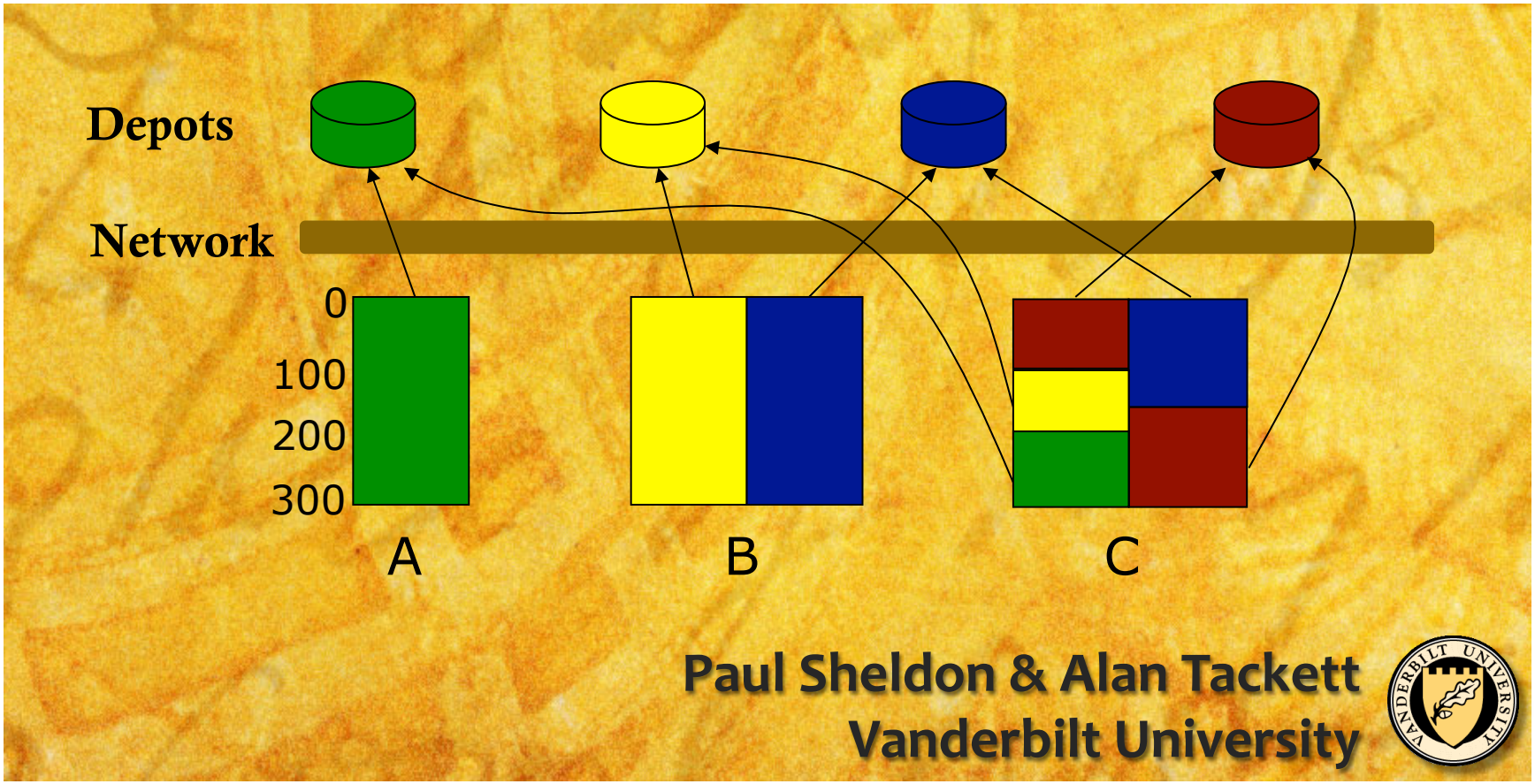
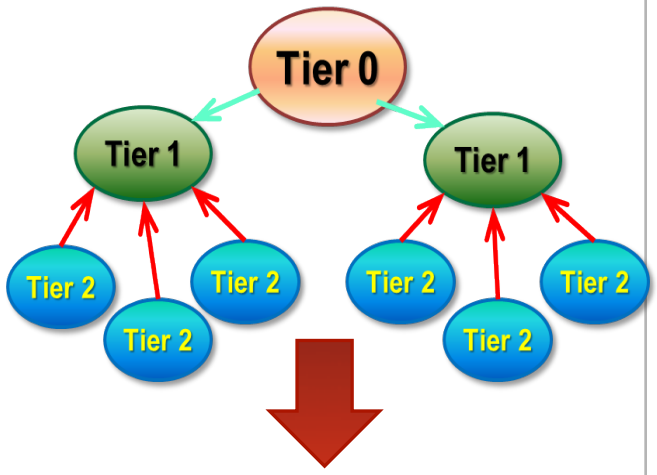
LHC and LSST Use Cases



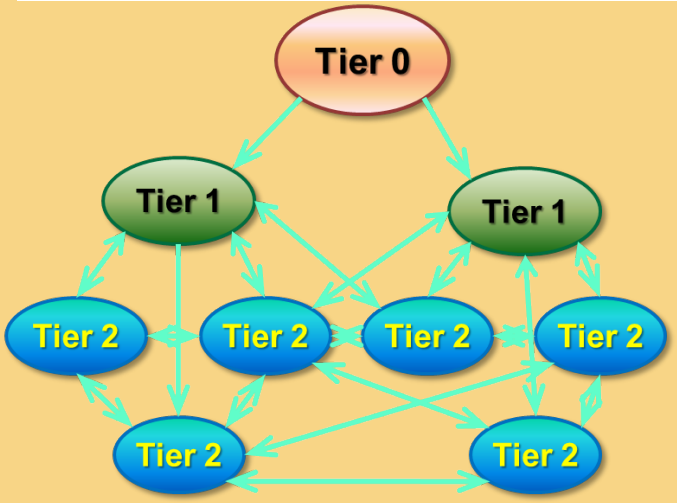
Paul Sheldon & Alan Tackett
Vanderbilt University

LHC Data Movement and Placement

- Model must evolve
- Was: Hierarchical, strategic pre-placement of data sets…



- **Meshed data flows:** Any site can use any other site to get data
- **Dynamic data caching:** Analysis sites will pull datasets from other sites "on demand"
- **Remote data access:** Jobs executing locally access data at remote sites in quasi-real time



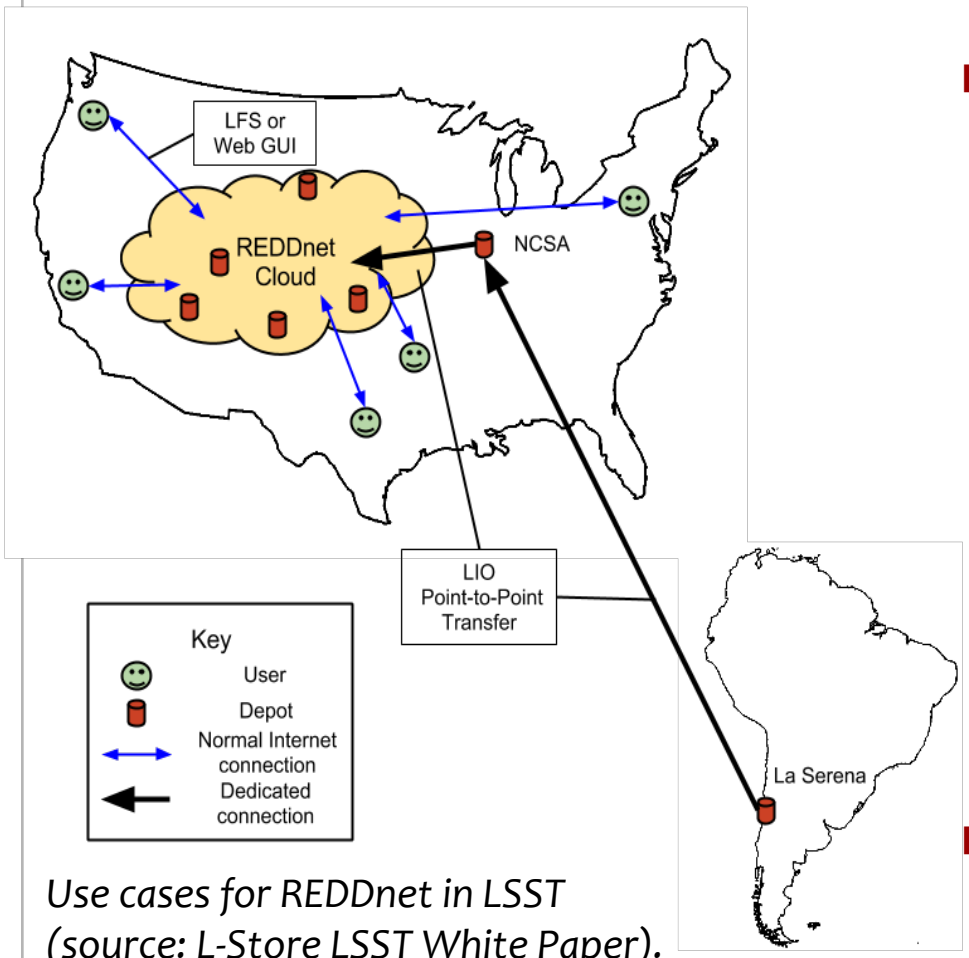
Fisk/Shank View of the Future

- Ian Fisk (CMS) and Jim Shank (ATLAS), HEP Snowmass Planning Exercise: *CMS Data Model must evolve*
- 100 Fold Increase in Data by 2020.
  - "We cannot afford another factor of 100 increase in storage, so we need to find ways of being more efficient in the use of space."
  - "We need to identify technology that allows a system to distribute and serve the data much more flexibly and dynamically."
  - "Data Management resources that deliver data on demand… content delivery networks."
- Networks are cheaper than storage
  - "Data delivery systems give a lot of flexibility… but they put strong requirements on networking." Tier 2's will need 100 Gb/s

Use Cases for 100G

- Brazilian Tier N centers are closely aligned with those in the US. Data Intensive CDN will need 100 Gbs OpenWave/AmLight.
- LHC overall will need international 100G networks.
  - Did not mention, but could have: projects such as DYNES and ANSE are important first steps for LHC use case.
  - Working to better integrate REDDnet technology (more on REDDnet later) into CMS "CDN"

LSST Use Case

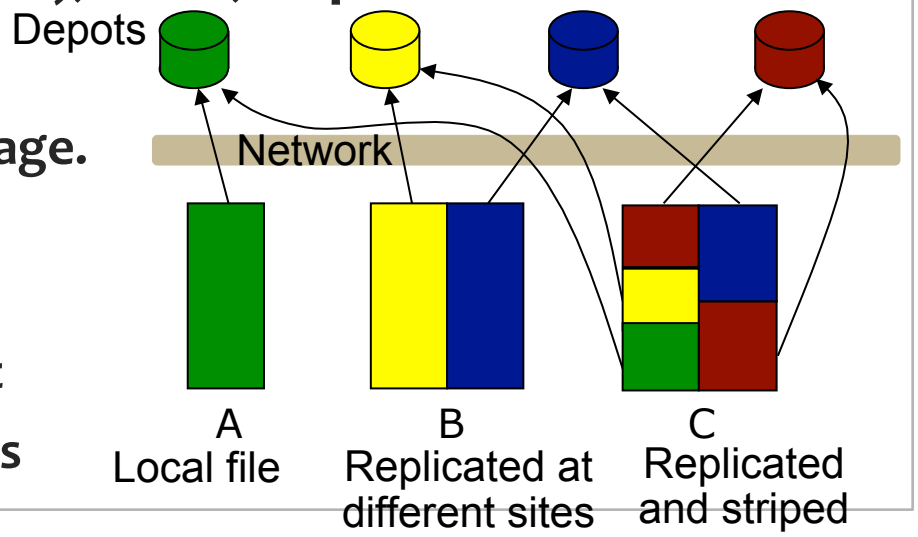


Use cases for REDDnet in LSST (source: L-Store LSST White Paper).

- LSST generates 15 TB each night. Each image must be moved from Chile to NCSA and analyzed within 60 seconds(?): was something interesting observed?
  - TCP SDF: Transport Control to Achieve Smooth Data Flow at a Stable Throughput – Chris Lu at Vanderbilt
- Proposed use of REDDnet technology for LSST project

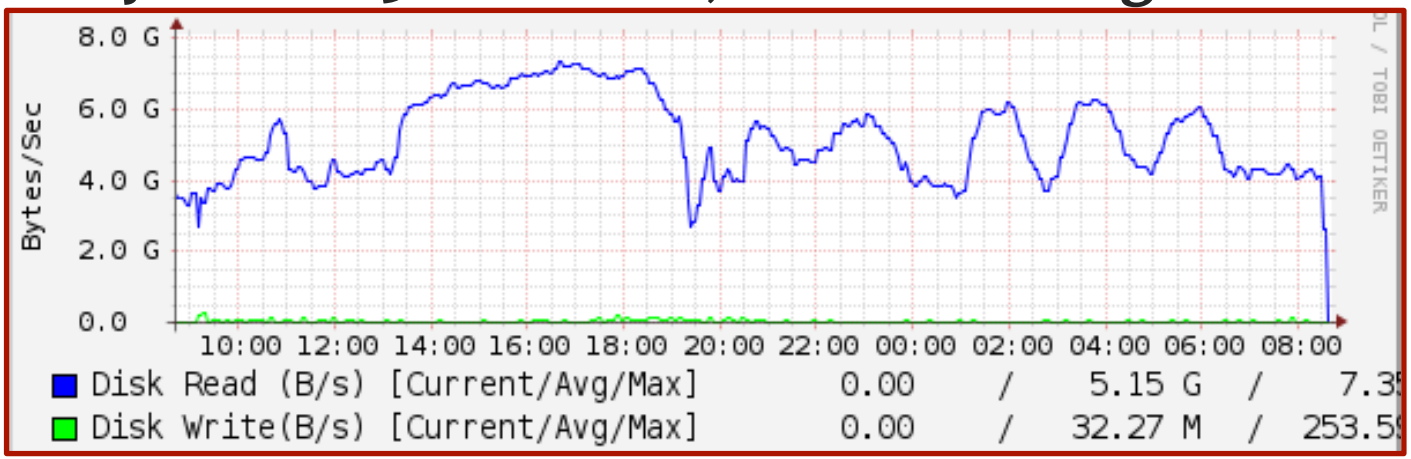
REDDnet / DLT

- NSF Funding
- *Collaborators:* Indiana, Tennessee, Stephen F. Austin, Texas Tech, Vanderbilt
- *Users:* Texas Tech Library, Vanderbilt CMS Tier-2 Computing Center, Vanderbilt TV News Archive
- *Proposed:* AmericaView (USGS), LSST, expanded CMS…
- REDDnet is:
  - IBP Depot: the basic unit of storage. Usually a server full of disks.
  - Distributed: Network of Depots
  - Flexible: Hardware Independent
  - Scalable: Heterogeneous Depots



… REDDnet / DLT

- REDDnet has:
  - High Capacity Storage
  - WAN Networking
  - Separately managed Raw data and Metadata
  - Multiple Access methods (interfaces) for different use cases
- Working to expand usage within CMS, proposed use in LSST.
- Currently curates 3 PB of data, often under significant load.

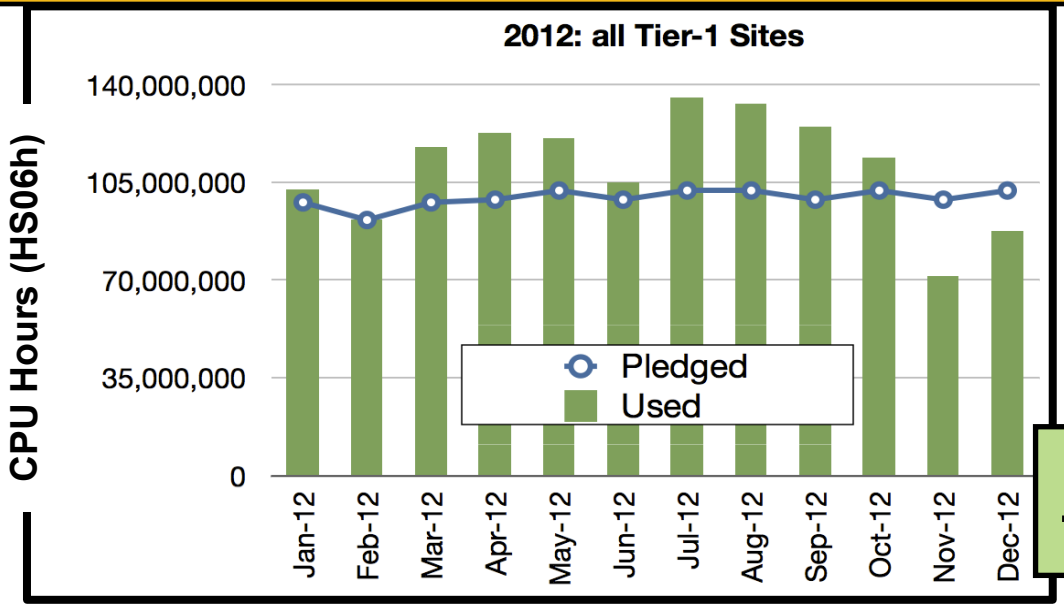


Backup Slides

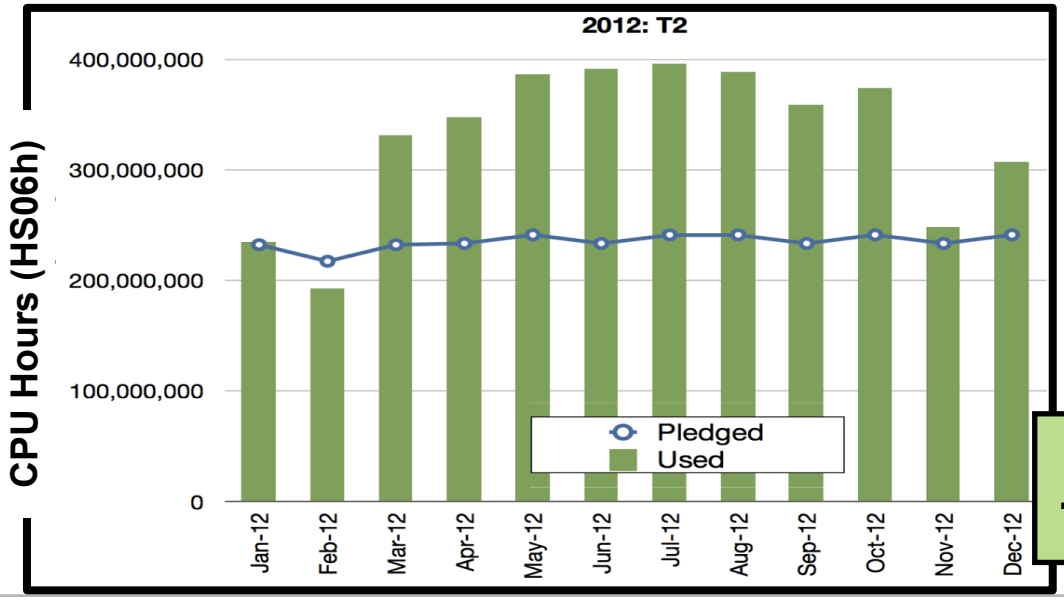
How is it Working?

200K – 300K LHC jobs launched/day

1 HS06h ~ 1 GFLOP hour
http://w3.hepix.org/benchmarks/doku.php/



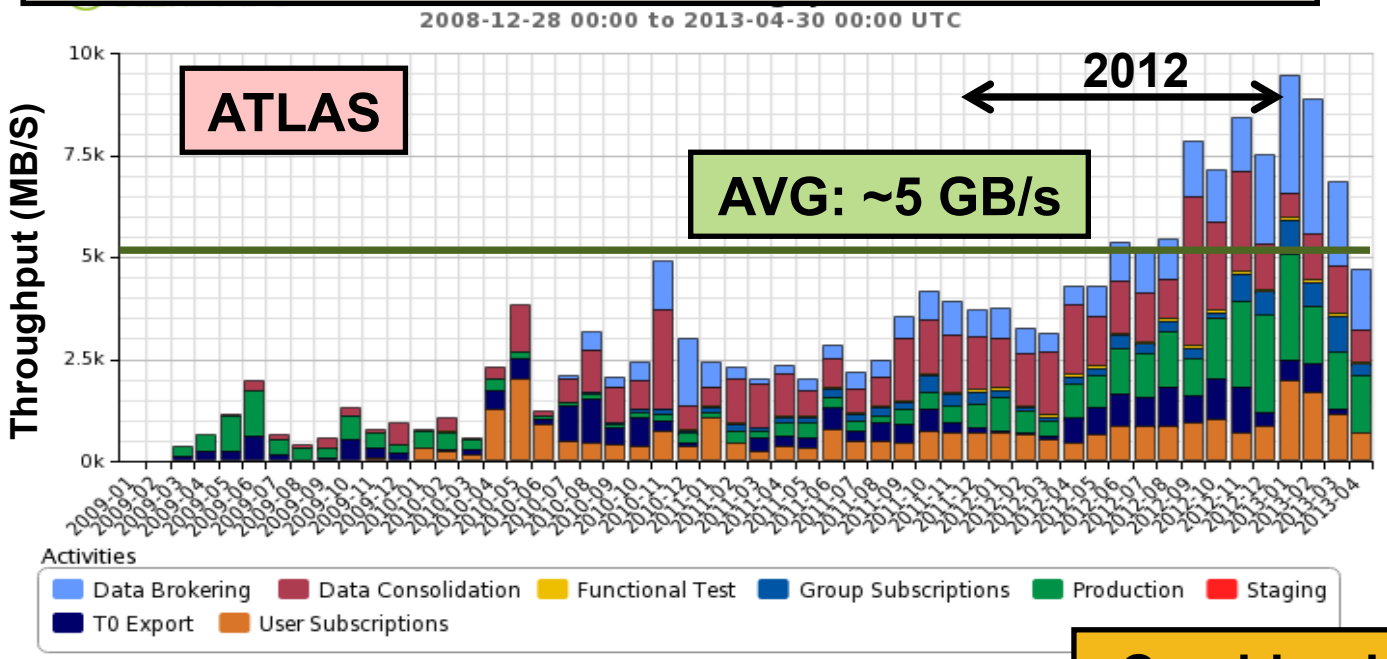
CMS Tier 1

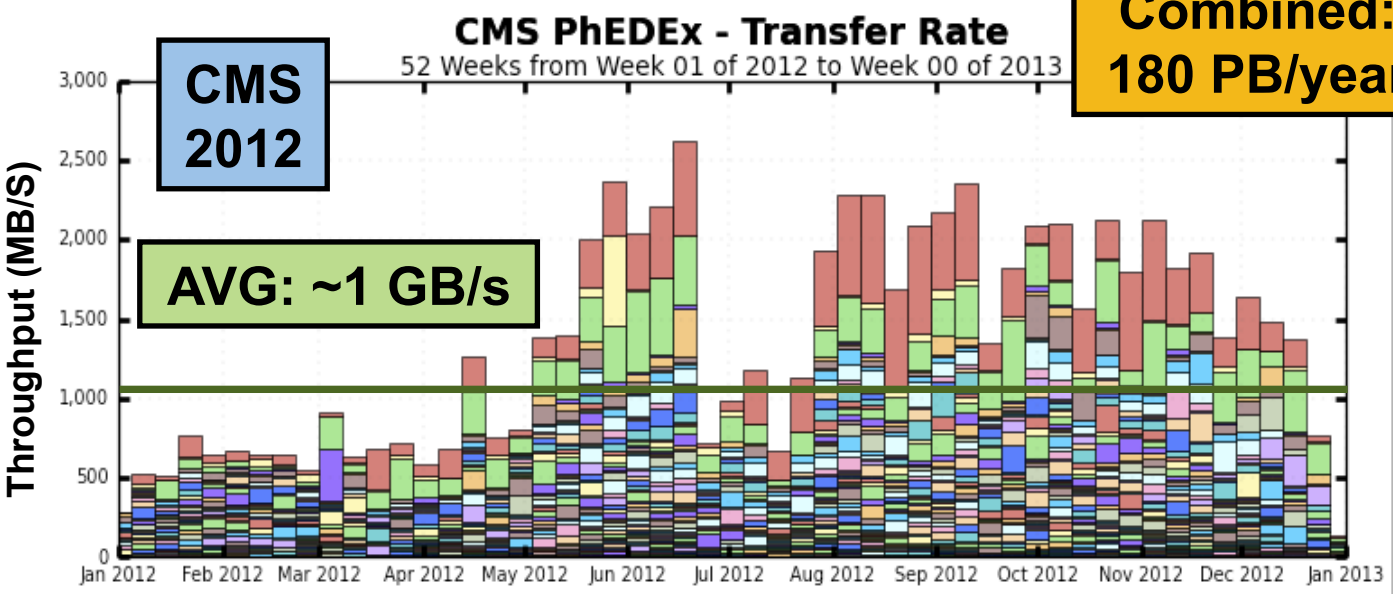


CMS Tier 2

How is it Working?

Aggregate Bandwidth on all Tier N Links

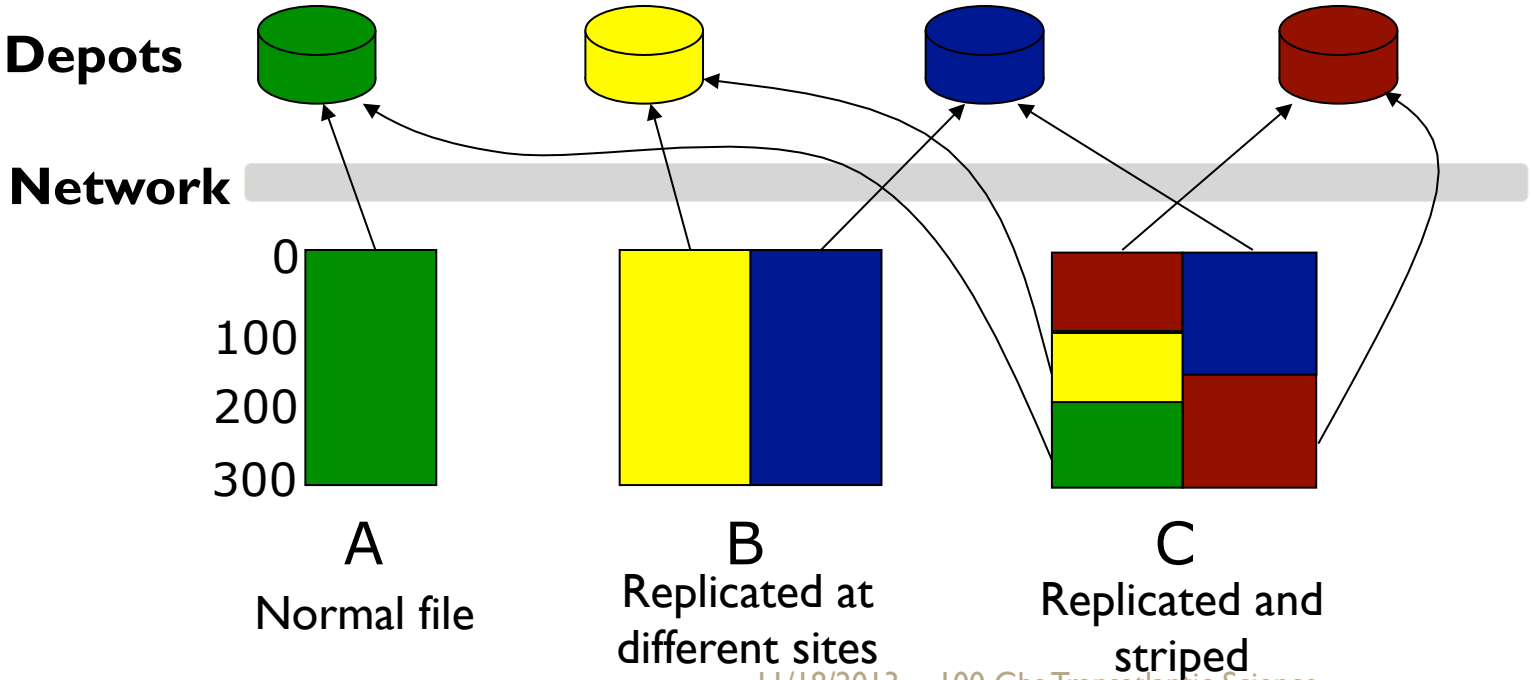




Combined: 180 PB/year

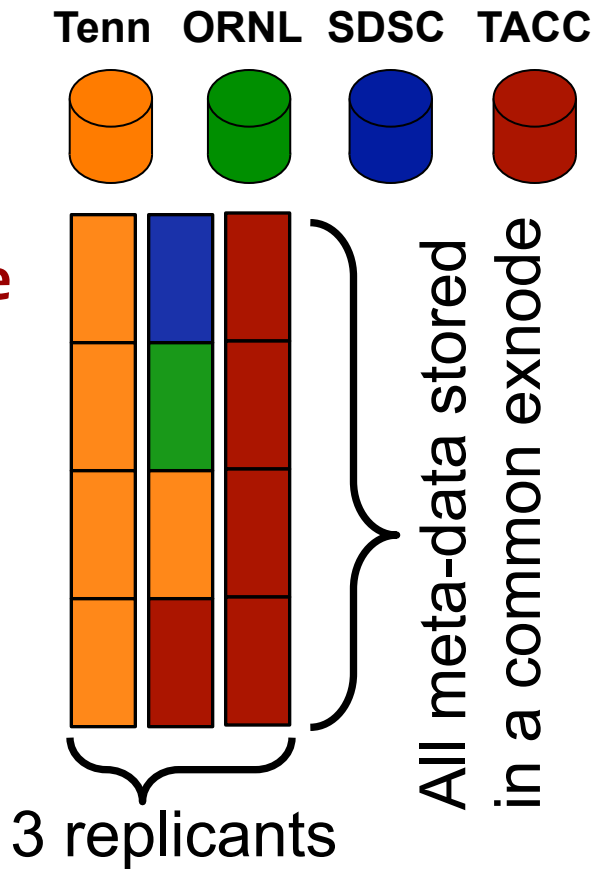
eXnodes: The "DNA" of REDDnet

- Analogous to an Unix inode and contains
  - List of allocations
  - File assembly instructions
  - Fault tolerance encoding scheme
  - Encryption keys



Core Elements: Replication/Movement

- Can replicate and/or migrate files automatically (set by policy)
- Replication info: **stored in same exnode** as the original file
- Example: 3 replicants across 4 depots
- If a depot (or its network) goes down, can still access the file
  - *User doesn't even have to know*
- Replication can provide ***proximity***



DYNES: Addressing Growing/ Evolving Network Needs

- NSF MRI-R2: DYnamic NEtwork System (DYNES)
- A nationwide cyber-instrument spanning ~40 US universities and ~14 Internet2 connectors
  - Extends Internet2/ESnet Dynamic Circuit tools (OSCARS, …)
- Who is it?
  - Project Team: Internet2, Caltech, Univ. of Michigan, and Vanderbilt Univ.
  - Community of regional networks and campuses
  - LHC, astrophysics community, OSG, WLCG, other virtual organizations
- What are the goals?
  - Support large, long-distance scientific data flows in the LHC, other data intensive science (LIGO, Virtual Observatory, LSST,…), and the broader scientific community
  - Build a distributed virtual instrument at sites of interest to the LHC but available to R&E community generally

DYNES Mission/Goals

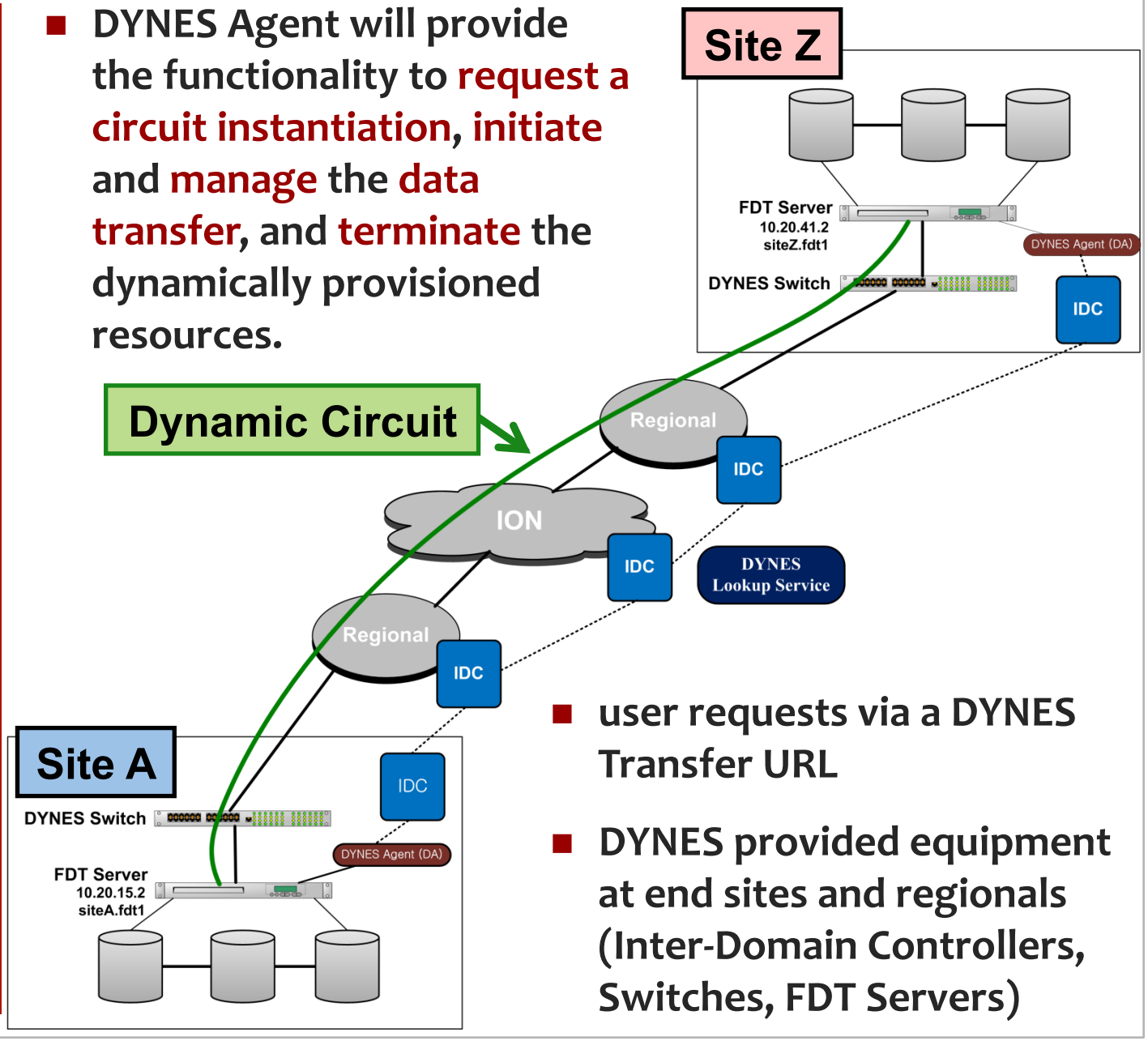
- Provide access to Internet2's dynamic circuit network in a manageable way, with fair-sharing
  - Based on a 'hybrid' network architecture utilizing both routed and circuit based paths.
- Enable such circuits w/ bandwidth guarantees across multiple US network domains and to Europe
  - Will require scheduling services at some stage
- Build a community with high throughput capability using standardized, common methods

Why Dynamic Circuits?

- **Internet2's and ESnet's dynamic circuits services provide:**
  - **Increased effective bandwidth capacity, and reliability of network access**, by isolating traffic from standard IP "many small flows" traffic
  - **Guaranteed bandwidth as a service** by building a system to automatically schedule and implement virtual circuits
  - **Improved ability of scientists to access network measurement data** for network segments end-to-end via perfSONAR monitoring
- **Static "nailed-up" circuits will not scale.**
- **GP network firewalls incompatible with enabling large-scale science network dataflows**
- **Implementing many high capacity ports on traditional routers would be expensive**

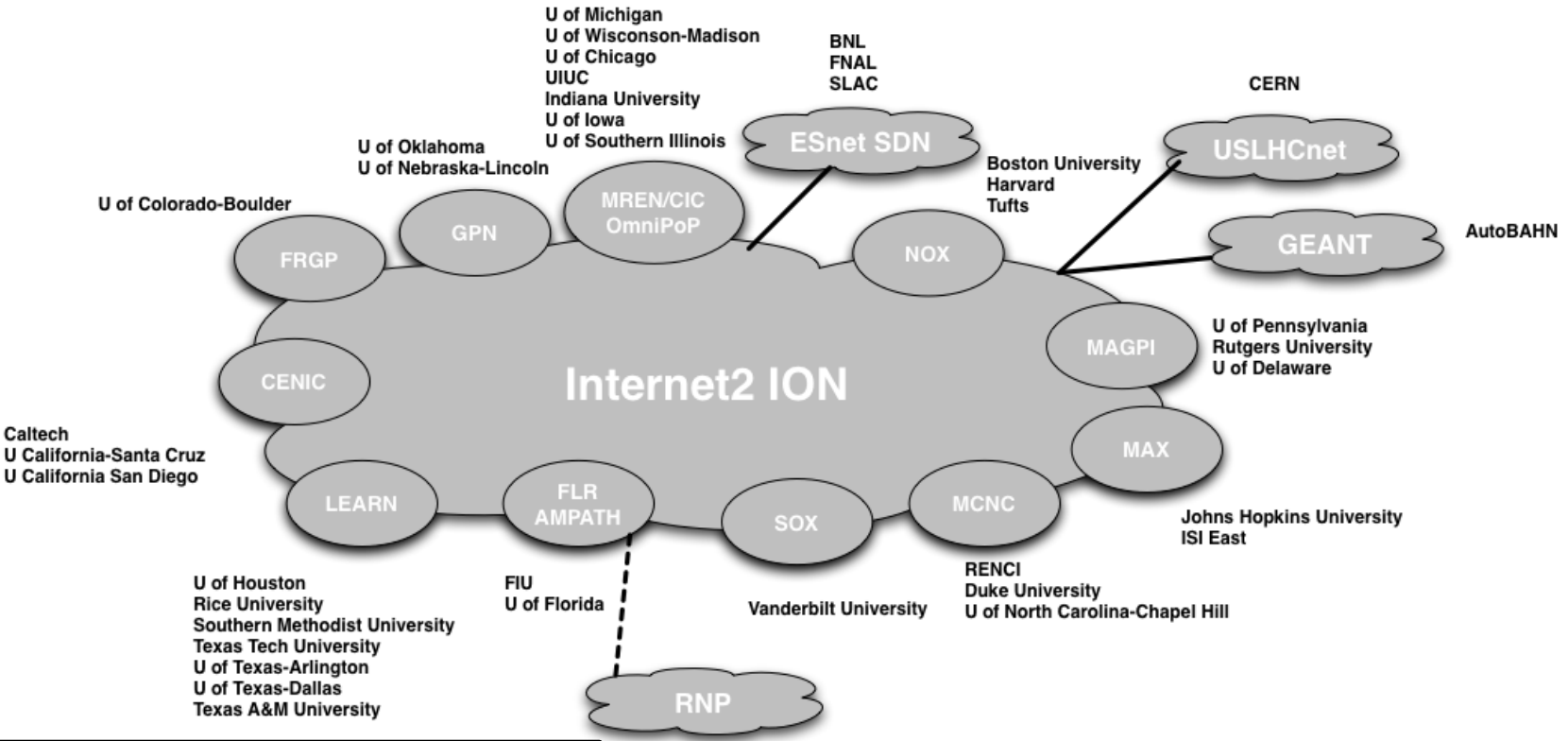
DYNES Dataflow

- DYNES Agent will provide the functionality to request a circuit instantiation, initiate and manage the data transfer, and terminate the dynamically provisioned resources.



- user requests via a DYNES Transfer URL
- DYNES provided equipment at end sites and regionals (Inter-Domain Controllers, Switches, FDT Servers)

DYNES Topology



Based on Applications Rcvd from Potential Sites

ANSE: Integrating DYNES into LHC Computing Frameworks

- NSF CC-NIE: Advanced Network Services for Expts (ANSE)
- Project Team: Caltech, Michigan, Texas-Arlington, & Vanderbilt
- Goal: Enable strategic workflow planning including network capacity as well as CPU and storage as a co-scheduled resource
- Path Forward:
  - Integrate advanced network-aware tools with the mainstream production workflows of ATLAS and CMS
  - Use in-depth monitoring and network provisioning
  - Complex workflows: a natural match and a challenge for SDN
- Exploit state of the art progress in high throughput long distance data transport, network monitoring and control

ANSE Methodology

- Use agile, managed bandwidth for tasks with levels of priority along with CPU and disk storage allocation.
  - define goals for time-to-completion, with reasonable chance of success
  - define metrics of success, such as the rate of work completion, with reasonable resource usage efficiency
  - define and achieve "consistent" workflow
- Dynamic circuits a natural match
- Process-Oriented Approach
  - Measure resource usage and job/task progress in real-time
  - If resource use or rate of progress is not as requested/planned, diagnose, analyze and decide if and when task re-planning is needed
- Classes of work: defined by resources required, estimated time to complete, priority, etc.