MANCHESTER 1824

Improving Sonar Rates: UK Experience

Alessandra Forti

Atlas S&C Week

6th April 2011

February Sonar Tests

ATLAS Sonar Tests: Large Files, All tests since 2011-01-01

Destination Site

Source Site	RAL-LCG2_DATADISK	FZK-LCG2_DATADISK	INFN-T1_DATADISK	PIC_DATADISK	NDGF-T1_DATADISK	TAIWAN-LCG2_DATADISK	SARA-MATRIX_DATADISK	TRIUMF-LCG2_DATADISK	IN2P3-CC_DATADISK	BNL-OSG2_DATADISK	Average
UKI-SCOTGRID-GLASGOW_DATADISK	14.07	5.18	7.68	3.01	10.37	0.90	7.88	1.91	8.87	1.95	6.18
UKI-NORTHGRID-MAN-HEP_DATADISK	3.06	3.48	3.48	2.00	0.59	0.67	4.13	1.78	6.30	1.20	2.67
UKI-NORTHGRID-LANCS-HEP_DATADISK	5.87	2.78	3.20	2.51	1.99	0.57	5.36	1.25	4.81	0.96	2.93
UKI-NORTHGRID-SHEF-HEP_DATADISK	9.77	6.48	8.03	5.25	1.81	1.56	9.09	2.76	15.88	6.16	6.68
UKI-NORTHGRID-LIV-HEP_DATADISK	8.26	5.64	7.80	5.40	4.13	2.16	9.08	2.79	14.10	5.27	6.46
UKI-SOUTHGRID-CAM-HEP_DATADISK	6.99	3.93	5.78	3.53	2.82	0.79	8.50	1.82	8.92	2.96	4.60
UKI-SOUTHGRID-OX-HEP_DATADISK	10.44	6.18	7.18	4.90	5.05	1.13	9.80	2.30	11.40	4.32	6.27
UKI-SOUTHGRID-RALPP_DATADISK	19.64	5.43	6.75	6.54	1.61	1.88	10.62	2.96	18.84	5.02	7.93
UKI-LT2-QMUL_DATADISK	14.24	3.31	12.75	11.73	14.90	4.36	31.18	10.47	21.41	14.47	13.88
Average	10.26	4.71	6.96	4.99	4.81	1.56	10.63	3.11	12.28	4.70	6.40

	UKI-SCOTGRID-GLASGOW_DATADISK	UKI-NORTHGRID-MAN-HEP_DATADISK	UKI-NORTHGRID-LANCS-HEP_DATADISK	UKI-NORTHGRID-SHEF-HEP_DATADISK	UKI-NORTHGRID-LIV-HEP_DATADISK	UKI-SOUTHGRID-CAM-HEP_DATADISK	UKI-SOUTHGRID-OX-HEP_DATADISK	UKI-SOUTHGRID-RALPP_DATADISK	UKI-LT2-QMUL_DATADISK	Average
RAL-LCG2_DATADISK	3.44	40.21	18.12	9.00	5.00	13.51	10.87	32.72	9.85	15.86
FZK-LCG2_DATADISK	5.76	9.07	11.23	4.57	11.94	12.23	10.22	4.18	5.37	8.28
INFN-T1_DATADISK	1.24	5.47	17.15	3.76	2.15	3.52	3.84	0.61	3.76	4.61
PIC_DATADISK	1.07	1.41	2.12	1.04	1.14	1.77	1.54	1.34	1.95	1.49
NDGF-T1_DATADISK	3.86	21.58	26.70	4.54	1.65	4.05	6.73	1.71	17.48	9.81
TAIWAN-LCG2_DATADISK	1.50	3.62	3.98	1.02	1.76	3.39	1.64	0.56	3.99	2.39
SARA-MATRIX_DATADISK	0.63	20.47	38.96	1.31	0.98	7.97	7.38	0.97	19.53	10.91
TRIUMF-LCG2_DATADISK	1.11	3.70	5.53	1.09	1.42	3.22	1.43	0.64	2.08	2.25
IN2P3-CC_DATADISK	2.13	24.20	28.31	3.04	2.13	8.29	6.02	5.31	11.37	10.09
BNL-OSG2_DATADISK	0.94	9.60	14.68	1.96	0.88	1.59	3.70		5.40	4.84
Average	2.17	13.93	16.68	3.13	2.90	5.96	5.34	5.34	8.08	7.05

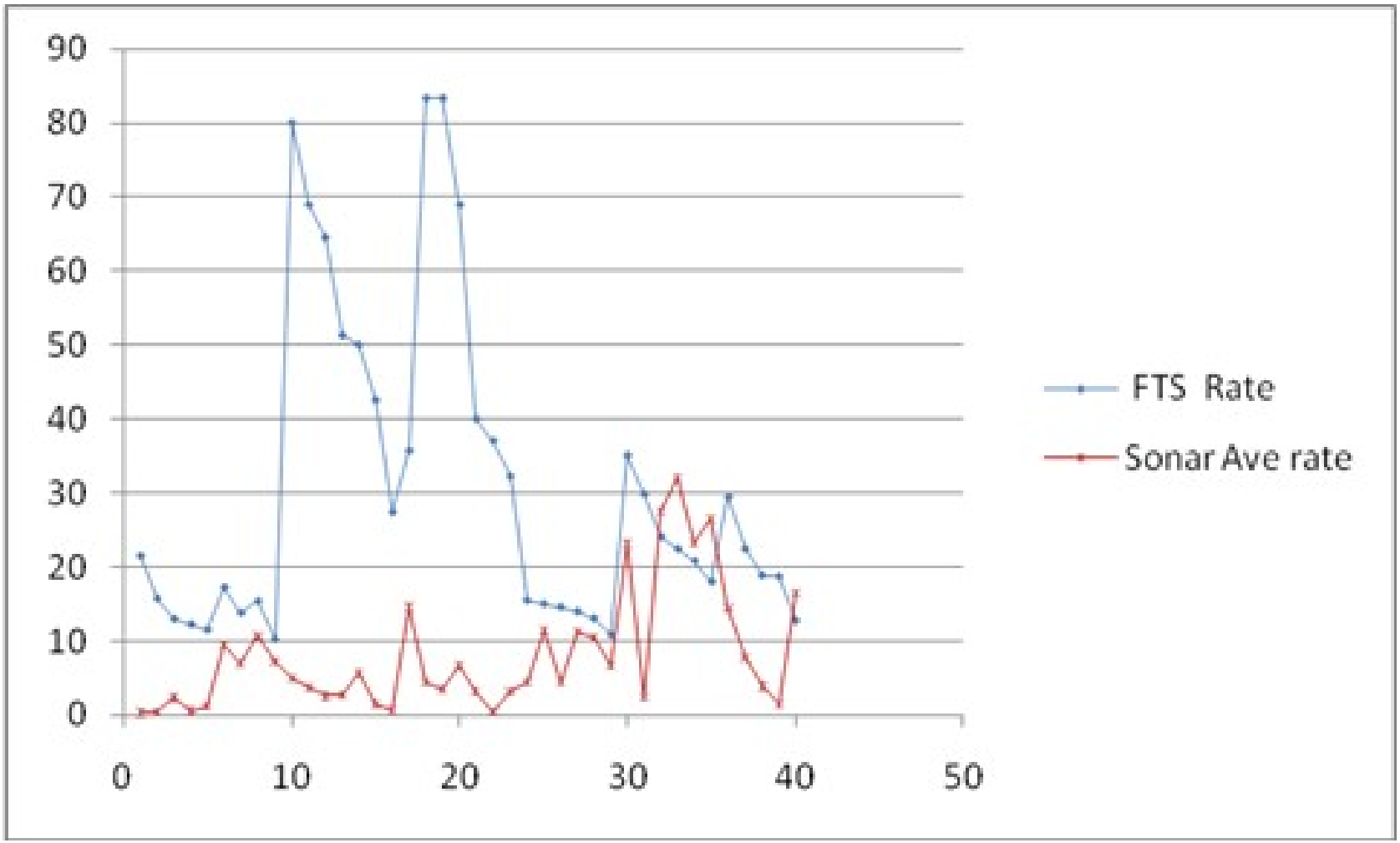
MANCHESTER 1824

Some Questions

- Do we have the network capability required?
- Does the network deliver the capability that is on paper?
- Does the monitoring report what we want?
- Do site admins see what the user see?

MANCHESTER 1824

FTS vs Sonar



Network capability is there

Network capabilities

SITE	BRUNEL	IC	QMUL	RHUL	UCL	LANCS	LIV	MAN	SHEF
T2 (Gbs)	1	10	1	0.4	1	2	1	10	1
JA (Gbs)	3X1	2X10	3X1	2X1	2X10	1X10	1X1	1X10	1X1

JA=JANET backbone
* 40 Gbit/s in the North
* 100 Gbits/s in the South

SITE	DURHAM	ECDF	GLASGOW	BHAM	BRIS	CAM	EFDA JET	OX	RALPP
T2 (Gbs)	1	1	6	1	1	1	0.1	1	20
JA (Gbs)	1X1	1X10	2X10	2X1	2X1	1X10	2X1	2X10	2X20

MANCHESTER 1824

Manchester, Glasgow and QMUL

- Manchester and Glasgow
  - Big networking capabilities
  - Big asymmetry in opposite directions
- QMUL
  - Smaller internal network
  - Better rates in sonar tests
- North – South divide?
  - The difference is between 40-100Gbs
    - Janet is big enough
  - Traceroute confirmed similar outside routes

MANCHESTER 1824

Tuning TCP Buffers

- Most sites had default DPM tuning borrowed from CERN castor
  - Some parameters too small for Gb size networks according to the experts online
  - All sites discussed this and tried to better tune their data servers
    - http://fasterdata.es.net/fasterdata/host-tuning/linux
- Some small improvement but problems persisted
  - The real bottleneck were still there

MANCHESTER 1824

Simplified Transfers

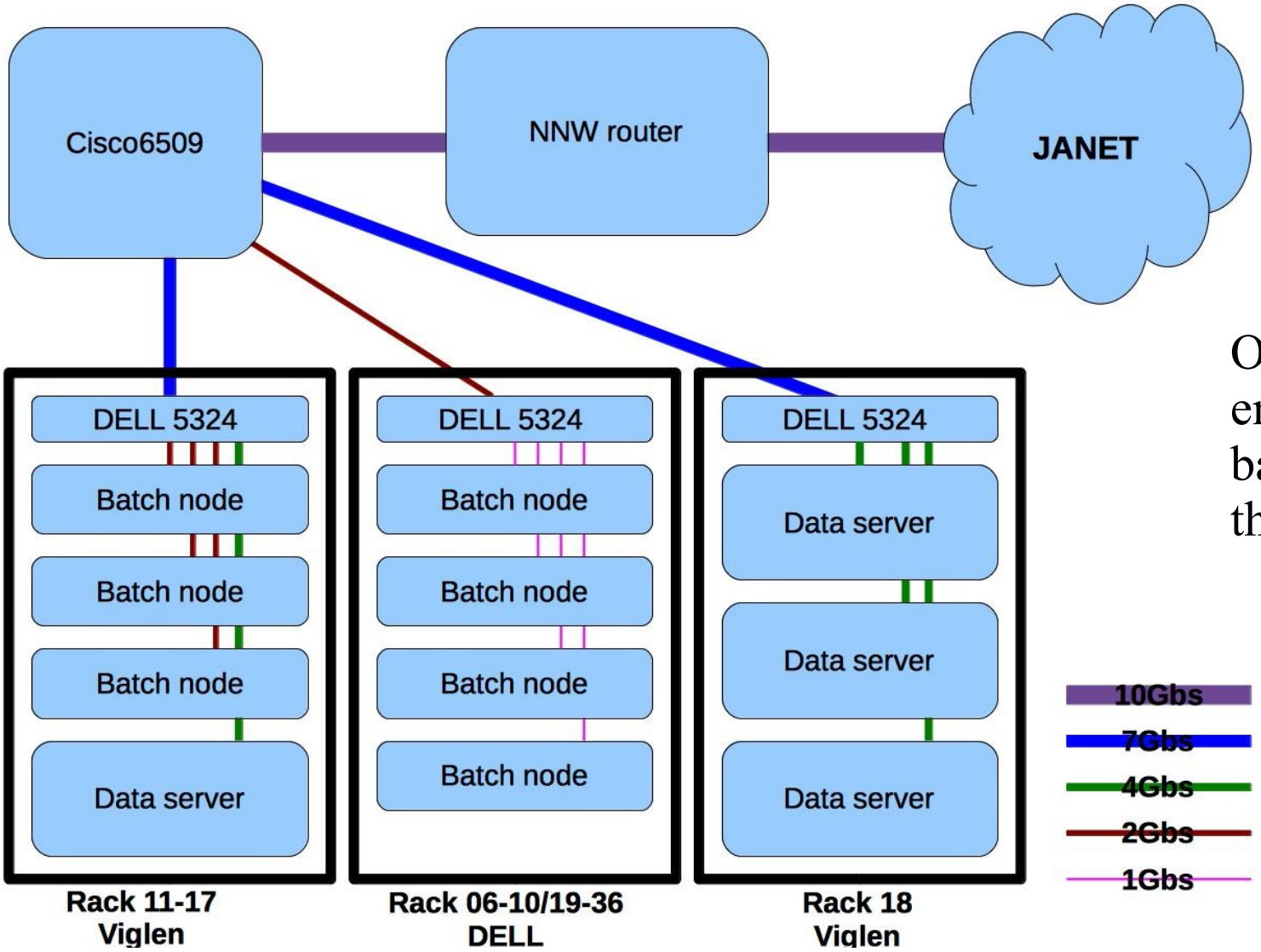
- Eliminate FTS overhead
  - (QMUL,MAN,GLA) ⊗ (BNL,SARA,RAL)
  - Only gridftp single stream unoptimised 2GB files
  - More agile than using the FTS framework
    - BUT for big files results only marginally better than sonar
- FTS looses out on smaller files when the overhead becomes notable
  - 80% of atlas transfers are files <100MB
    - But we decided to concentrate on the 2GB which scored worst.

MANCHESTER 1824

Iperf&tcptrace&ifstat

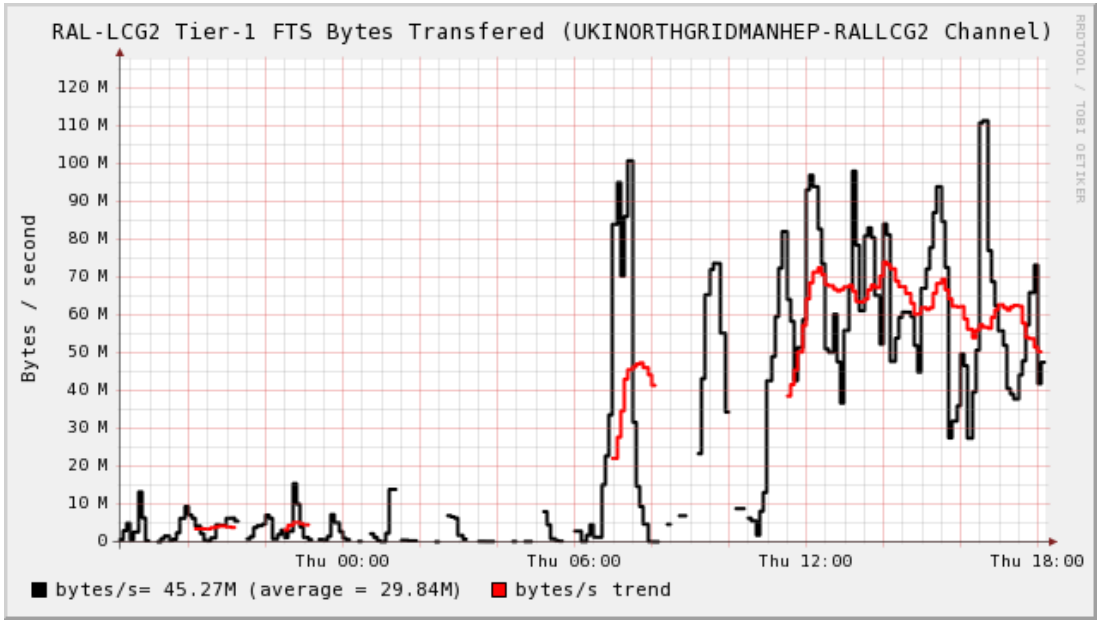
- A series of iperf tests were carried out between the three sites and RAL
  - Helped to pinpoint Glasgow problem in a campus firewall.
    - Removal of the firewall has improved incoming traffic for Glasgow T2
      - BNL-GLA : 1952 KB/sec avg BEFORE
      - BNL-GLA : 19828.68 KB/sec avg AFTER
  - Didn't help Manchester
    - tcpdump/tctrace that confirmed the problem for outgoing traffic but not the reason
    - ifstat was seeing 100MBs outgoing traffic on the bonded interfaces

Looking locally



On paper we have enough local bandwith. Is it the NNW router?

Other Monitoring

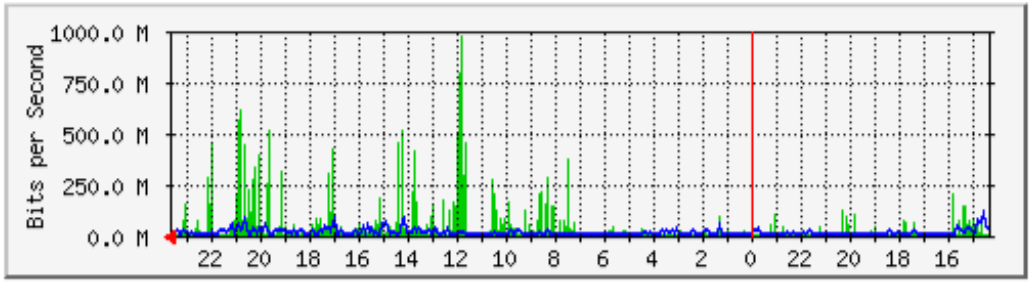


se09 -- nsw151

System: nsw151 in
Maintainer:
Description: se09
ifType: ethernetCsmacd (6)
ifName: ch7
Max Speed: 125.0 MBytes/s

The statistics were last updated **Tuesday, 5 April 2011 at 23:44**, at which time '**nsw151**' had been up for **40 days, 21:16:01**.

`Daily' Graph (5 Minute Average)



	Max	Average	Current
In	971.9 Mb/s (24.3%)	50.9 Mb/s (1.3%)	1432.0 b/s (0.0%)
Out	114.9 Mb/s (2.9%)	10.6 Mb/s (0.3%)	204.2 kb/s (0.0%)

Aggregate traffic is not comparable with single transfers. Agreed!
But sonar tests were consistently bad and consistently bad transfers should make consistently bad aggregate.

MANCHESTER 1824

Plugging a machine

desperate times desperate measures

- Assuming it local between machines and NNW router
  - Plugged a machine with Gb ethernet in each router
  - Got good rates also from the racks
  - To make sure it was not a disk servers problem we got a long cable and plug it into the cisco: it was ok!
  - Only thing that remained was the bonding.
    - After disconnecting 3 our of 4 links transfers shot up
      - MAN-BNL : 207.41 KB/sec avg BEFORE
      - MAN-BNL : 33616.61 KB/sec avg AFTER

MANCHESTER 1824

Glasgow sonar

Show 25 entries — Search all columns: glasgow

					SMALL FILES			MEDIUM FILES			LARGE FILES		
Prio	Source	SCloud	Destination	DCloud	MB/s	MB	#Ev	MB/s	MB	#Ev	MB/s	GB	#Ev
7	FZK-LCG2	DE - T1	UKI-SCOTGRID-GLASGOW	UK - T2	0.5+-0.24	20.0+-0.0	10	2.46+-1.96	484.31+-299.61	52	3.97+-3.43	1.83+-0.67	37
7	INFN-T1	IT - T1	UKI-SCOTGRID-GLASGOW	UK - T2	0.56+-0.18	20.0+-0.0	10	1.65+-0.56	276.01+-129.78	19	6.18+-7.21	1.53+-0.46	25
7	NDGF-T1	NG - T1	UKI-SCOTGRID-GLASGOW	UK - T2	0.37+-0.22	18.28+-5.72	11	3.36+-2.65	430.52+-284.16	22	3.97+-3.44	1.75+-0.51	19
7	UKI-SCOTGRID-GLASGOW	UK - T2	FZK-LCG2	DE - T1	0.58+-0.57	20.0+-0.0	10	1.58+-1.15	200.0+-0.0	10	3.1+-1.68	2.0+-0.0	10
7	UKI-SCOTGRID-GLASGOW	UK - T2	INFN-T1	IT - T1	0.59+-0.26	20.0+-0.0	10	2.95+-1.12	200.0+-0.0	10	10.97+-11.57	2.0+-0.0	10
7	UKI-SCOTGRID-GLASGOW	UK - T2	NDGF-T1	NG - T1	1.39+-0.15	20.0+-0.0	10	4.16+-3.64	200.0+-0.0	10	13.68+-16.55	2.0+-0.0	15
7	UKI-SCOTGRID-GLASGOW	UK - T2	TW-FTT	TW - T1	0.31+-0.14	20.0+-0.0	10	1.38+-0.52	200.0+-0.0	10	3.29+-1.91	2.0+-0.0	10
7	UKI-SCOTGRID-GLASGOW	UK - T2	PIC	ES - T1	0.66+-0.11	20.0+-0.0	10	2.88+-2.43	200.0+-0.0	10	14.09+-11.02	2.0+-0.0	10
7	UKI-SCOTGRID-GLASGOW	UK - T2	SARA-MATRIX	NL - T1	1.04+-0.35	20.0+-0.0	15	3.93+-2.23	200.0+-0.0	15	20.42+-19.67	2.0+-0.0	15
7	UKI-SCOTGRID-GLASGOW	UK - T2	TAIWAN-LCG2	TW - T1	0.16+-0.01	20.0+-0.0	10	0.86+-0.45	200.0+-0.0	10	1.86+-1.17	2.0+-0.0	10
7	UKI-SCOTGRID-GLASGOW	UK - T2	TRIUMF-LCG2	CA - T1	0.5+-0.04	20.0+-0.0	10	0.82+-0.31	200.0+-0.0	10	1.19+-0.11	2.0+-0.0	10
7	UKI-SCOTGRID-GLASGOW	UK - T2	IN2P3-CC	FR - T1	0.71+-0.35	20.0+-0.0	10	3.08+-1.3	200.0+-0.0	10	6.74+-3.8	2.0+-0.0	10
7	UKI-SCOTGRID-GLASGOW	UK - T2	NIKHEF-ELPROD	NL - T1	0.79+-0.17	20.0+-0.0	15	3.75+-2.22	200.0+-0.0	15	25.06+-23.36	2.0+-0.0	15
7	UKI-SCOTGRID-GLASGOW	UK - T2	BNL-OSG2	US - T1	0.4+-0.23	20.0+-0.0	15	2.15+-0.69	200.0+-0.0	15	8.31+-8.52	2.0+-0.0	15
7	UKI-SCOTGRID-GLASGOW	UK - T2	CERN-PROD	CERN - T0	0.61+-0.26	20.0+-0.0	15	2.9+-1.3	200.0+-0.0	15	15.02+-14.82	2.0+-0.0	14
7	TW-FTT	TW - T1	UKI-SCOTGRID-GLASGOW	UK - T2	0.21+-0.08	25.35+-2.97	55	0.53+-0.6	694.99+-326.79	38	1.41+-1.1	1.86+-0.23	15
7	PIC	ES - T1	UKI-SCOTGRID-GLASGOW	UK - T2	0.38+-0.23	20.0+-0.0	10	0.99+-0.48	593.66+-272.53	37	1.33+-0.53	2.04+-0.21	16
7	SARA-MATRIX	NL - T1	UKI-SCOTGRID-GLASGOW	UK - T2	0.69+-0.21	20.0+-0.0	10	5.25+-4.9	611.87+-321.72	36	6.7+-6.04	1.82+-0.57	20
7	TAIWAN-LCG2	TW - T1	UKI-SCOTGRID-GLASGOW	UK - T2	0.23+-0.15	20.0+-0.0	10	0.75+-0.53	200.0+-0.0	10	2.15+-0.86	2.0+-0.0	10
7	TRIUMF-LCG2	CA - T1	UKI-SCOTGRID-GLASGOW	UK - T2	0.38+-0.15	20.0+-0.0	10	1.1+-1.09	614.16+-267.61	45	0.77+-0.33	1.72+-0.3	27
7	IN2P3-CC	FR - T1	UKI-SCOTGRID-GLASGOW	UK - T2	0.49+-0.3	20.0+-0.0	10	2.87+-1.5	428.89+-265.57	29	3.87+-3.06	1.83+-0.25	29
7	NIKHEF-ELPROD	NL - T1	UKI-SCOTGRID-GLASGOW	UK - T2	0.48+-0.08	20.0+-0.0	10	2.8+-1.67	200.13+-0.44	11	9.22+-7.27	2.0+-0.0	10
7	BNL-OSG2	US - T1	UKI-SCOTGRID-GLASGOW	UK - T2	0.36+-0.2	20.0+-0.0	10	1.76+-2.19	588.9+-368.37	29	2.07+-2.46	1.64+-0.6	30
7	CERN-PROD	CERN - T0	UKI-SCOTGRID-GLASGOW	UK - T2	0.5+-0.17	20.0+-0.0	10	1.52+-0.79	200.0+-0.0	10	9.94+-7.59	2.0+-0.0	10
7	Filter source	Source cloud	Filter dest	Dest cloud									

Showing 1 to 24 of 24 entries (filtered from 6,321 total entries) — First Previous 1 Next Last

Manchester sonar

Only DATADISK to DATADISK transfers are shown (Period: 2011-03-11 - 2011-04-01)

Show 50 entries — Search all columns: man-hep

					SMALL FILES			MEDIUM FILES			LARGE FILES		
Prio	Source	SCloud	Destination	DCloud	MB/s	MB	#Ev	MB/s	MB	#Ev	MB/s	GB	#Ev
7	FZK-LCG2	DE - T1	UKI-NORTHGRID-MAN-HEP	UK - T2	0.91+-0.37	20.0+-0.0	10	3.04+-2.41	200.0+-0.0	10	6.82+-2.14	2.0+-0.0	10
7	INFN-T1	IT - T1	UKI-NORTHGRID-MAN-HEP	UK - T2	0.94+-0.08	20.0+-0.0	10	6.51+-1.49	200.0+-0.0	10	29.66+-17.31	2.0+-0.0	10
7	NDGF-T1	NG - T1	UKI-NORTHGRID-MAN-HEP	UK - T2	1.42+-0.15	20.0+-0.0	10	10.37+-2.04	200.0+-0.0	10	18.88+-12.74	2.0+-0.0	10
7	TW-FTT	TW - T1	UKI-NORTHGRID-MAN-HEP	UK - T2	0.37+-0.03	20.0+-0.0	10	1.23+-0.46	200.0+-0.0	10	1.88+-0.8	2.0+-0.0	10
7	PIC	ES - T1	UKI-NORTHGRID-MAN-HEP	UK - T2	0.88+-0.24	20.0+-0.0	10	1.97+-0.67	200.0+-0.0	10	1.59+-0.56	2.0+-0.0	10
7	SARA-MATRIX	NL - T1	UKI-NORTHGRID-MAN-HEP	UK - T2	1.16+-0.13	20.0+-0.0	10	9.05+-2.65	200.0+-0.0	10	32.01+-19.62	2.0+-0.0	10
7	TAIWAN-LCG2	TW - T1	UKI-NORTHGRID-MAN-HEP	UK - T2	0.37+-0.06	20.0+-0.0	10	0.9+-0.46	200.0+-0.0	10	2.51+-1.0	2.0+-0.0	10
7	TRIUMF-LCG2	CA - T1	UKI-NORTHGRID-MAN-HEP	UK - T2	0.61+-0.11	20.0+-0.0	10	4.71+-1.31	200.0+-0.0	10	12.79+-8.83	2.0+-0.0	10
7	IN2P3-CC	FR - T1	UKI-NORTHGRID-MAN-HEP	UK - T2	1.05+-0.15	20.0+-0.0	10	7.13+-2.85	200.0+-0.0	10	17.91+-16.1	2.0+-0.0	10
7	NIKHEF-ELPROD	NL - T1	UKI-NORTHGRID-MAN-HEP	UK - T2	1.19+-0.07	20.0+-0.0	10	8.64+-1.84	200.0+-0.0	10	40.26+-11.48	2.0+-0.0	10
7	BNL-OSG2	US - T1	UKI-NORTHGRID-MAN-HEP	UK - T2	0.5+-0.2	20.0+-0.0	10	4.35+-2.04	200.0+-0.0	10	17.46+-8.8	2.0+-0.0	10
7	CERN-PROD	CERN - T0	UKI-NORTHGRID-MAN-HEP	UK - T2	0.88+-0.32	20.0+-0.0	10	6.82+-1.95	200.0+-0.0	10	22.94+-12.86	2.0+-0.0	10
7	UKI-NORTHGRID-MAN-HEP	UK - T2	FZK-LCG2	DE - T1	0.53+-0.06	20.0+-0.0	10	0.99+-0.21	200.0+-0.0	10	0.91+-0.16	2.0+-0.0	10
7	UKI-NORTHGRID-MAN-HEP	UK - T2	INFN-T1	IT - T1	0.71+-0.05	20.0+-0.0	10	2.55+-1.26	200.0+-0.0	10	7.32+-9.14	2.0+-0.0	10
7	UKI-NORTHGRID-MAN-HEP	UK - T2	NDGF-T1	NG - T1	0.86+-0.64	20.0+-0.0	15	3.64+-4.51	200.0+-0.0	15	10.52+-17.0	2.0+-0.0	15
7	UKI-NORTHGRID-MAN-HEP	UK - T2	TW-FTT	TW - T1	0.07+-0.01	20.0+-0.0	10	0.08+-0.02	200.0+-0.0	10	0.53+-0.04	2.0+-0.0	5
7	UKI-NORTHGRID-MAN-HEP	UK - T2	PIC	ES - T1	0.57+-0.1	20.0+-0.0	10	1.65+-1.34	200.0+-0.0	10	4.03+-5.02	2.0+-0.0	10
7	UKI-NORTHGRID-MAN-HEP	UK - T2	SARA-MATRIX	NL - T1	1.24+-0.08	20.0+-0.0	15	5.45+-2.3	200.0+-0.0	15	15.04+-14.71	2.0+-0.0	15
7	UKI-NORTHGRID-MAN-HEP	UK - T2	TAIWAN-LCG2	TW - T1	0.06+-0.01	20.0+-0.0	10	0.12+-0.08	200.0+-0.0	10	0.23+-0.18	2.0+-0.0	12
7	UKI-NORTHGRID-MAN-HEP	UK - T2	TRIUMF-LCG2	CA - T1	0.11+-0.02	20.0+-0.0	10	0.14+-0.02	200.0+-0.0	10	0.16+-0.01	2.0+-0.0	15
7	UKI-NORTHGRID-MAN-HEP	UK - T2	IN2P3-CC	FR - T1	0.64+-0.08	20.0+-0.0	10	1.22+-0.34	200.0+-0.0	10	1.27+-0.33	2.0+-0.0	10
7	UKI-NORTHGRID-MAN-HEP	UK - T2	NIKHEF-ELPROD	NL - T1	0.75+-0.14	20.0+-0.0	15	3.23+-2.68	200.0+-0.0	15	10.69+-11.8	2.0+-0.0	15
7	UKI-NORTHGRID-MAN-HEP	UK - T2	BNL-OSG2	US - T1	0.32+-0.14	20.0+-0.0	15	2.26+-2.57	200.0+-0.0	15	9.18+-10.35	2.0+-0.0	11
7	UKI-NORTHGRID-MAN-HEP	UK - T2	CERN-PROD	CERN - T0	0.63+-0.28	20.0+-0.0	15	3.1+-2.26	200.0+-0.0	15	8.66+-8.95	2.0+-0.0	15
7	Filter source	Source cloud	Filter dest	Dest cloud									

Showing 1 to 24 of 24 entries (filtered from 6,321 total entries) — First Previous 1 Next Last

MANCHESTER 1824

Next steps

- Manchester needs to solve the bonding problem
  - We tried to keep it simple.....
- Gridftp mini sonar tests will be extended to other T1s with which we still have bad rates
  - PIC and TRIUMF
- In parallel FTS tuning for the now 4 T2D candidates
  - Channels have been already setup
  - Other tuning such as increasing the number of threads will be applied
  - Monitoring of sonar and FTS instant rates

MANCHESTER 1824

Conclusions

- Do we have the network capability required?
  - Some sites definitely do.
- Does the network deliver capability that is on paper?
  - Not yet. There are clearly problems that need to be solved as Manchester and Glasgow stories show.
- Does the monitoring report what we want?
  - Not always TCP packet loss and retries are not detected by the monitoring we look at but they contribute heavily to low throughput.
- Do site admins see what the user see?
  - No. Until the sonar tests we didn't think we had a problem. Such tests are important because they systematically give a user perspective.

MANCHESTER 1824

Acknowledgements

- GridPP storage group
  - Sam Skipsey and Brian Davis
- DANTE
  - Richard Hugh-Jones
- NNW
  - Michael Robson

MANCHESTER 1824

Questions?