

CCRC08-1 report

CCRC f2f, March 4 2008

Kors Bos, NIKHEF/CERN

Courtesy: many slides from ATLAS Jamboree talk by Simone Campana

Week 1:FDR-1

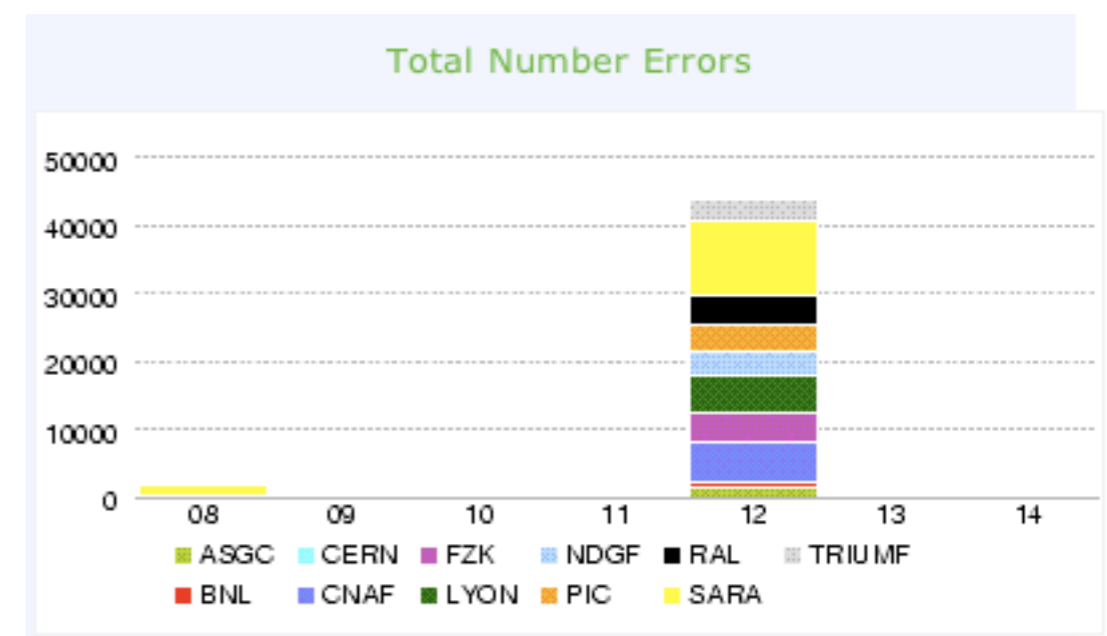
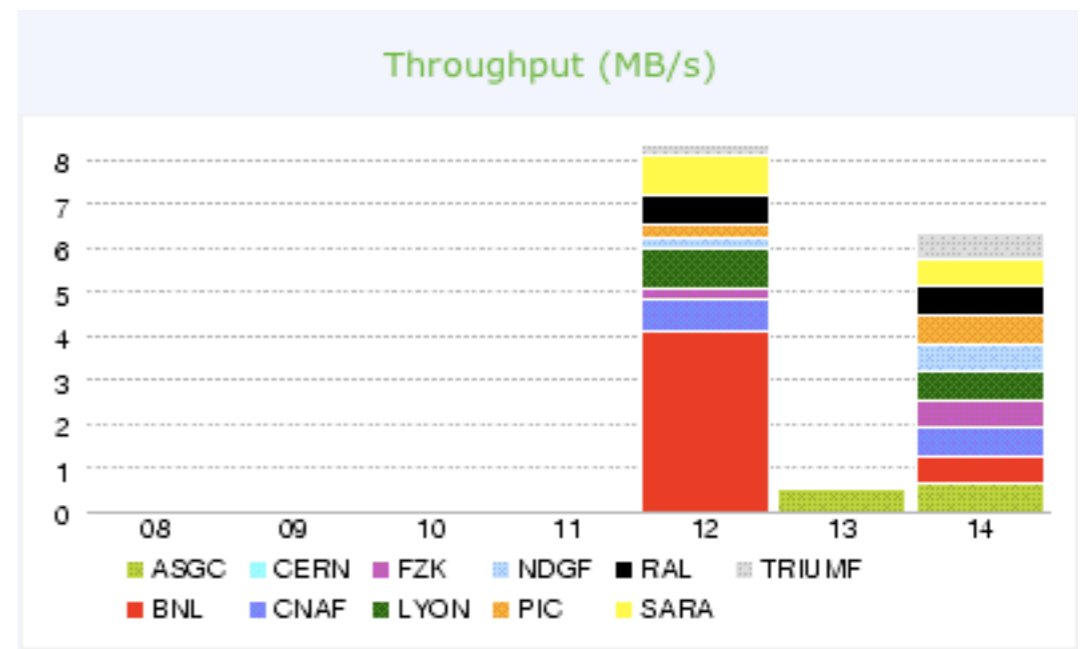
- Use simulated data with physics mix of 10^{31} Luminosity
- Run 10 hrs @200 HZ from SFO's
- Full T0 operation:
 - ✓ Calibration, calibration database may take 24 hrs
 - ✓ Data quality, sign off end of the next day
 - ✓ First pass reconstruction may take 24 hrs
 - ✓ If necessary re-calibration make take another 24 hrs
 - ✓ Subscribe data for transfers to T1&2's
- We had planned to concentrate on T0 operations first
- But it took longer than was expected
- Moreover generated data sample was smaller: no fakes

Week 1 what we also did

- Setup of ATLAS site services
 - ✓ Storage spaces definitions in TiersOfATLAS
 - ✓ usage of FTS delegation
- Testing of SRM2.2 endpoints
 - ✓ Space tokens, ACLs
 - ✓ BDII publication
 - ✓ FTS channels configurations
- Several configuration issues spotted and cured.

Week 2

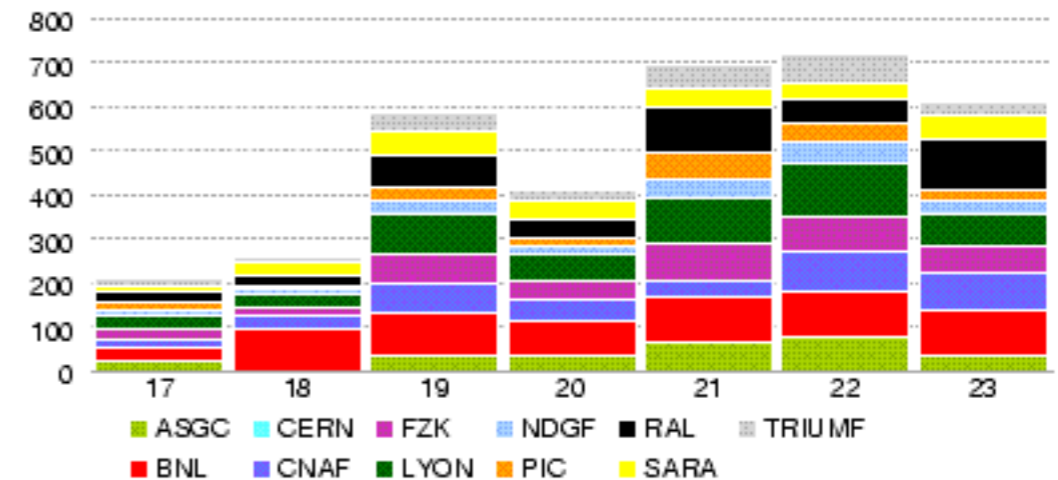
- FDR exports
 - ✓ 2 bulks of data: day 1+2 and day 3
- Very high failure in the first bulk
 - ✓ Certificate-to-pool mapping problem at T0
 - ➔ Triggering CASTOR disk-to-disk internal replication
 - ➔ Problem not on BNL (separate VOBOX)
 - ✓ Basically 100% success in second bulk



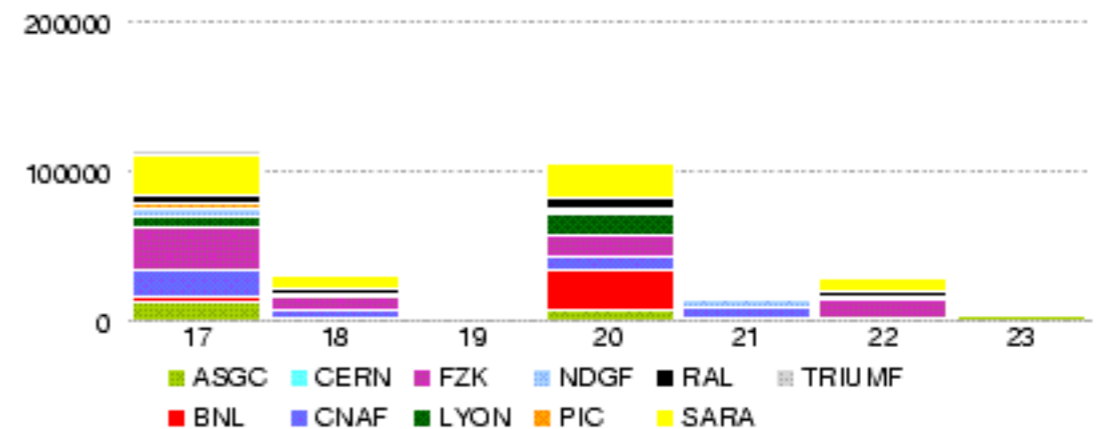
Week 3

- Changed to other data sample
 - ✓ FDR sample too small for TT
 - ✓ Used generated files: nominal size
 - ✓ Ramped up to nominal rates
 - ✓ Full Computing Model, MoU shares
- Target Rate: ~900 MB/s
 - ✓ Full steam for 24h
- Relatively good throughput achieved
 - ✓ Sustained 700 MB/s for 2 days
 - ✓ Peaks above 1.1GB/s for several hours

Throughput



Errors



Week 3 compared

- What happened to THIS?

- ✓ T0 exercise in September

- Much more precise in data transfers

- ✓ The T0 buffer is recycled after 1 day, now after 2.5 days

- ✓ If data did not make it in 1 day ... tant pis!

- ✓ In the past, datasets were cleaned daily from T0

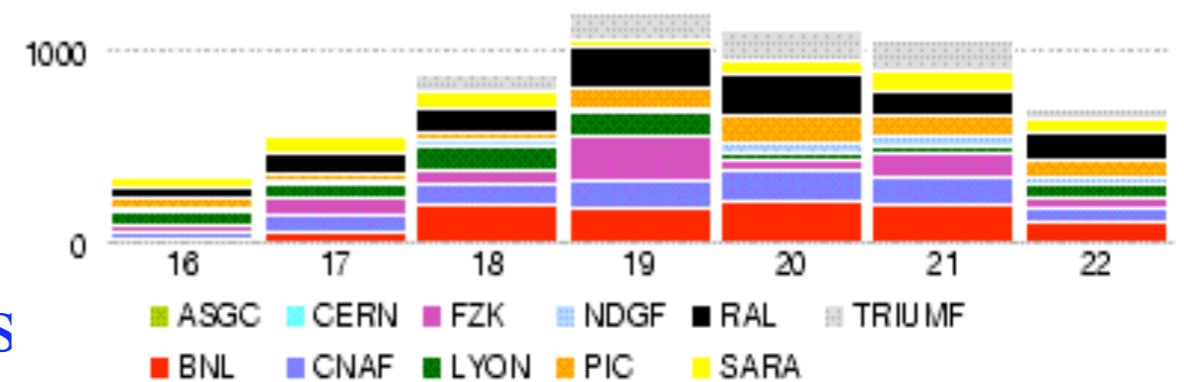
- ➔ now we control the cleaning

- ✓ In September we have been “oversubscribing” data : ESD everywhere

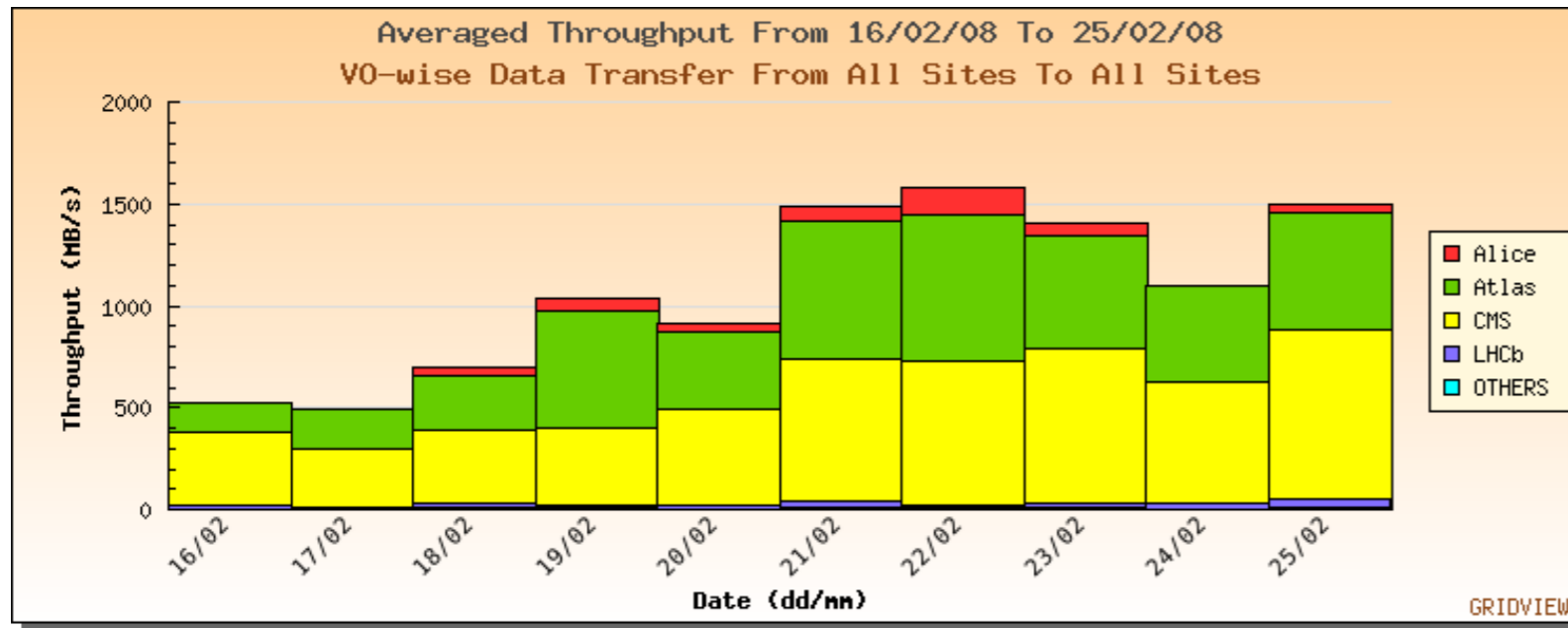
- ➔ Now we strictly followed the CM and the MoU ratio's

- Sites request more TT tests

- We want to do more FT tests

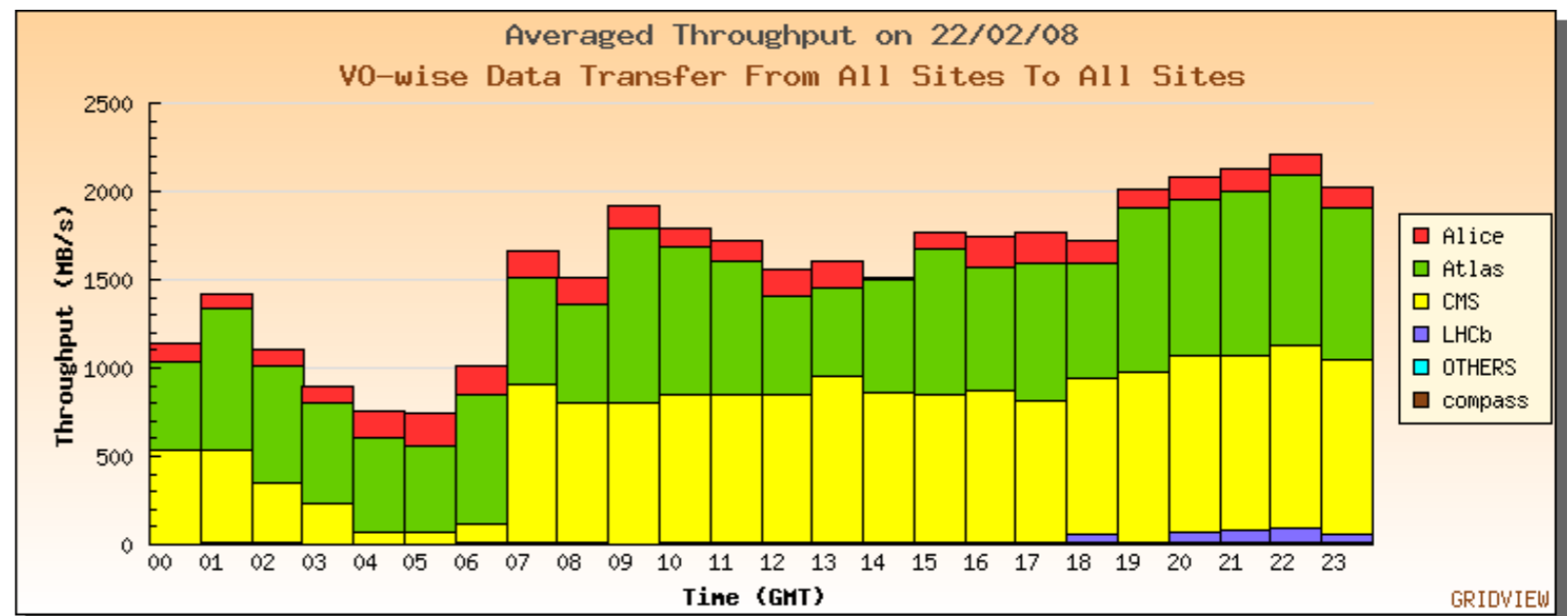


You'll never walk alone



Weekly
Throughput

2.1 GB/s out
of CERN



T1-T1 transfers

TRIU MF									
BNL		X	100	100	20		100		25
15	20	100		100					
FZK		X	X	100	30		100		40
50	30	100		100					
INFN		100	90	X	0		100		25
30	X	100		100					
IN2P3		X	100	100	X		100		35
20	30	100		100					
NDGF		100	100	100	0		X		40
50	30	100		100					
NIKHEF		0	0	0	0		0		X
0	0	0		0					
PIC		95	100	95	3		100		15
X	5	90		95					
RAL		100	100	100	10		0		30
10	10	100		100					
TAIWAN		100	100	100	25		0	40	70
80	X		100						
TRIU MF		100	100	30	20		100		30
30	X		100						

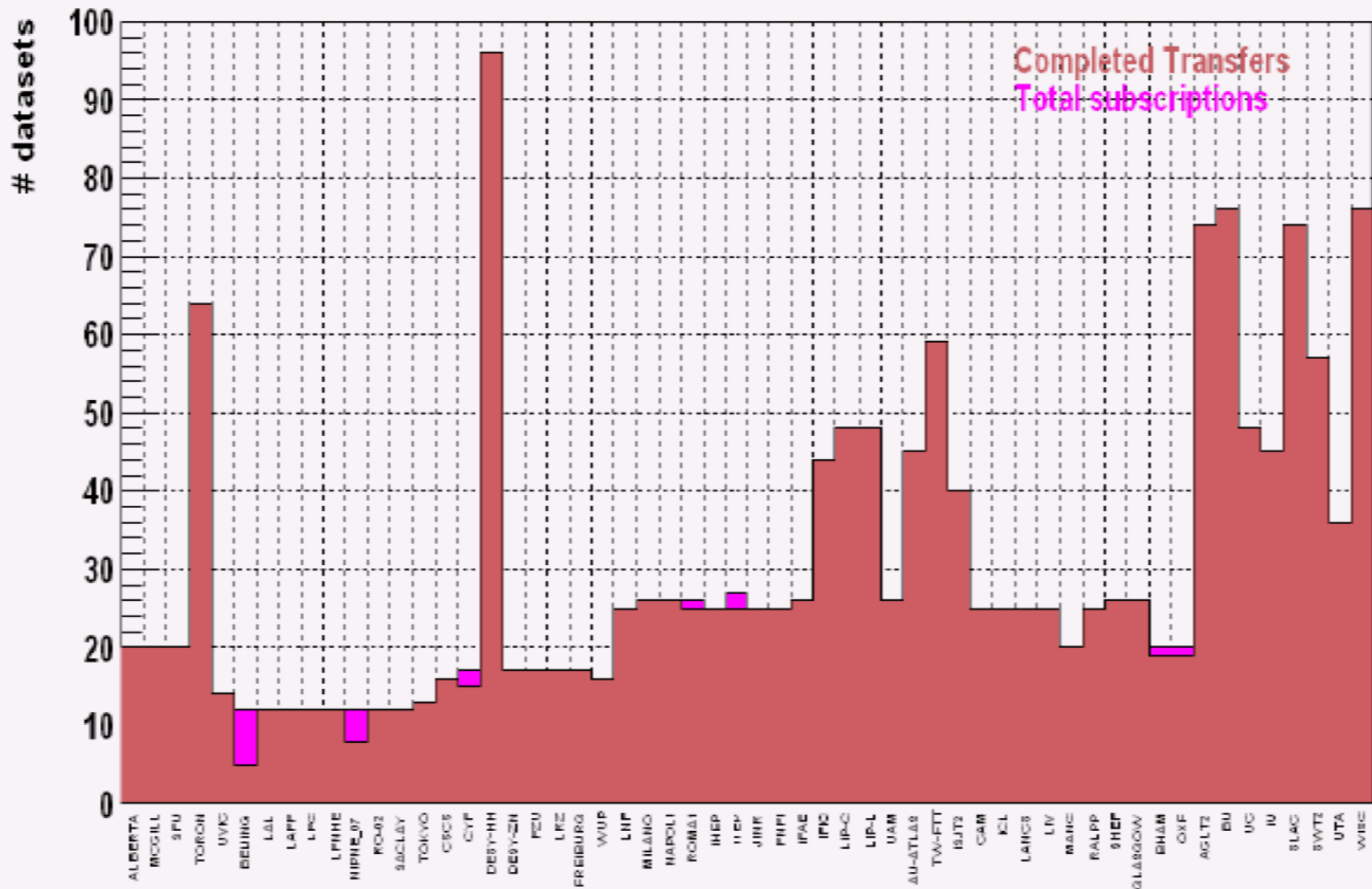
- Varying results, not all understood
- NIKHEF unable to export data (understood?)
- Several issues of FTS channels configuration
 - ✓ Many still being investigated (e.g. NDGF exporting via SRM1)
- Need to repeat the exercise (next weeks)

Week 4

- T1-T2 replication within cloud
 - ✓ Ongoing exercise also in production, see next slide
- Replication of M5 data to T1 tapes
 - ✓ Needed for re-processing exercise
- Reprocessing at T1's: Triumf, NDGF, BNL
- FDR-2 simulation production validation started
- T0→T1 transfers stopped on Thursday because of M6 preparation
 - ✓ Clear CCRC data from all sites
 - ✓ Set up and test T0-T2 channels for Muon calibration
 - ✓ Re-arrange CERN Castor pools

FDR Feb 2008. Data Transfer to Tier-2s

FDR AOD datasets replication within clouds (56 Tiers)



SRM2.2: general considerations

- SRM2.2 testing was (one of the) main goals of ATLAS CCRC08
 - ✓ Delivered “just in time”
 - ➔ Most of the sites delivered during last pre-FDR week
 - ✓ But, still, in time
- IMHO, it has been working very well
 - ✓ No major problems observed in all implementations
 - ➔ Various bugs fixed promptly.
 - ✓ Generally good storage setup at sites
 - ✓ Various issues in the interaction between different components (SRM-FTS-SRM) have been identified and overcome

SRM2.2 space tokens

- Space tokens at T1s
 - ✓ ATLASDATADISK and ATLASDATATAPE correctly configured in all T1s
 - ➔ ATLASMCDISK and ATLASMCTAPE begins being tested
 - ✓ Should start thinking about T1D1
 - ➔ ATLASDATADISKTAPE and ATLASMCDISKTAPE
 - ➔ It is clear how this will be implemented at every T1?
 - ✓ Put more pressure on group analysis at T1
 - ➔ ATLASGRM<group> need to be setup for a (very) few groups
- Space tokens at T2s
 - ✓ Many T2s are still on SRMv1 (50%)
 - ✓ SRMv2 T2s set up ATLASDATADISK and ATLASMCDISK
 - ➔ Being tested right now.

Space Tokens and Disk Pools

- CCRC08 Storage Solution Working Group
 - ✓ Technical discussion about prioritized evolution of SRM
 - ➔ <http://indico.cern.ch/conferenceDisplay.py?confId=29191>
 - ✓ On top of the list every experiment put “protection of space tokens from generic user access, based on VOMS groups and roles”
 - ➔ CASTOR and dCache
 - ✓ This is true for both “READ” and “WRITE”
- Timescale of implementation is by CCRC09
 - ✓ How to protect in the meanwhile?
 - ➔ Write+Read: @T0, static DN(now Group)-to-pool mapping. @T1s?
 - ➔ Read: at T0, interesting “service class fallback” capability
 - Can protect from users “recall from tape” into critical pools.

Communication Channels

- ELOG: <https://prod-grid-logger.cern.ch/elog/CCRC'08+Logbook/>
 - ✓ Found very useful.
 - ➔ Need an ATLAS specific one also?
 - ✓ But not a ticketing system
- E-mail your favorite T1 (or T0) contact
 - ✓ Many times very effective. But your favorite contact might be skiing on the Alps
- GGUS is a ticketing system
 - ✓ But the latency is too high
- Need to define a “hotline”
 - ✓ At CERN: atlas-grid-alarm@cern.ch is seen by the 24 hour operator who may escalate
 - ➔ Very restricted list of people can post (need to revisit)
 - ✓ Asked Jamie to also discuss this with T1s

Metrics pre-FDR

- **Week 4: Jan 21 – 27**

- ✓ Setting up sites with SRMv2 production endpoints
- ✓ Testing central bulk data deletion at sites
- ✓ RDO file mixing into Byte Stream data
- ✓ T0 – T1 transfers, storage space can be re-cycled
- ✓ *Milestone 2: T0 and T1 space tokens tested*

- **Week 5: Jan 28 – Feb 3**

- ✓ Setting up sites with SRMv2 production endpoints
- ✓ Testing central bulk data deletion
- ✓ RDO file mixing into Byte Stream data
- ✓ T0 – T1 transfers, storage space can be re-cycled
- ✓ Upload Byte Stream data into the SFO's

- **Week 4: Jan 21 – 27**

- ✓ **Many ToA changes**
- ✓ **Not started yet**
- ✓ **Not started yet**
- ✓ **Wait for srmv2**
- ✓ ***Milestone 2: failed***

- **Week 5: Jan 28 – Feb 3**

- ✓ **Many more ToA changes**
- ✓ **Simple deletions**
- ✓ **started**
- ✓ **started**
- ✓ **Not started yet**

Ref. ATLAS Planning for FDR-1 and CCRC-1 vs.1 Jan.28
<http://www.nikhef.nl/~bosk/documents/d2008/ATLAS-Document-for-FDR-and-CCRC.pdf>

Metrics week 1&2

- **Week 6: Feb 4 – 10**

- ✓ **Streaming Data read from SFO's into CASTOR**

- ✓ **Full T0 operation:**

- Data stored onto tape
- Calibration
- First pass reconstruction
- Data subscriptions

T0 – T1 transfers, storage space re-cycled

- *Milestone 4: Data streaming from SFO's tested*

- *Milestone 5: Full T0 operation tested*

- **Week 7: Feb 11 – 17**

- ✓ **Without SFO's from now on**

- ✓ **Full T0 operation from streaming data on CASTOR**

- ✓ **Data subscriptions to Tier-1's and Tier-2's**

- ✓ **Test SRMv2.2 bring-online functionality in at 2 sites**

- ✓ **Central bulk data deletion on the last day**

- *Milestone 6: Data subscription to all Tiers and storage classes tested*

- **Week 6: Feb 4 – 10**

- ✓ **Done**

- ✓ **Full T0 operation: Done**

- ✓ **Done**

- ✓ **Done**

- ✓ **Done**

- ✓ **Done**

- ✓ **Done**

- *Milestone 4: Done*

- *Milestone 5: Done*

- **Week 7: Feb 11 – 17**

- ✓ **True**

- ✓ **Done**

- ✓ **Done**

- ✓ **Done**

- ✓ **Requested to leave the data**

- *Milestone 6: Done*

Ref. ATLAS Planning for FDR-1 and CCRC-1 vs.1 Jan.28

<http://www.nikhef.nl/~bosk/documents/d2008/ATLAS-Document-for-FDR-and-CCRC.pdf>

Metrics week 3&4

- **Week 8: Feb 18 – 24**

- ✓ Full T0 operation from streaming data on CASTOR
- ✓ Data subscription to all Tiers and storage classes
- ✓ Test SRMv2.2 bring-online functionality
- ✓ Group Analysis in at least two Tier-2's
- ✓ DPD production in at least two Tier-1's
- ✓ Central bulk data deletion on the last day

- *Milestone 7: staging of data from tape tested in 2 Tier-1's tested*

- *Milestone 8: DPD production in Tier-1 tested*

- *Milestone 9: Group analysis in Tier-2 tested*

- **Week 9: Feb 25 – 29**

- ✓ Full T0 operation from streaming data on CASTOR
- ✓ Data subscription to all Tiers and storage classes
- ✓ Re-processing in 2 Tier-1 sites
- ✓ Group Analysis in Tier-2's
- ✓ DPD production in Tier-1's

- *Milestone 10: Re-processing in two Tier-1's tested*

- **Week 8: Feb 18 – 24**

- ✓ Done. but with other data
- ✓ Dones
- ✓ Done
- ✓ Started
- ✓ Not done
- ✓ Done but done dayly

- *Milestone 7: done*

- *Milestone 8: not done*

- *Milestone 9: started*

- **Week 9: Feb 25 – 29**

- ✓ Done but with different data
- ✓ done
- ✓ Done, even at 3
- ✓ started
- ✓ Not done

- *Milestone 10: done*

Ref. ATLAS Planning for FDR-I and CCRC-I vs. I Jan.28

<http://www.nikhef.nl/~bosk/documents/d2008/ATLAS-Document-for-FDR-and-CCRC.pdf>

Some more observations

- CCRC was a very useful exercise for ATLAS
- Achieved most milestones in spite of external dependencies
- The small size of the FDR data sample was a complication
- It's difficult to serve the Detector, Physics and IT community
- Maybe we should hope FDR2 does not coincide with CCRC2
- A T2 workshop in Moscow was a major disruption
- The same holds for S&C workshop: we don't have enough people
- We do not want to stop but carry on March + April with CCRC 1.5
- But first we have to do M6 this week

CCRC08-1.5 Planning

TT=Throughput Test FT=Functional Test T1-T1 FTS channel tuning

March

- 03 -07 M6
- 10 – 14 FT T1-T1 Lyon, PIC
- 17 – 20 TT

March 25 deadline SRMv2 @ CERN

- 21 -24 Easter
- 25 – 28 FT T1-T1 BNL, SARA

April 02 deadline for re-proc @ all T1's

April 02 deadline for SRMv2 @ all T2's

April

- 31 – 04 TT
- 07 – 11 FT T1-T1 RAL, NDGF
- 14 – 16 TT
- 21 – 25 FT T1-T1 CNAF, FZK
- 28 – 30 TT
- 01 – 04 Ascension

May 02 deadline for re-proc FDR data

May 02 deadline for all T1-T1 tested

May 05 – 30 CCRC08-2

Back-up slides

Week III site-by-site

- **ASGC (CASTOR)**

- ✓ **Permission problems on the disk pools for a few days**

- ➔ **Credential mapping**

- ✓ **Smooth running once this has been fixed**

- ➔ **Except for a few hours of SRM frontend downtime on Feb. 24th**

- **BNL (dCache)**

- ✓ **Run very smooth, nothing to report**

- ➔ **Feeling is that we are far from breaking it...**

Week III site-by-site

- **CNAF (CASTOR+StoRM)**

- ✓ **StoRM had a bug preventing >2GB files to be exported**

- ▶ **Took long time to figure out. Fixed right after.**

- ➔ **No major issue but several service glitches**

- ▶ **Mis-configured gridftp frontends etc ...**

- ✓ **CASTOR running smoothly**

- **FZK (dCache)**

- ✓ **NO_SPACE_LEFT problem. How is this possible for D0T1 areas?**

- ➔ **Need clarification**

Week III site-by-site

- **LYON, NDGF and PIC (dCache)**

- ✓ Quite smooth. Occasional SRM frontend downtimes.

- **RAL (CASTOR)**

- ✓ Frequent glitches in SRM frontend (SRM timeout)

- ✓ Configuration issues on space token definition

- ➔ Specified space token does not exist: atlasdatadisk

- **SARA+NIKHEF (dCache+DPM)**

- ✓ Nikhef suffered authentication problems (DPM bug)

- ✓ A lot of DISK FULL problems, partially due to extreme slowness in deleting files in Nikhef

- ➔ 0.2Hz, see Stephane talk.

- **TRIUMF**

- ✓ Running smoothly all over the exercise

Week 3 -Overall Issues

- Some sites are tight in disk space
 - ✓ Remember data were supposed to retain for 1 week
 - ➔ Necessary for subsequent exercises (T1-T1 and T1-T2)
 - ✓ Retained for 2/3 days only
 - ✓ Excellent test for bulk deletion!
- Bug in the FTS proxy delegation mechanism
 - ✓ Appears as authentication problem (public/private key mismatch)
 - ➔ Workaround at server and client side now in place
 - ✓ Hit very hard the T0 FTS on Feb. 17-18 and T1 FTSES on Feb. 20
 - ➔ This affects both T0 and T1s FTSES servers

Week 3 -Overall Issues

- Problem with CASTOR disk server at T0
 - ✓ Hardware problem on Sat 17th and Sun 24th Feb
 - ➔ Service degradation
 - ➔ CASTOR team is working on automatic detection and cure
 - ✓ Need to revisit the problem notification process (also for T1s)
- dCache race condition in directory creation
 - ✓ “File Exists” on first attempt
 - ➔ Minor impact, rare condition (but significant under heavy load)

More General Considerations

- **FTS service**
 - ✓ Delegation problem already discussed
 - ✓ Different configuration for “shares” enforcement. Comments from RAL?
 - ✓ FTS monitoring from every cloud (including CERN)
- **LFC service**
 - ✓ Has been running very reliably.
- **DDM Central Catalog and Site Services**
 - ✓ Operations for T0 exports carried on by “non experts”
 - ➔ **Developers intervention in case of need**
 - ✓ Now reliable and simpler to operate (very useful DB peek tool)

Monte Carlo with SRM2.2

- Need to start using SRM2.2 for MC production
 - ✓ Panda Server and Panda job are SRM2.2 aware
 - ➔ Running in “SRM2.2 mode” in NL cloud. Encouraging results
 - ✓ Missing pieces:
 - ➔ SRM2.2 is in every T1, but what about T2s? Space tokens?
 - ➔ DPM needs a fix for ACL in directory creation.
 - Backported to DPM1.6.7 (current version). Under certification.
 - ➔ Need lcg_util 1.6.7, available only for SL4
 - All T1s are on SL4, but T2s? An ATLAS deadline (March 15) has been announced
- My feeling: FDR2 production with mixed SRMv1/v2 scenario
 - ✓ Not a problem for T2 (production “buffers”).