown	155471[998]	plattidpxn04		default		High_writes_gpfsfpo	28964	43.00 MB	1.00 Sec	2017-03-06 12:07:42



Phase 2: Performance

Catching rogue jobs before they are run

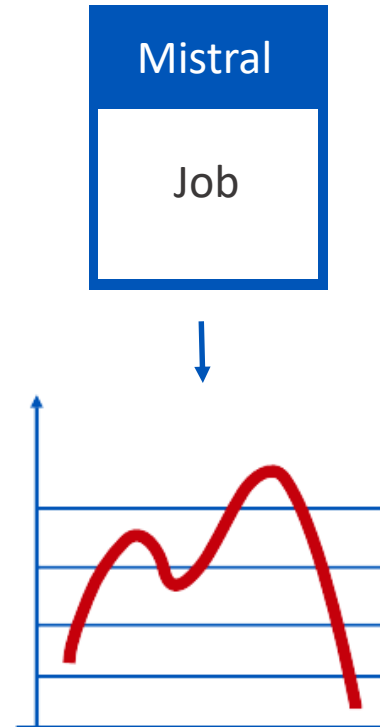
R&D jobs often seg fault and this exposes Mistral.
How to protect the storage without changing the R&D environment?

Solution

1. Guess which job is causing the problem the old way
2. Check problem jobs by profiling I/O in detail

Set up a special LSF queue with a detailed Mistral I/O profiling contract.

Make every user an I/O expert



Phase 2: I/O Profiling-as-a-Service

Introduce methodologies for finding and fixing I/O issues

Step 1: Qualify new jobs on test queue

Step 2: Run at scale

Step 3: Return to profiling queue if there is a problem

Mistral: IT and QA raise I/O tickets

Breeze: R&D to fix the problem.



Summary

Profile profile profile.

Protect your storage by monitoring the whole cluster for rogue jobs.
Prevent bad I/O patterns by running health checks on new work flows.

Optimise applications for shared storage and optimise the compute and storage infrastructure for the jobs.



Thank you

Please get in touch for more information.

Dr Rosemary Francis
CEO and Co-founder

Ellexus Ltd

St John's Innovation Centre, Cowley Road, Cambridge CB4 0WS, UK
info@ellexus.com
01223 123456



Mistral: wide-scale I/O profiling and load balancing for shared storage
www.ellexus.com