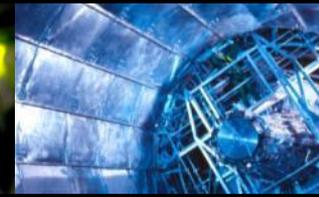
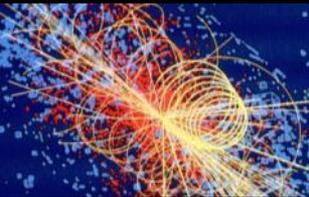


Project Status Report

Ian Bird

Computing Resource Review Board

12th October, 2010

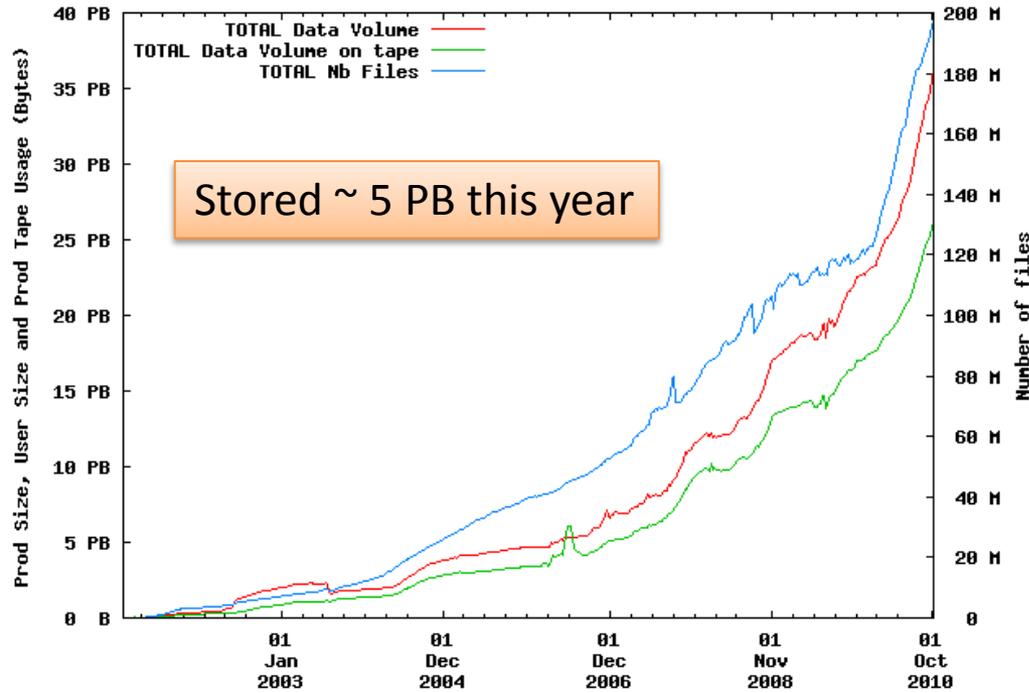


Project status report

- Overall status – first 6 months of data
- Planning and milestones
 - Status of planning for Tier 0
 - EGEE → EGI
 - Evolution of the WLCG service

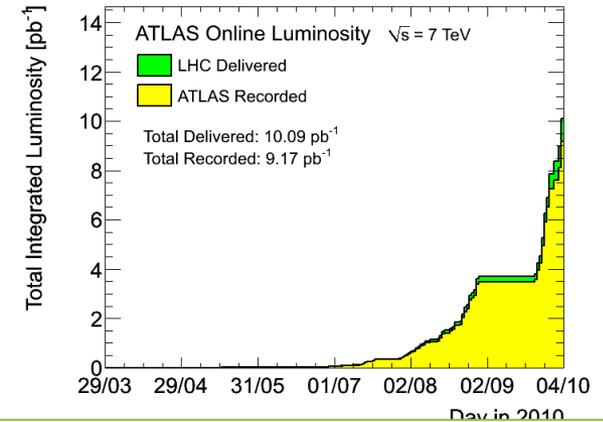
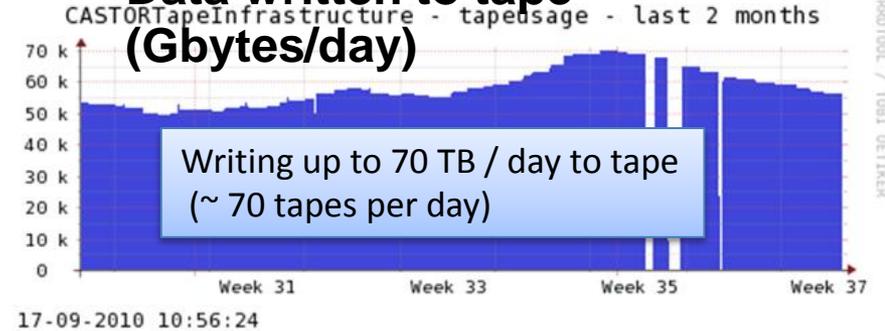
6 months of LHC data

Experiments Production Data and Experiments User Data in CASTOR

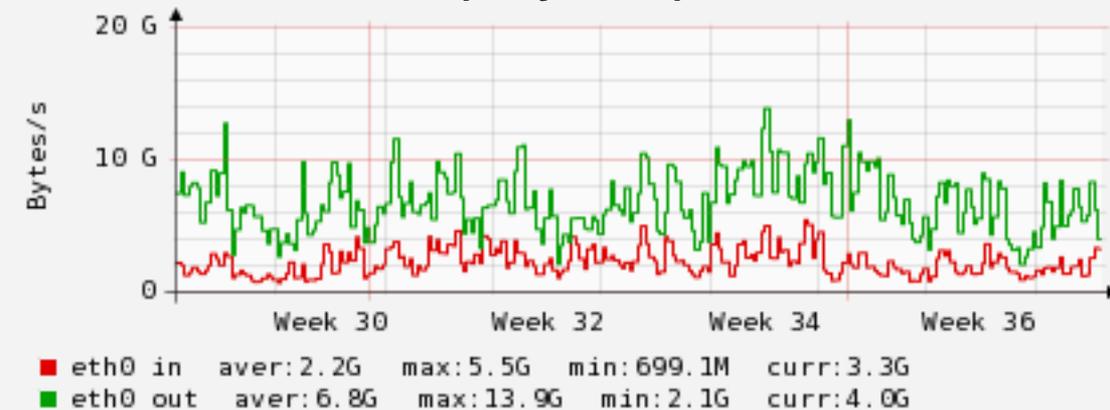


Generated Oct 05, 2010 CASTOR (c) CERN/IT

Data written to tape (Gbytes/day)



Disk Servers (Gbytes/s)

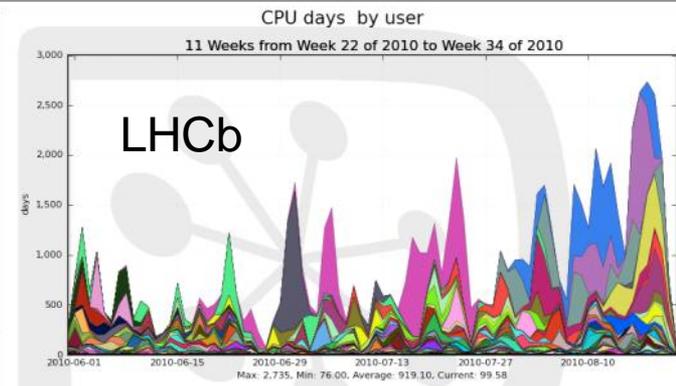
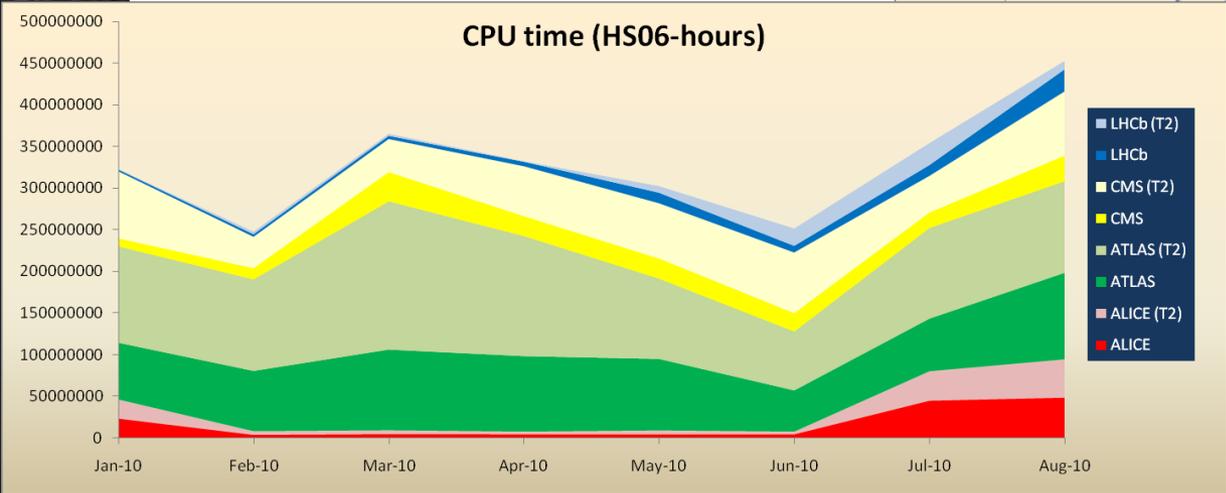
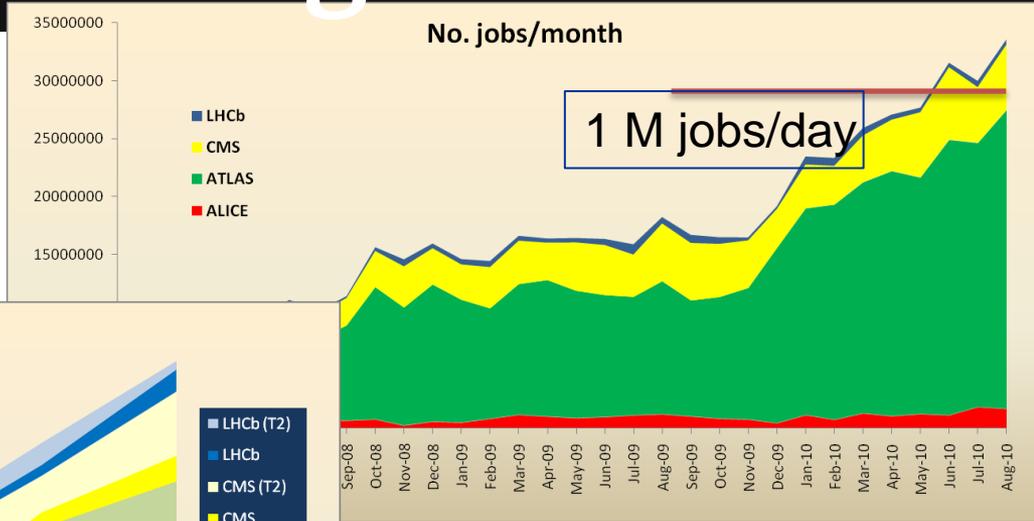


Tier 0 storage:

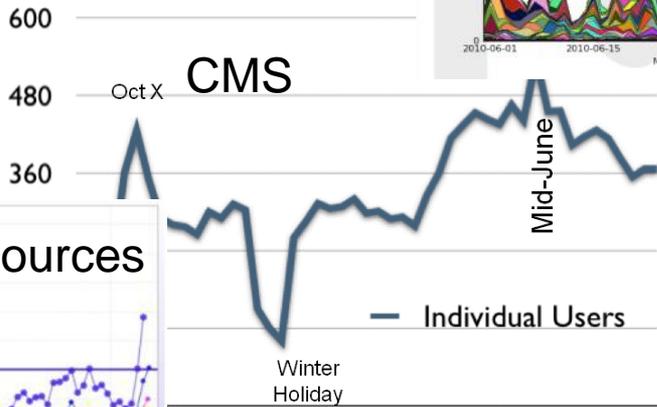
- Accepts data at average of 2.2 GB/s; peaks > 6 GB/s
- Serves data at average of 7 GB/s; peaks > 14 GB/s
- **CERN Tier 0 moves ~ 1 PB data per day**

WLCG Usage

- Use remains consistently high
 - 1 M jobs/day; 100k CPU-days/day
 - Pilot jobs → real task load higher



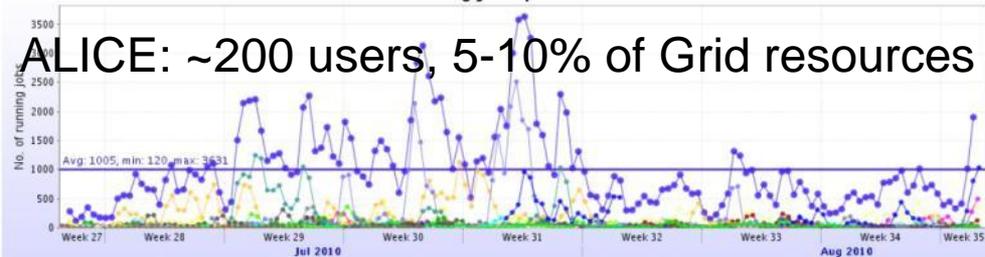
As well as LHC data, large simulation productions ongoing



Large numbers of analysis users

CMS ~500,
ATLAS ~1000,
LHCb/ALICE ~200

WLCG Computing Grid

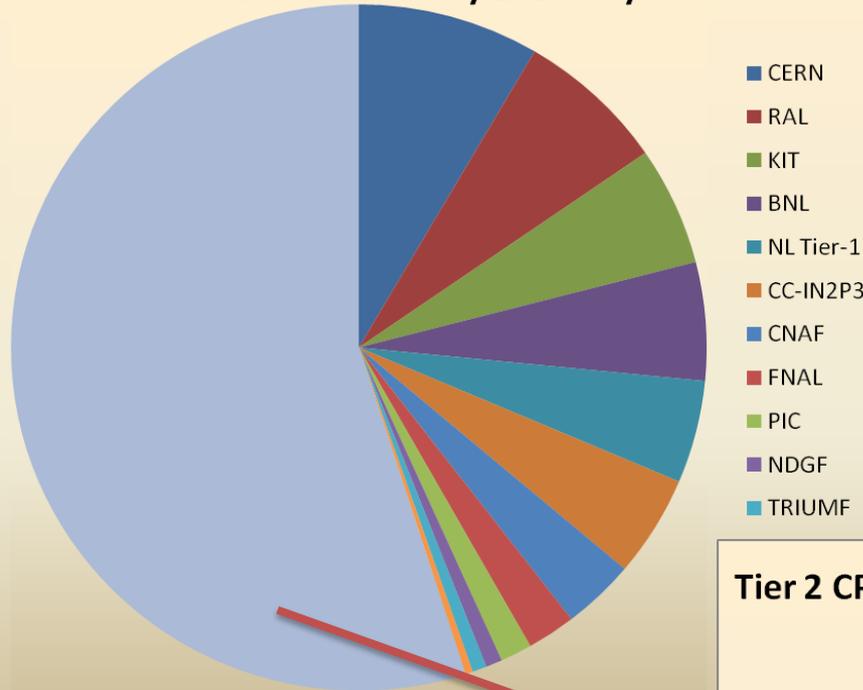


45 48 51 2 5 8 11 14 17 20 23 26 29 32

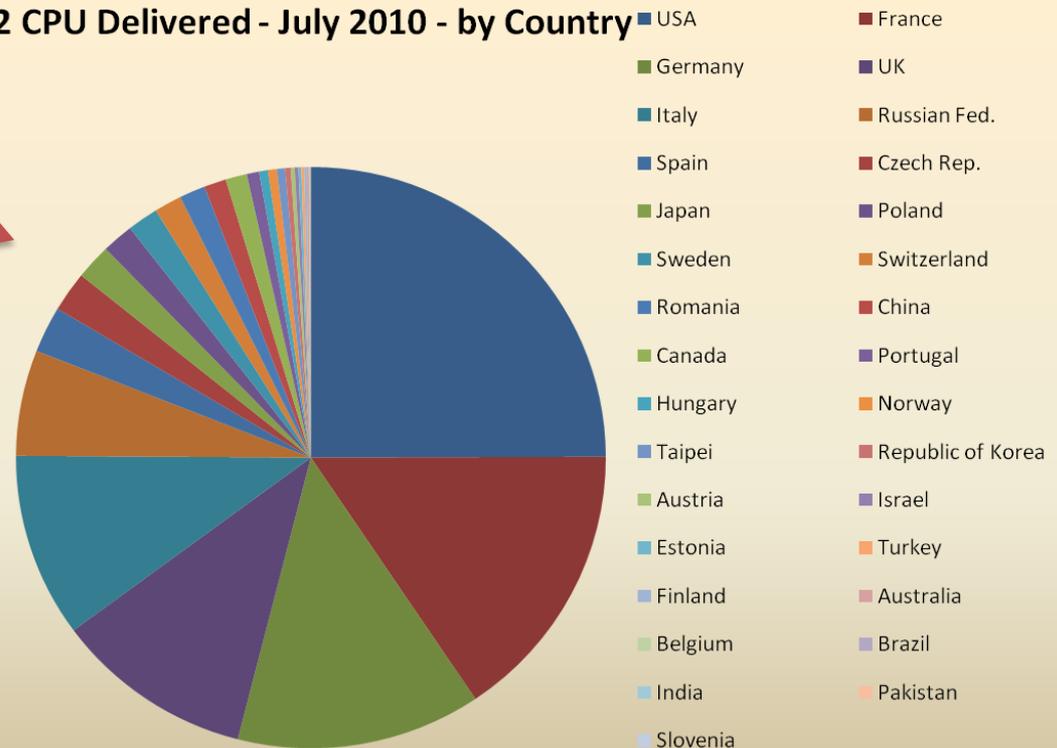
CPU – July

- Significant use of Tier 2s for analysis
 - frequently-expressed concern that too much analysis would be done at CERN is not reflected

CPU Delivered - July 2010 - by Tier



Tier 2 CPU Delivered - July 2010 - by Country

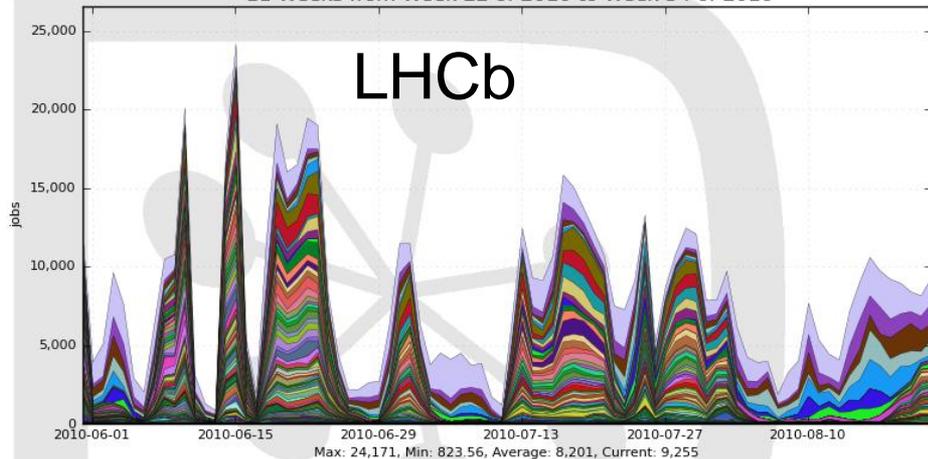


- Tier 0 capacity underused in general
 - But this is expected to change as luminosity increases

Job workloads

Running jobs at all sites

11 Weeks from Week 22 of 2010 to Week 34 of 2010

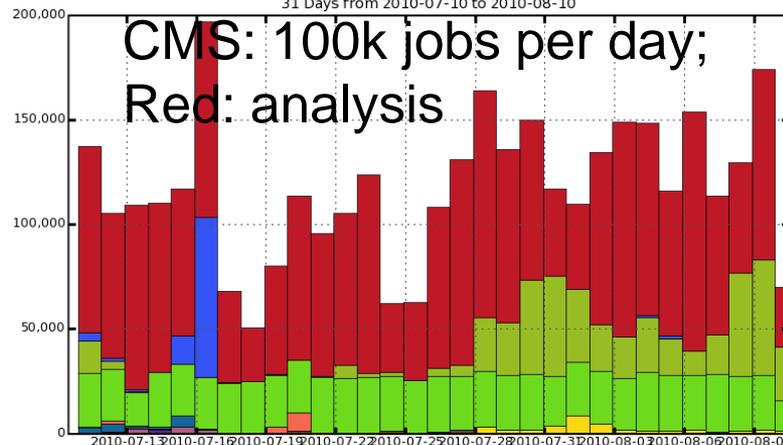


Max: 24,171, Min: 823.56, Average: 8,201, Current: 9,255



Terminated jobs

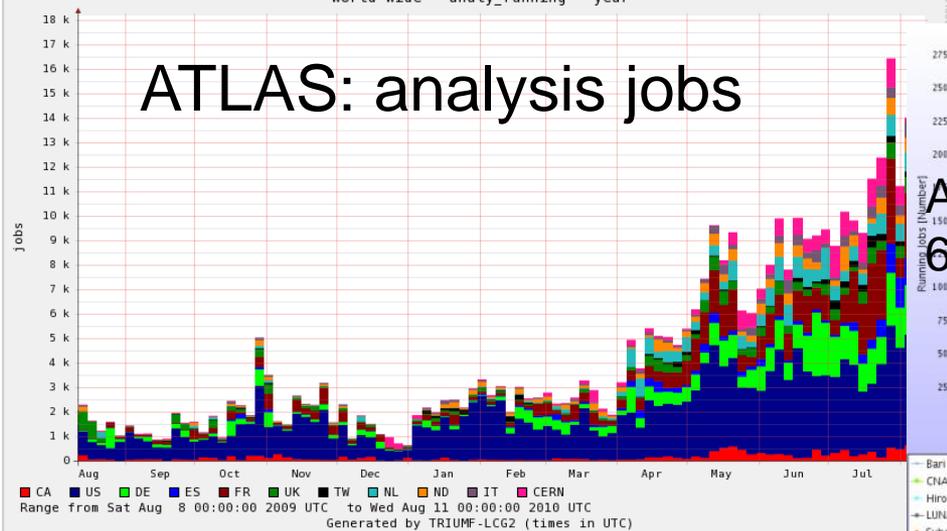
31 Days from 2010-07-10 to 2010-08-10



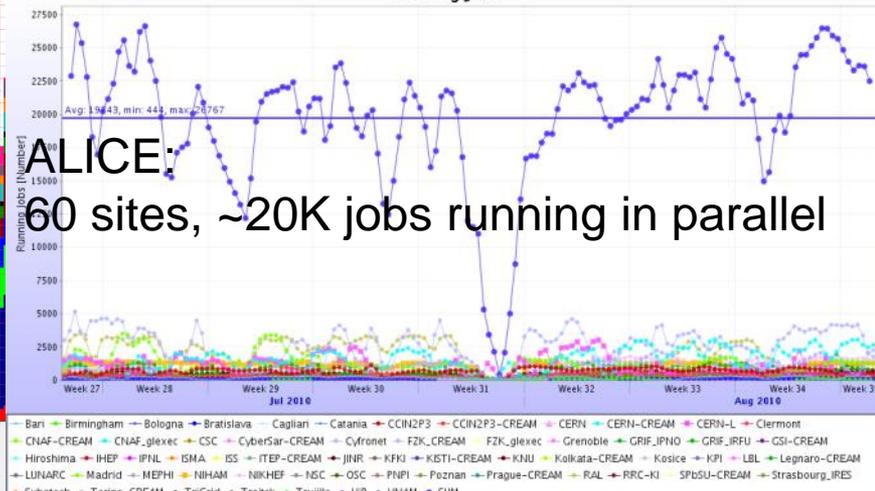
Maximum: 196,944, Minimum: 0.00, Average: 113,779, Current: 70,004

World Wide - analy_running - year

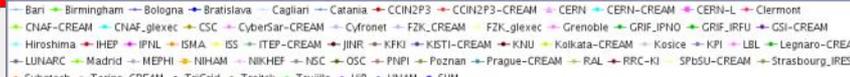
Generated on 2010-08-22 09:04:15 UTC



Running Jobs

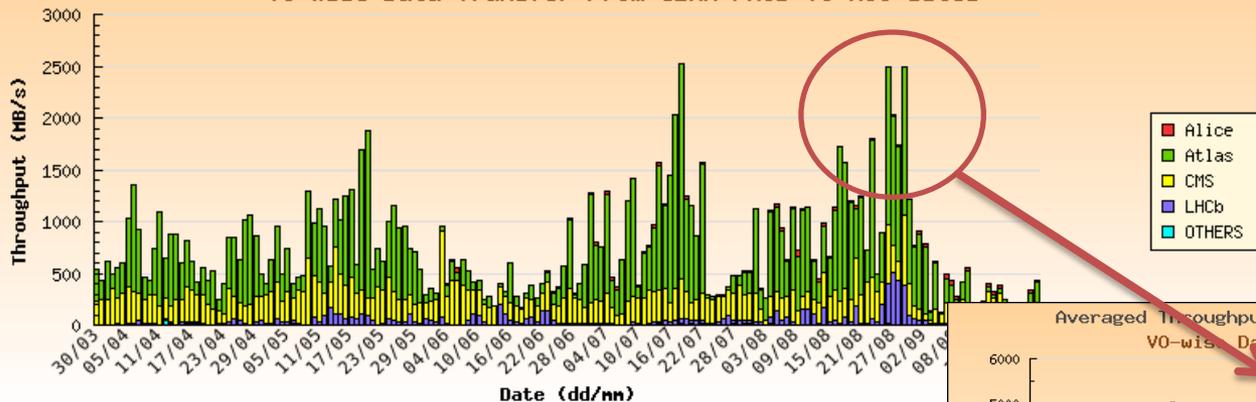


ALICE:
60 sites, ~20K jobs running in parallel

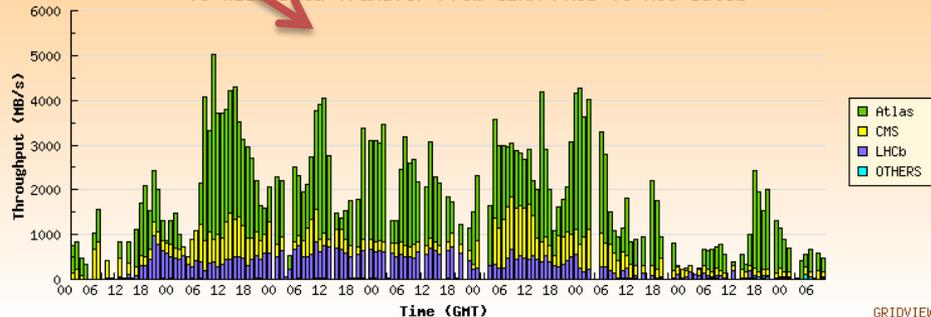


Data transfers: T0 – T1

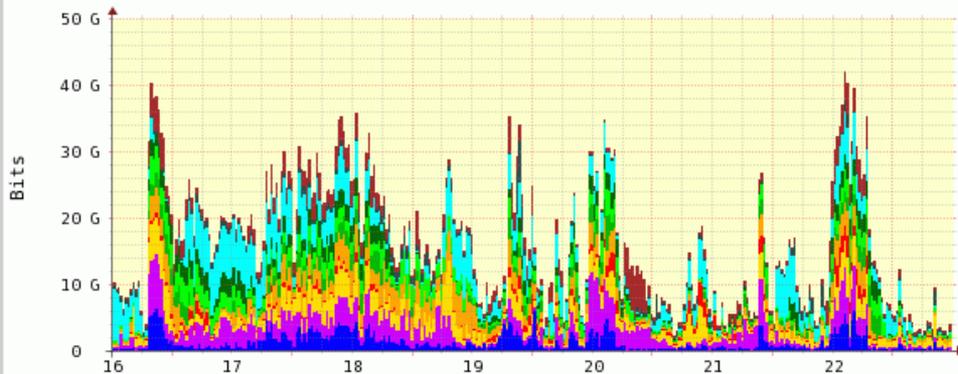
Averaged Throughput From 30/03/10 To 23/09/10
V0-wise Data Transfer From CERN-PROD To All Sites



Averaged Throughput from 00 Hrs on 25/08/10 to 08 Hrs on 01/09/10
V0-wise Data Transfer From CERN-PROD To All Sites



LHCOFN TOTAL Traffic Flow 1 (Out-bound)



Traffic on OPN up to 70 Gb/s!
- ATLAS early reprocessing campaigns

- Data transfers (mostly!) at rates anticipated

For all experiments:
early data has been
available for analysis
within hours of data
taking

Reliabilities

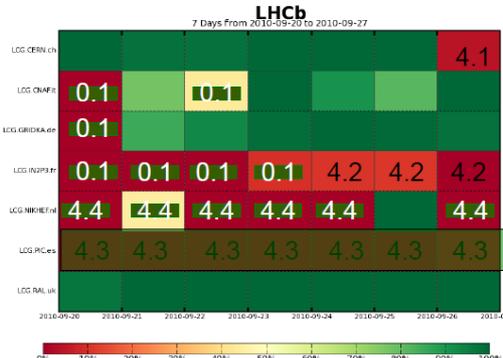
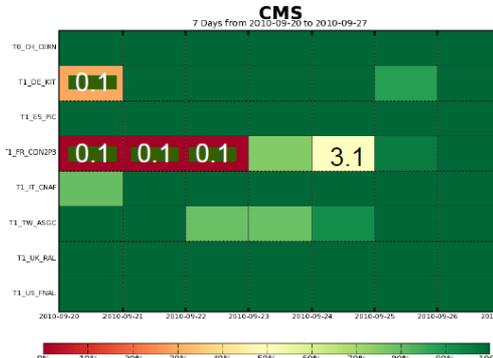
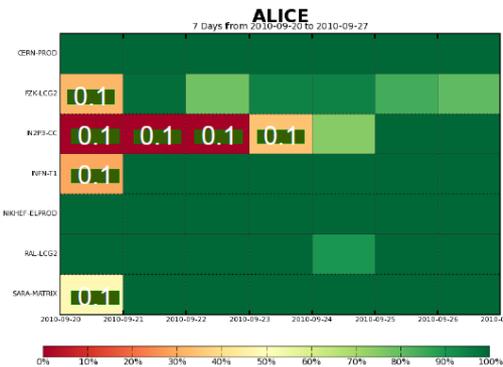
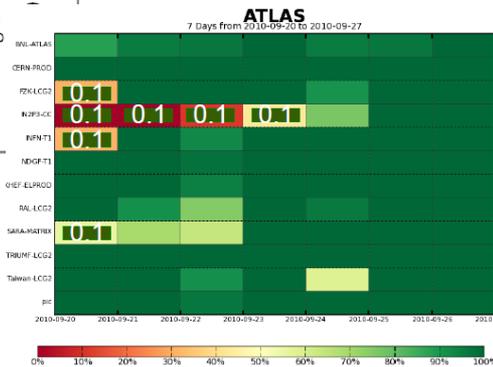
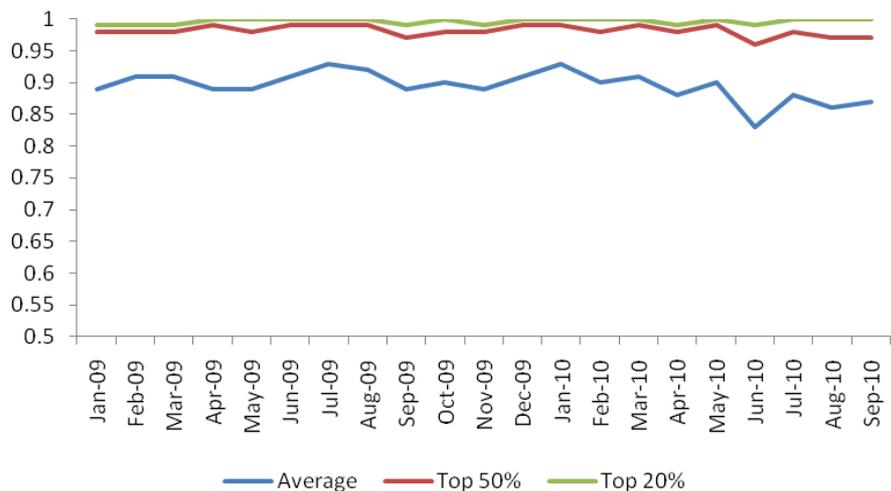
Site Reliability: CERN + Tier 1s



Experiment-measured site availabilities:

Includes down times during security patching;
At times ~50% of resources were unavailable.

Tier 2 Reliabilities



Service Incidents

- Operational problems at sites requiring an incident report (i.e. That have a user-visible service degradation), remain largest stability issue.
 - ~5 -6 per month; some few hours, some more extended
 - Databases, power/cooling, hardware failures, etc.
 - Not grid middleware
- End Sept. had a security problem – not grid specific – requiring urgent reactions from sites to protect themselves
 - Net result was significant reduction in overall capacity
 - This is unavoidable at some level
- Have to become more resilient to these;
 - evolution of computing models (cannot assume all sites will always be available)
 - Ensure lessons are learned by other sites

ALICE data loss (May)

- A configuration error in Castor resulted in data being directed across all available tape pools instead of to the dedicated raw data pools
 - For ALICE, ATLAS, CMS this included a pool where the tapes were re-cycled after a certain time
- The result of this was that a number of files were lost on tapes that were recycled
- For ATLAS and CMS the tapes had not been overwritten and could be fully recovered (fall back would have been to re-copy files back from Tier 1s)
- For ALICE 10k files were on tapes that were recycled, inc 1700 files of 900 GeV data
- Actions taken:
 - Underlying problem addressed; all recycle pools removed
 - Software change procedures were reviewed
 - Tapes sent to IBM and SUN for recovery – have been able to recover ~97% of critical (900 GeV sample) files, ~50% of all ALICE files
 - Action to improve user-facing monitoring in Castor
 - As this was essentially a procedural problem: a review of Castor operations procedures (sw dev, deployment, operation etc) was organised together with experiments and outside experts.
 - Work with ALICE to ensure that always 2 copies of data available
 - In HI running there is a risk for several weeks until all data is copied to Tier 1s; several options to mitigate this risk under discussion

Milestones

- Pilot jobs, glxec, SCAS, etc
 - No pressure at the moment as data taking takes priority
- CREAM deployment:
 - >100 sites have a CREAM CE;
 - Ongoing testing of submission from condor_g
 - Reliability and scalability still improving

Milestones – 2

- Storage accounting reports:
 - For Tier 1 sites done (as part of installed capacity)
 - For Tier 2 sites, for end Oct
 - Before these are regularly published must finish validation of the published data
- Gathering of installed capacity →

Installed capacity

- Action has been pending for a long time:
- Significant effort was put into defining and agreeing how to publish and collect the data – via the information system.
 - Document has been available for some time
- Tool now has the ability to present this information →
 - Validation for Tier 0 and Tier 1s done
 - Significant effort still required to validate this data for Tier 2 s before it is publishable
 - Tier 1s tasked with validation for Tier 2s
 - Publication of shares between experiments still to be validated fully

Installed capacities reporting

Topology Pledges **Capacities**

Capacities > Pledges Comparison

VO: --SELECT A VO-- Year: 2010 Month: 10

All Tiers Tier 0 Tier 1 **Tier 2**

JSON CSV Search:

Country	Federation	Physical CPU	Logical CPU	HEPSPEC06	CPU Pledge	Total Online Storage (GB)	Disk Pledge	Total Nearline Storage (GB)	Tape Pledge
Canada	CA-TRIUMF	464	1,208	13,460	10,800	1,100,000	1,095,000	750,000	710,000
France	FR-CCIN2P3	1,614	9,072	78,019	44,186	5,393,512	5,109,000	22,534,710	5,300,000
Germany	DE-KIT	2,645	9,770	88,712	58,730	6,844,258	6,924,000	9,292,000	8,932,000
Italy	IT-INFN-CNAF	2,252	8,192	85,516	44,000	4,570,571	5,300,000	9,510,000	5,450,000
Netherlands	NL-T1	941	4,512	48,241	47,296	3,995,414	4,186,000	1,263,330	3,629,000
Nordic	NDGF	10,584	10,584	61,122	13,710	3,076,621	1,720,000	1,464,000	1,770,000
Spain	ES-PIC	592	2,368	24,144	17,238	3,086,334	1,968,000	3,852,371	2,136,000
Taiwan	TW-ASGC	990	3,960	40,154	28,000	2,400,000	3,500,000	2,133,000	3,500,000
UK	UK-T1-RAL	1,300	5,200	49,140	44,376	5,977,227	4,638,000	3,627,649	4,877,000
USA	US-FNAL-CMS	1,692	6,768	44,400	44,400	4,100,000	4,100,000	11,000,000	11,000,000
USA	US-T1-BNL	1,658	5,652	58,000	49,680	6,100,000	5,037,000	4,000,000	3,266,000
Total		24,732	67,286	590,908	402,416	46,643,937	43,577,000	69,427,060	50,570,000

Showing 1 to 11 of 11 entries

Planning for Heavy Ion run

- The conditions for HI running are only now becoming clear:
 - ALICE:
 - will run at 200 Hz (cf 100) and collect 200 TB/day (cf 100); data rate < 2.5 GB/s (cf 1.25)
 - Same total data volume as original planning; but much higher rate
 - ATLAS:
 - Similar to pp running conditions: ~ 200-300 MB/s; 10-20 TB/day
 - CMS:
 - 150 Hz but no zero suppression or trigger → 1.8 GB/s (150 TB/day)
 - LHCb:
 - No HI running

- Doubling of ALICE rate
 - (same integrated data volume as planned)
- New large CMS request

Heavy Ion considerations

- These requests significantly exceed the anticipated (and tested) data rates
 - And there has been no readiness test for combined HI operations at this scale
- Data export:
 - ALICE Tier 1s aggregate acceptance cannot match data taking rate; CMS intends to export full sample to FNAL (ATLAS as for pp)
 - Export will take much longer than data taking period
- Strategy
 - Working with experiments to match resources to needs – we have some resources from 2011 pre-purchases
 - For ALICE: have planned to increase disk pools to ensure 2 copies of data at CERN until all has been copied to Tier 1s
 - A readiness test for ALICE at new higher rate is already scheduled
 - Arrange a combined test for ALICE and CMS at higher rates
 - Look at mechanisms to absorb the load (e.g. make use of disk buffers at the experiments to decrease the instantaneous rate if needed)

Budget at CERN

- Part of MTP reduction comes from the LCG materials:
 1. Move to 4-year equipment replacement cycles
 - Most savings in 2011, very little afterwards
 - Potentially negative effect on power budget
 - Could impact service quality
 2. Reduce slope of Tier 0 computing resource increase
 - Save ~ 1 MCHF/year on average
 - This is the main mechanism we have to reduce costs. Current assumption was ~30%/year growth.
 - Helps delay the need for additional computing infrastructure
- Implications:
 - May limit the overall growth of Tier 0 resources
 - NB As yet no experience with HI needs; long term needs of experiments still not clear
- Detailed planning requires further study and input

Tier 0 planning – status

- Plan for a new Tier 0 building at CERN is cancelled
- Delayed need for containers (or other alternate solution) by ~1 year or more (→2014):
 - Accelerator stop in 2012, 2016 (cf original power plan)
 - Efforts invested in recent years have benefitted total power needs
 - Ongoing plans for upgrading B513 from 2.9 to 3.5MW including up to 600 kW diesel-backed power, together with use of local hosting (100 kW today, potentially more)
 - Budget reduction will have implications for total power needs
 - Reanalysis of strategy for next few years will be done

Tier 0 evolution – remote hosting

- For the longer term, or as an alternative to containers, investigate **hosting outside of the CERN territory**
 - equipment would be installed and operated in a remote facility
- Several member states have expressed their interest in proposing such a solution
- The Council was invited to solicit informal bids for co-hosting of data centre computing equipment from institutes/commercial consortia:
 - in order to determine the feasibility, interest and possible economic advantages
 - Submitters were invited to estimate the usable electrical power (kW) that would be available for hosting for an annual charge of 4 MCHF.
 - Should the proposals be of economic interest, further more formal steps could then be proposed.

- Coordination for European Grid resources
 - Established February 8th 2010
 - Central policy & services needed to run a grid
 - Sustainable small coordinating organisation
- Governance & ownership by its participants
 - EGI Council votes linked to fees
 - Resources from within its participants
- Located in Amsterdam with approx. 40 staff
 - Coordinating core (~20 people) in Amsterdam

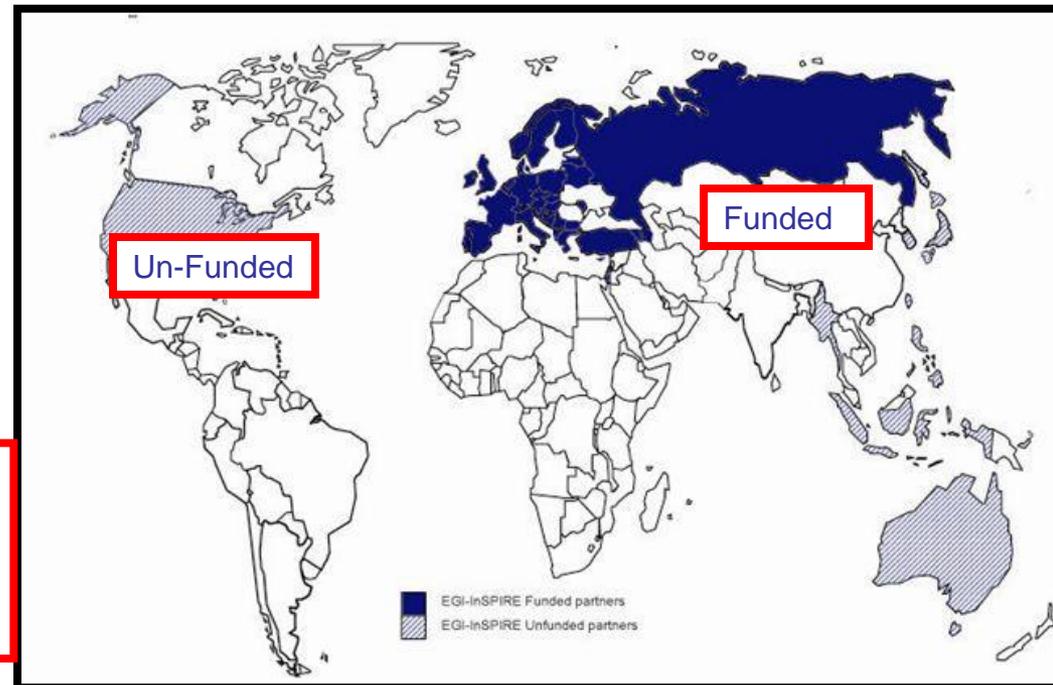
EGI and EGI.eu supported by EGI-InSPIRE project

Integrated Sustainable Pan-European Infrastructure for Researchers in Europe

- A 4 year project with €25M EC contribution
 - Project cost €72M
 - Total Effort ~€330M
 - Effort: 9261PMs
- Started 1st May '10

Project Partners (50)

- EGI.eu, 39 NGIs, 2 EIROs
- Asia Pacific (8 partners)



- Virtual Research Community MoU
 - User Community Board: Technical user input
 - Contact points for community representation
- External Advisory Board
 - Strategic input to the PMB
- Services for Heavy User Communities
 - HEP User Support
 - GANGA & DIANE
 - Dashboards

EGEE-EGI: Summary

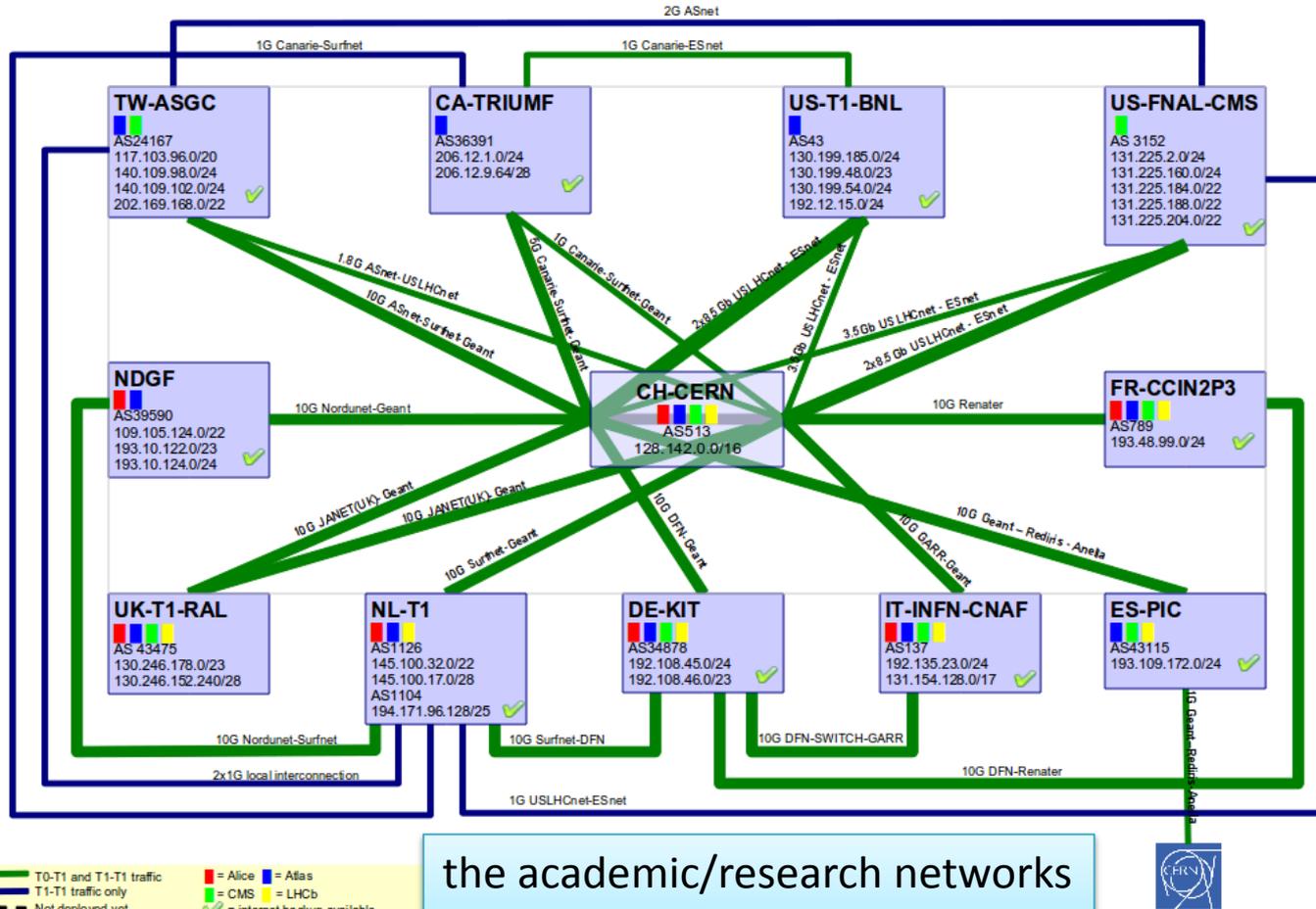
- EGI and EMI in place for several months now
- Some teething issues to be expected, but no major problems
- However, very many areas where procedures and processes and contact points are still undefined
- Main concern is that we (WLCG) need to keep the operation going and have had to be pragmatic (e.g. New release processes, pushing for processes for new sites,...)
 - Operationally EGI is low key and not very visible – main contact is via Tier 1s and NGIs

Service evolution strategy

- Have started discussion on several aspects –
 - Data management:
 - Discussion in March, workshop in June, follow ups in GDBs
 - Multi-core and virtualisation:
 - 2nd workshop was held in June (1st was last year)
- Long term:
 - Clarify the distinction between the WLCG distributed computing infrastructure and the software that we use to implement it
 - Understand how to make the grid middleware supportable and sustainable
 - Use of other components (e.g. Industrial messaging, tools like Nagios, etc);
 - what does cloud technology bring us?
 - and how to integrate/maintain our advantages – global trust, VOs, etc.

e.g. Networking

LHCOPN



the academic/research networks for Tier1/2!

- Requirements – working group set up to discuss with network communities for the future needs and expectations

Summary

- After 6 months of data taking WLCG is working as planned
 - Significant numbers of people doing analysis
 - Tier 2s really used for analysis
 - Network traffic close to what was planned
 - Operational load is manageable ...
- Planning for Heavy Ion run in hand – but some concerns over last minute increases in requirements
 - Need to fit in testing
- EGI transition has not disrupted operations
- Considerations for evolution of various aspects of the system
 - Make more effective use of networks