'Robust Data Transfer' Service Challenge Meeting

"A Data Movement Service for the LHC"

James Casey, IT-GD, CERN

NIKHEF/SARA, 12 October 2004

A Data Movement Service for the LHC

- Service Overview
- CERN Configuration
- Tier-1 Network and Storage Setup
- Transfer software

Service Overview

- Aim is to prototype the data movement system needed for LCG startup
  - Progressive milestones to have all components in place 3 months before first data from LHC
  - Details in Les's slides
- Many of the components already exist
  - But have not been proven together
  - Nor at the required data rates and reliability levels
- Need to get the service teams who already look after infrastructure connected

Projected data rates and bandwidth requirements

	RAL	Fermilab	Brookhaven	Karlsruhe	IN2P3	CNAF	PIC
Data Rate (MB/sec)	182.49	69.29	173.53	317.69	317.69	317.69	182.49
Total Bandwidth (Gb/sec)	4.38	1.66	4.16	7.62	7.62	7.62	4.38
Assumed provisioned bandwidth	10.00	10.00	10.00	10.00	10.00	10.00	10.00

	Taipei	Tokyo	Nordugrid	TRIUMF	NIKHEF
Data Rate (MB/sec)	176.15	106.87	106.87	106.87	113.20
Total Bandwidth (Gb/sec)	4.23	2.56	2.56	2.56	2.72
Assumed Provisioned bandwidth	10.00	10.00	10.00	10.00	10.00

*** Projections as of 30-09-04**

Note: Total Required bandwidth = (total * 1.5 *(headroom)*) * 2 *(capacity)*

Milestones

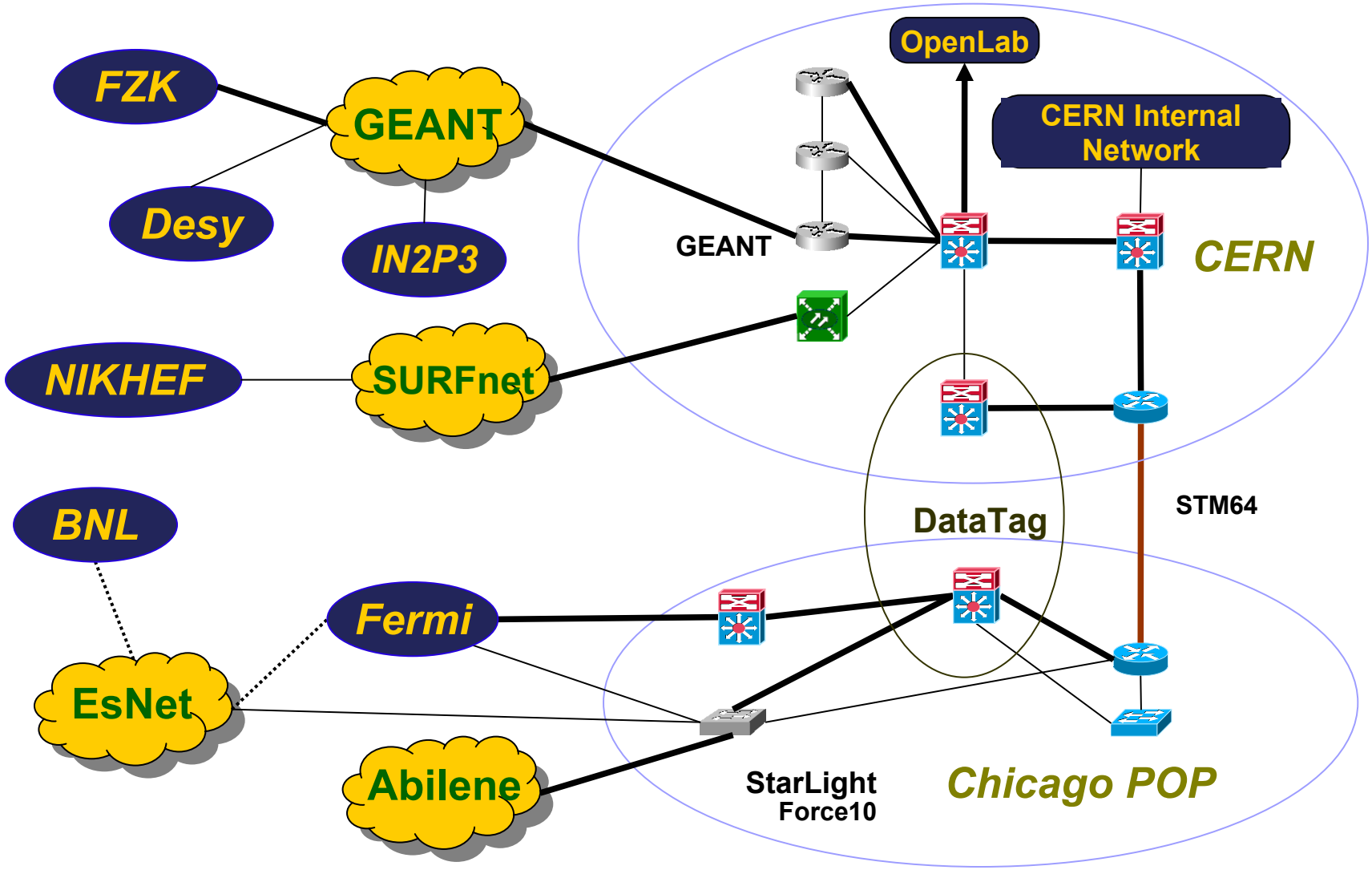
- Dec04 - Service Challenge 1 complete
  - mass store-mass store, CERN+3 sites, 500 MB/sec between sites, 2 weeks sustained
- Mar05 - Service Challenge 2 complete
  - reliable file transfer service, mass store-mass store, CERN+5 sites, 500 MB/sec between sites, 1 month sustained

- We will try and use first version of reliable file transfer service for Service Challenge 1
  - But focus is on network and storage infrastrucure

CERN Configuration

- 10x Itanium2 machines
  - Each has 1Gb link that gets aggregated into a 10Gb switch
  - All nodes run CERN SLC3
- Configured to run the following services:
  - 5 gridftp servers, non-load balanced
  - A 4-node load-balanced SRM/Gridftp system (not yet complete)
  - 1 control node, upon which the transfer management software will run
- Direct connections to external network
  - 10 Gb connection to GEANT
  - 10 Gb link to Chicago (via Starlight)

Current Network Layout



Required Resources at a Tier-1

- Network
  - Preferably a dedicated network connection to CERN
  - A defined dedicated subnet which will be routed to the CERN transfer cluster
- Storage
  - Gridftp will be the initial transfer protocol
  - Desirable for disk pool manager
    - to manage space
    - to load balance between storage nodes

Steps to add a new Tier-1

- 4 steps to getting a new Tier-1 connected and tested
  1. Initial connectivity and routing configuration of a site to the test cluster at CERN
  2. Deployment of managed storage at the site
     - short term transfers carried out to test and measure untuned performance
  3. Site tuning (at network and storage level)
  4. "Service Challenge". Can last
     - a few days for a site with non-dedicated network
       - Aim is to show that the network is the bottleneck
     - A few weeks for a site with sufficient bandwidth and hardware
       - Aim is to show reliable transfers at peak data rate (500MB/s disk-to-disk)

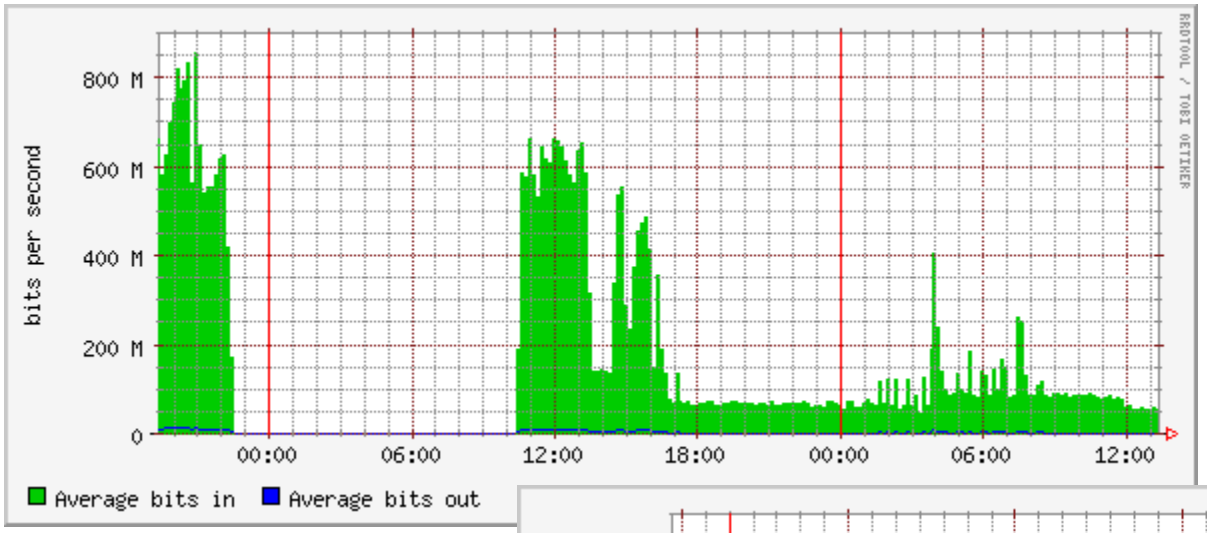
Current Status

	RAL	Fermilab	Brookhaven	Karlsruhe	IN2P3	CNAF	PIC
1. Network Configuration		√	√	…	√		
2. Storage Configation		√	√		…		
3. Site Tuning		…					
4. “Service Challenge”							

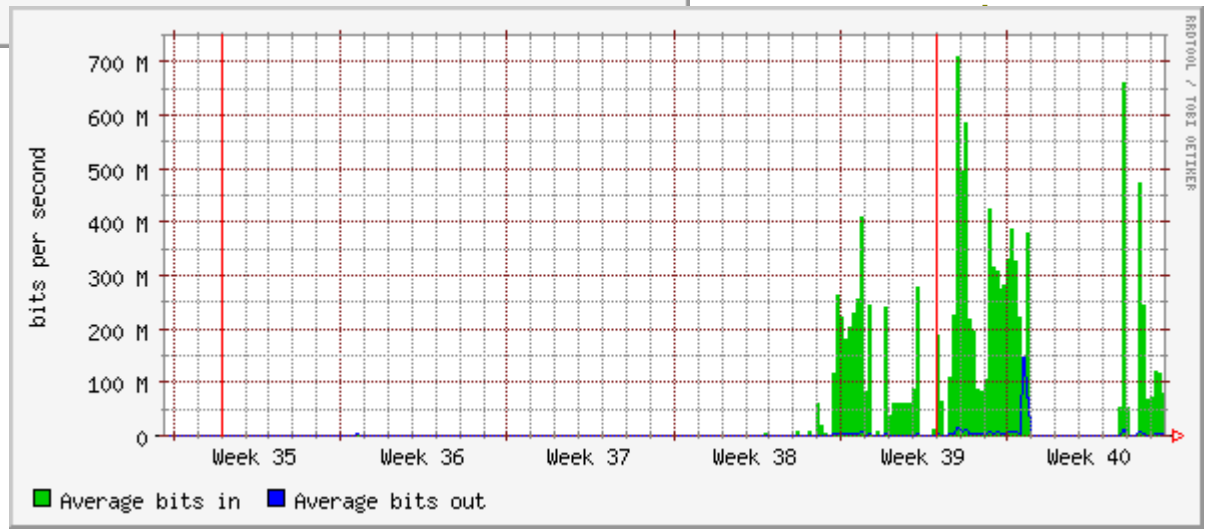
	Taipei	Tokyo	Nordugrid	TRIUMF	NIKHEF	DESY
1. Network Configuration					√	…
2. Storage Configation					√	
3. Site Tuning						
4. “Service Challenge”						

KEY
√ = Complete
… = In Progress

First measurements



- Measurements taken at CERN Router
- Most traffic is to FNAL
  - ~ 100MB/s



Next steps

- Each step of process takes approx. 2 weeks to complete
  - Aim to go through process with all Tier-1s by early 2005
  - First draft of document -
  - 'Robust Data Transfer' Service Challenge Tier-1 Information"
    - Describes procedure for joining
    - Details on CERN configuration
  - Sent out on the hep-forum-wan-tier1 mailing list
- Next steps
  - Start process with other European Tier-1s
  - Then tackle non-European ones
    - Additional problems due to long network path length
- Start to schedule 'slots' for December/January for sites to have Service Challenge

Transfer Software

- Proposal for File Movement System created by LCG-GD and EGEE-DM teams. Aim:
  - Create architecture that will let components from both projects interoperate
  - Allow sites to use their own existing storage and transfer tools where possible
- First draft of architecture – July 2004
  - https://edms.cern.ch/document/490347/2
  - Based on ideas from current systems
    - CMS TMDB, Globus RFT, Condor Stork

Status

- Simple CLI Tools created
  - lcg-move-start [--credentials cred] [--verbose]
  - lcg-move-submit [--id jobId]
    [--source src --dest dest || --file inputs]
  - lcg-move-job-ls [-l] [-a]
  - lcg-move-cancel [--id jobId] [--verbose]
  - lcg-move-get-job-summary [--id jobId] [--verbose]
  - lcg-move-get-job-status [--id jobId] [--verbose]
- Simple Transfer daemon exists
  - Started work with Condor Stork, but limitations showed up
  - Now use a simple perl multi-process daemon to transfer files
- gLite working on WS interfaces and Web-based UI

Summary

- Service Challenges have aggressive deadlines
- Basic infrastructure in place at several sites
  - Need to work on next layer of infrastructure – transfer service tools
  - Need to get some other sites onboard
- Scheduling of Service Challenge slots starting now
  - NIKHEF - ???