DATA ACCESS and DATA MANAGEMENT CHALLENGES in CMS

Challenges

- CMS produces ~20PB of raw and derived data per year
  - An average replication factor of ~3
- 70 Computing sites that are globally distributed
- How to deliver samples to 150k processor cores as directed by the experiment centrally and thousands of scientists

Network

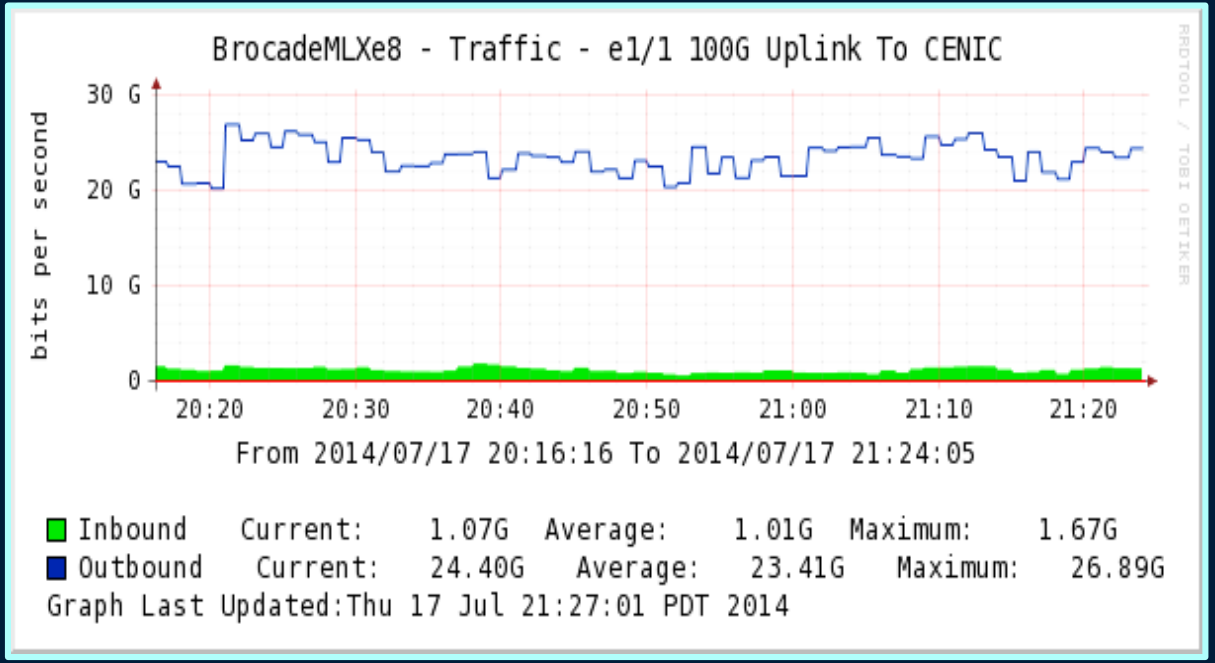
- The network capacity itself is keeping pace (just barely) due to the availability of 100Gb/s links
  - However we have a factor of 100 between are best and worst connected sites
  - Our ability to drive the network efficiently is still an issue, we use a lot of hardware to fill the pipes

Transfer Rates: Caltech Tier2 to Europe July 2014

One Day after commissioning the 1st 100G TA research link

Upload rate: 27 Gbps; 20Gbps to CNAF (Italy) Alone

- By Spring 2015: 12 – 40 Gbps Downloads were Routine to US CMS Tier2 Sites with 100G Links



US CMS university based Tier2s have moved to ~100G now

Caltech	100 Gbps
Florida	100 Gbps
MIT	100 Gbps
Nebraska	100 Gbps
Purdue	100 Gbps
UCSD	80 Gbps
Wisconsin	100 Gbps

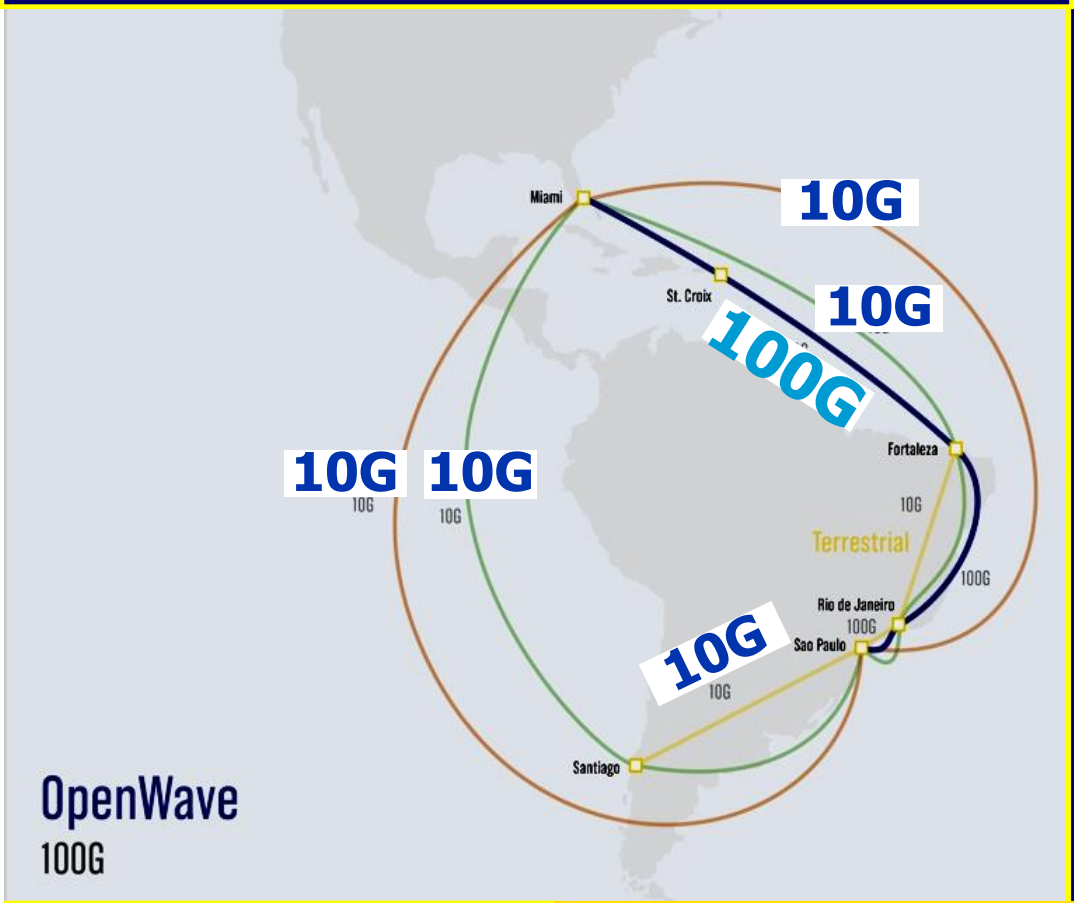
Harvey Newman

The move to 100G is timely and matches current needs, also at Tier2s. Backbones should continue to advance to meet the needs during Run2.

OpenWave: First 100G Link to Latin America in 2015. Connecting LSST

AmLight (US NSF) with RNP, ANSP

Total Capacity for Next Two Years: 140G



- An "Alien Wave" at 100G on the Undersea Cable
- ➡ Precedent-setting access to the frequency spectrum by the academic community
- ➡ Sao Paulo-Rio-Fortaleza -St. Croix-Miami backbone
- ➡ Scheduled to start soon
- 100G extensions by RNP in Rio and ANSP in Sao Paulo
- Will be extended to Chile at 100G then N X 100G
- Will be heavily used by LSST into the 2030s

Harvey Newman

February 2015

J. Ibarra, AmLight

➡ Using Padtec (BR) 100G equipment. Demonstrations with the HEP team (Caltech et al) at SC2013 and 2014

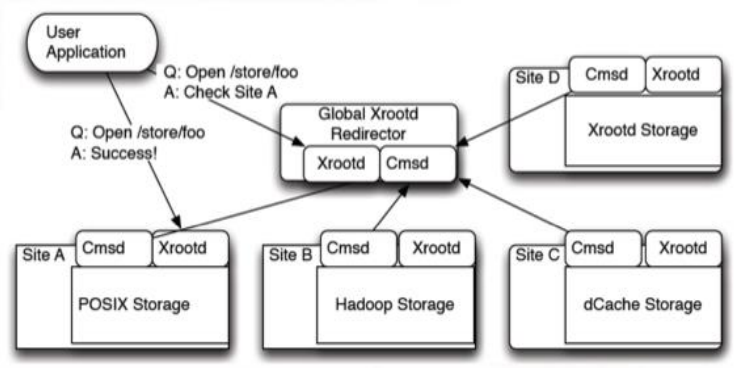
Coupling and Decoupling Services

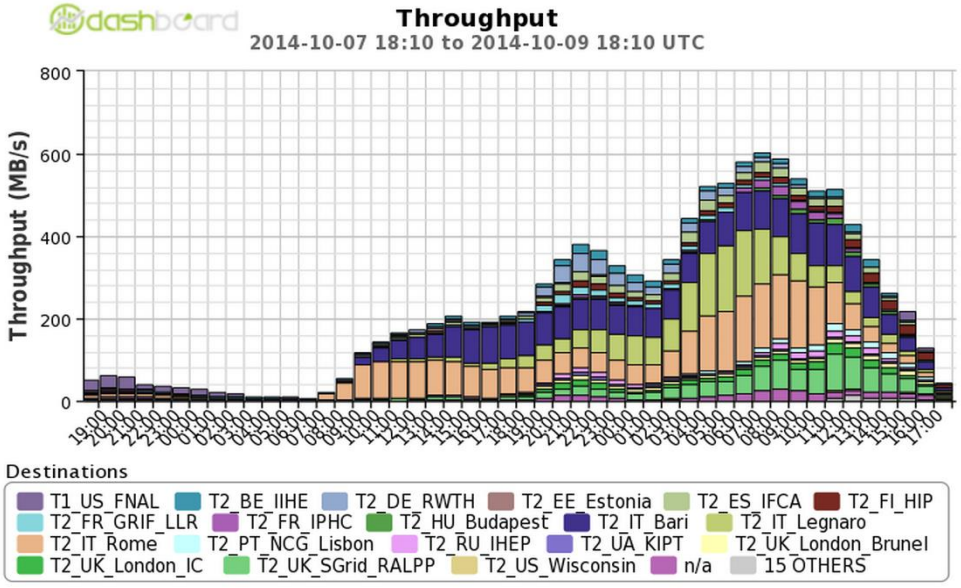
- We have spent the last several years trying to allow the processing and storage services to be more independent
  - Disk is expensive and normally has move IO capability than the amount of local processing services
  - Before this there was a lot of worry about the balance of CPU and storage
    - CPU can be scheduled more dynamically
    - CPU can be used opportunistically

Data Federation in Run II

Any Data Anytime Anywhere has been a primary focus area

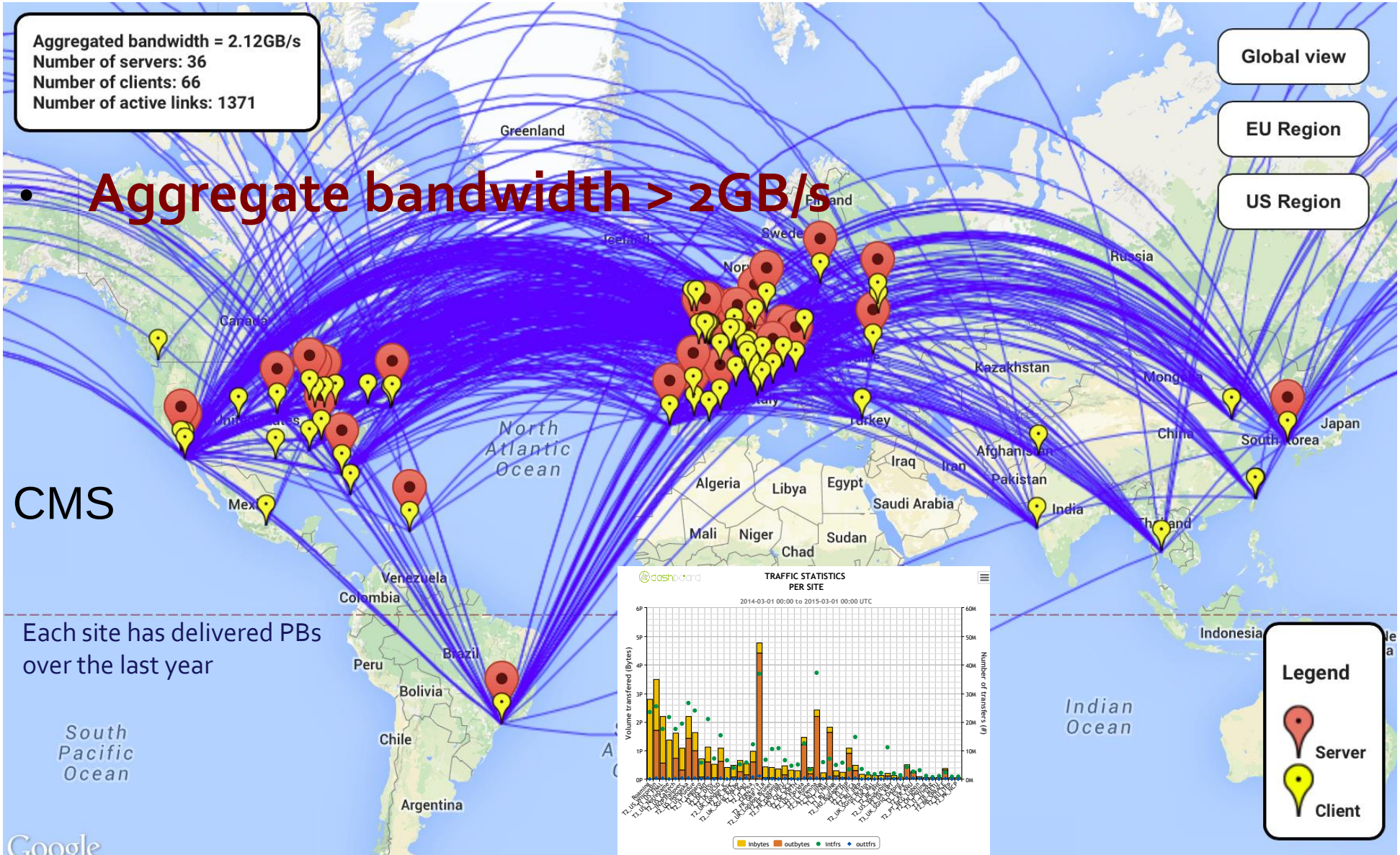
- We validated small scale use of non-local data access in the summer
  - Fall-back when analysis jobs do not find data
  - Very good feedback by users
- After summer scale tests were performed in Europe and the US
  - 20% of jobs were able to access data over the wide area (60k files/day, O(100TB)/day)
- Production system for Run2 enabling
  - Interactive access
  - Fail over protection
  - The ability to share production workflows





Successes in Connectivity

- Aggregate bandwidth > 2GB/s



Integration of network and storage

- Data Federation is not a content delivery network (CDN)
  - It has only basic network awareness
  - Integration of more intelligent caching and intermediate storage
- We see interesting opportunities in development of advanced data management that begins to close the gap between data federation and CDN
  - End goal would be to care a lot less about the actual location of the data
- Looking forward we would like to investigate Named Data Networks where more of the data management is integrated with the network itself