Possible Advances in Computing

Torre Wenaus, BNL

August 14, 2013

APS DPF Annual Meeting

UC Santa Cruz

BROOKHAVEN NATIONAL LABORATORY

a passion for discovery

Office of Science
U.S. DEPARTMENT OF ENERGY

My perspective…

- US ATLAS Physics Support & Computing Manager
- ATLAS Distributed Computing software development co-coordinator
- PanDA distributed processing and analysis system project co-leader & developer
- Remote participant in Snowmass

Outline

Drawing heavily on Snowmass but not an official summary

- Introduction
- Computing challenges at the physics frontiers
  - Highly abbreviated and condensed
- Computing advances: addressing the challenges
  - Networking
  - Data management
  - Processing
  - Distributed computing
  - Access and usability
- Conclusions

Snowmass Computing Frontier Study

Conveners: Lothar Bauerdick (FNAL), Steven Gottlieb (Indiana U)

The charge in brief:

- What are the **computational requirements for carrying out the experiments**
- What are the **computational requirements for theoretical computations and simulations**
- What **facility and software infrastructure** is needed to meet the requirements, and what **research investments** are needed in computing, storage, networking, application frameworks, algorithms, programming, etc.
- What are the **training requirements**

Subgroups for “user needs” **Computing needs of physics**

- CpF E1 Cosmic Frontier
  - Alex Szalay (Johns Hopkins), Andrew Connolly (U Washington)
- CpF E2 Energy Frontier
  - Ian Fisk (Fermilab), Jim Shank (Boston University)
- CpF E3 Intensity Frontier
  - Brian Rebel (Fermilab), Mayly Sanchez (Iowa State), Stephen Wolbers (Fermilab)
- CpF T1 Accelerator Science
  - Estelle Cormier (Tech-X), Panagiotis Spentzouris (FNAL); Chan Joshi (UCLA)
- CpF T2 Astrophysics and Cosmology
  - Salman Habib (Chicago), Anthony Mezzacappa (ORNL); George Fuller (UCSD)
- CpF T3 Lattice Field Theory
  - Thomas Blum (UConn), Ruth Van de Water (FNAL); Don Holmgren (FNAL)
- CpF T4 Perturbative QCD
  - Stefan Hoeche (SLAC), Laura Reina (FSU); Markus Wobisch (Louisiana Tech)

Subgroups for “infrastructure” **Technical capability mapping to needs**

- CpF I2 Distributed Computing and Facility Infrastructures
  - Ken Bloom (U.Nebraska/Lincoln), Sudip Dosanjh (LBL), Richard Gerber (LBL)
- CpF I3 Networking
  - Gregory Bell (LBNL), Michael Ernst (BNL)
- CpF I4 Software Development, Personnel, Training
  - David Brown (LBL), Peter Elmer (Princeton U.); Ruth Pordes (Fermilab)
- CpF I5 Data Management and Storage
  - Michelle Butler (NCSA), Richard Mount (SLAC); Mike Hildreth (Notre Dame U.)

… and participation from the wider community

Their work is reflected in this talk… thanks to all

Snowmass Computing Frontier Study

Conveners: Lothar Bauerdick (FNAL), Steven Gottlieb (Indiana U)

Tue 30/7

Computing Frontier: E1 Cosmic	Alex SZALAY et al.	Blegen 415, University of Minnesota	08:00 - 08:30
Computing Frontier: E2 Energy	Dr. Ian FISK et al.	Blegen 415, University of Minnesota	08:30 - 09:00
Computing Frontier: T1 Accelerator Science	Dr. Panagiotis SPENTZOURIS et al.	Blegen 415, University of Minnesota	09:00 - 09:30
Computing Frontier: T3 Lattice Field Theory	Dr. Ruth VAN DE WATER et al.	Blegen 415, University of Minnesota	09:30 - 10:00
Computing Frontier: T2 Astrophysics and Cosmology	Salman HABIB et al.	Blegen 415, University of Minnesota	10:00 - 10:30
Computing Frontier: E3 Intensity	Dr. Brian REBEL et al.	Blegen 415, University of Minnesota	10:30 - 11:00
Computing Frontier: T4 Perturbative QCD	Mr. Stefan HOECHE et al.	Blegen 415, University of Minnesota	11:00 - 11:30
Computing Frontier: Discussion and Wrap Up		Blegen 415, University of Minnesota	11:30 - 12:00

Wed 31/7

Computing Frontier: I2 Distributed Computing and Facility Infrastructures	Ken BLOOM et al.	Blegen 415, University of Minnesota	08:00 - 08:45
Computing Frontier: I3 Networking	Michael ERNST et al.	Blegen 415, University of Minnesota	08:45 - 09:30
Computing Frontier: I4 Software Development, Personnel, Training	David BROWN et al.	Blegen 415, University of Minnesota	09:30 - 10:15
Computing Frontier: I5 Data Management and Storage	Richard MOUNT et al.	Blegen 415, University of Minnesota	10:15 - 11:00
Computing Frontier: Larry Price, Department of Energy: Agency View of the Computing Frontier		Blegen 415, University of Minnesota	11:00 - 11:30
Computing Frontier: Discussion and Wrap Up		Blegen 415, University of Minnesota	11:30 - 12:00

Subgroups for "user needs" — Computing needs of physics

- CpF E1 Cosmic Frontier
  - Alex Szalay (Johns Hopkins), Andrew Connolly (U Washington)
- CpF E2 Energy Frontier
  - Ian Fisk (Fermilab), Jim Shank (Boston University)
- CpF E3 Intensity Frontier
  - Brian Rebel (Fermilab), Mayly Sanchez (Iowa State), Stephen Wolbers (Fermilab)
- CpF T1 Accelerator Science
  - Estelle Cormier (Tech-X), Panagiotis Spentzouris (FNAL); Chan Joshi (UCLA)
- CpF T2 Astrophysics and Cosmology
  - Salman Habib (Chicago), Anthony Mezzacappa (ORNL); George Fuller (UCSD)
- CpF T3 Lattice Field Theory
  - Thomas Blum (UConn), Ruth Van de Water (FNAL); Don Holmgren (FNAL)
- CpF T4 Perturbative QCD
  - Stefan Hoeche (SLAC), Laura Reina (FSU); Markus Wobisch (Louisiana Tech)

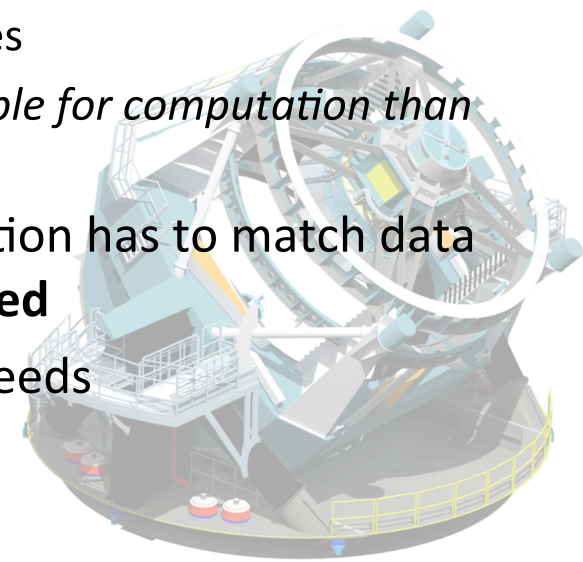
Subgroups for "infrastructure" — Technical capability mapping to needs

- CpF I2 Distributed Computing and Facility Infrastructures
  - Ken Bloom (U.Nebraska/Lincoln), Sudip Dosanjh (LBL), Richard Gerber (LBL)
- CpF I3 Networking
  - Gregory Bell (LBNL), Michael Ernst (BNL)
- CpF I4 Software Development, Personnel, Training
  - David Brown (LBL), Peter Elmer (Princeton U.); Ruth Pordes (Fermilab)
- CpF I5 Data Management and Storage
  - Michelle Butler (NCSA), Richard Mount (SLAC); Mike Hildreth (Notre Dame U.)

Two ½-day Snowmass sessions + informal discussions
Reports in progress: https://github.com/SnowMassComputing

Computing Challenges at the Physics Frontiers

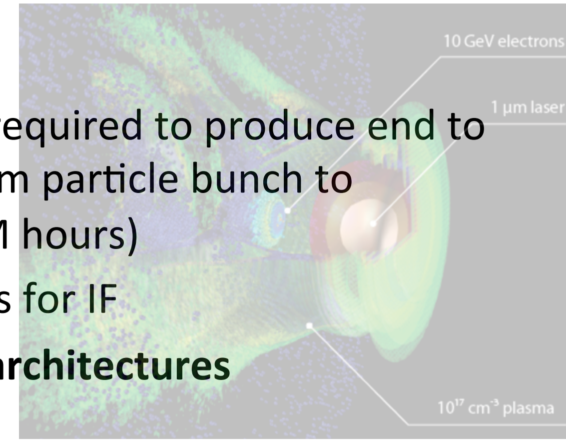
- **Cosmic Frontier**
  - **Sky survey, radio telescope instruments have large scale computing needs, comparable with Energy Frontier**
  - Growing detectors, data volumes, complexity, simulation needs
  - >1 PB data today, 50 PB total in 10 yrs, 400 PB/yr in 10-20 yrs
  - Already today, simulation processing is a **data-intensive computing** challenge
    - Data management tools, large scale databases, data analytics tools
    - Already users of innovative database technologies
    - *National resources are much more readily available for computation than for data management*
  - Growing data- and compute-intensive computation has to match data rates; **new distributed computing models needed**
  - Growing **data analytics, sustainable software** needs
  - **Data preservation and archiving** has to scale



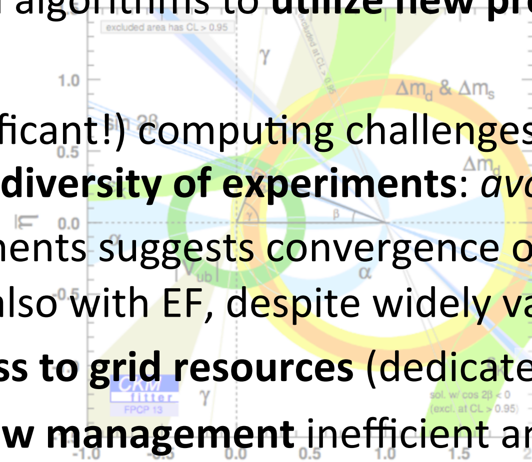
Computing Challenges at the Physics Frontiers

- **Energy Frontier**
  - LHC trigger rates and event complexity increasing through next 10+ years of machine and detector upgrades. **~15 PB LHC raw data now; ~130 PB in 2021**
  - Present total ATLAS dataset is 140 PB (~70 PB disk) – must be more efficient by 2021, **cannot afford 100x more storage**
  - Programs discussed at Snowmass all have potential for **10-fold increases** in trigger rate and complexity
  - **Storage is largest cost**: ATLAS spends ~1.6x more money on disk than on CPU
  - Cost constrains the data rate, driving choices on triggers and analyses
  - **Write a fraction to tape only**, pending a physics case to analyze?
    - a penny puts 1000 CMS events on tape, and disk price decline is slowing
    - *Requires efficient large scale tape based processing*
  - To control processing cost, **must track Moore's law effectively**
    - *Adapting to new processors is much more challenging than in the past*
  - Use **diverse resources** to maximize throughput: supercomputers, clouds, opportunistic

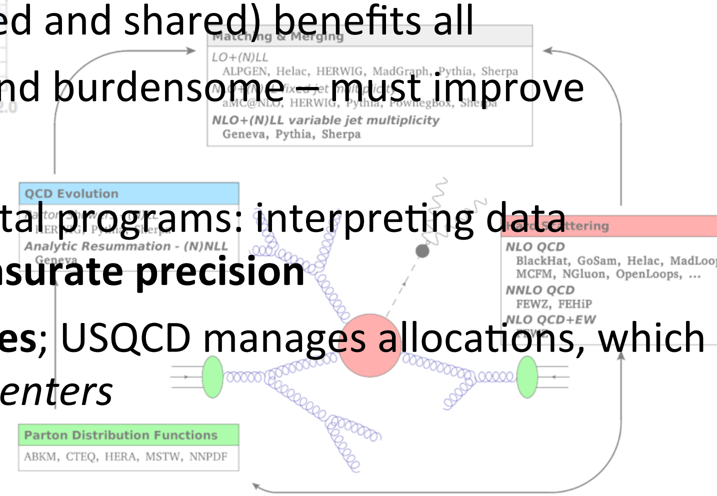
Computing Challenges at the Physics Frontiers



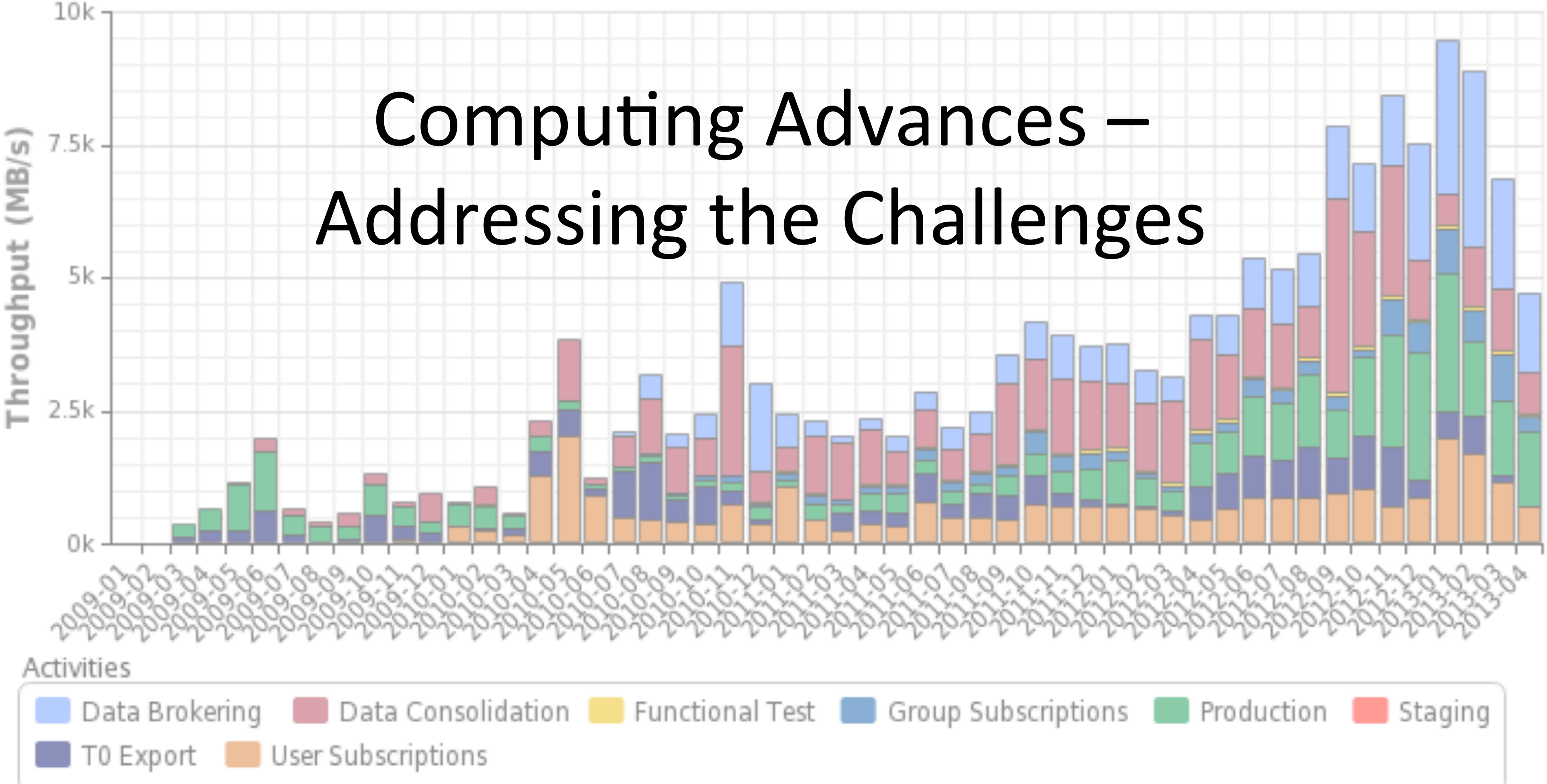
- **Accelerator Science**
  - **Petascale computing** (supercomputers in many cases) required to produce end to end designs across a wide range of modeling scales, from particle bunch to accelerator complex; **10-100x** current allocations (140M hours)
  - Necessary for achieving high gradients for EF, low losses for IF
  - R&D on advanced algorithms to **utilize new processor architectures**
- **Intensity Frontier**
  - Smaller (still significant!) computing challenges (Belle II estimates LHC Run 2-like data rates) but a **large diversity of experiments**: *avoid duplication, re-invention*
  - Survey of experiments suggests convergence on a **common computing model**; a lot of commonality, also with EF, despite widely varying scale and workflow
  - Transparent **access to grid resources** (dedicated and shared) benefits all
  - **Data and workflow management** inefficient and burdensome – must improve



- **Non-Perturbative and Perturbative QCD**
  - pQCD, Lattice QCD crucial for EF, IF experimental programs: interpreting data requires **theoretical predictions with commensurate precision**
  - Continued reliance on **supercomputing facilities**; USQCD manages allocations, which are *among the largest at US supercomputing centers*
  - Benefit from new technologies such as **GPUs**



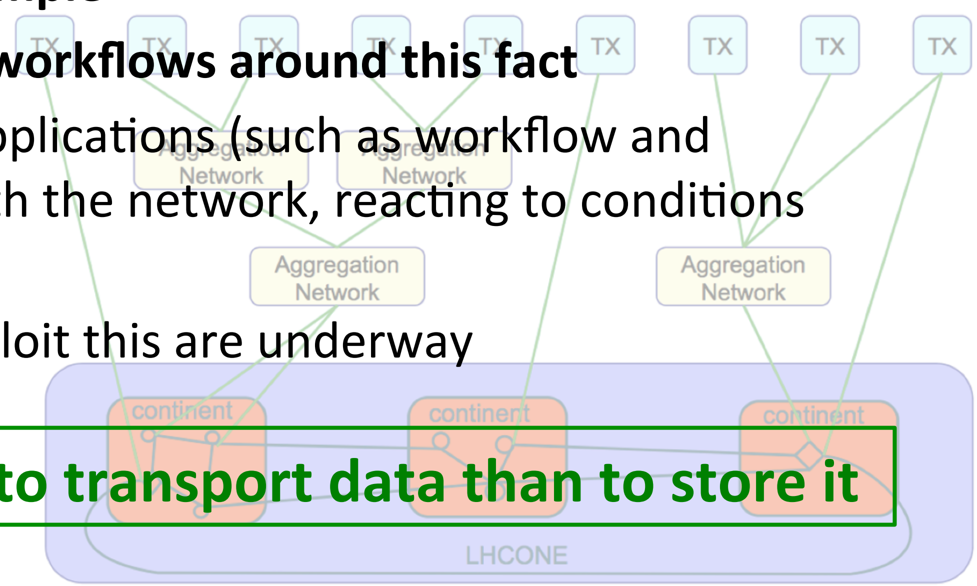
Computing Advances – Addressing the Challenges



ATLAS global transfers 1/2009 – 4/2013

Networking

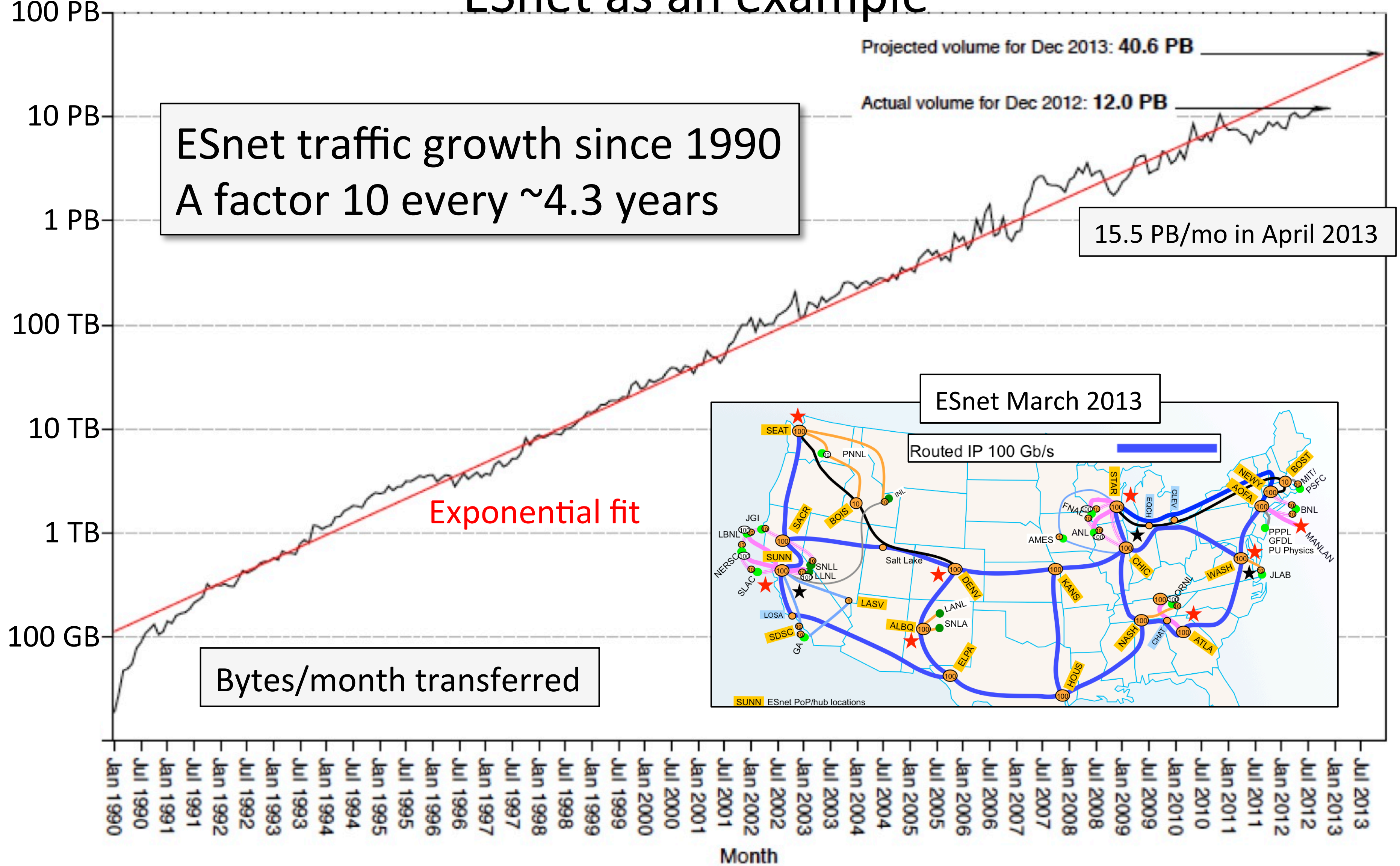
- HEP computing depends on reliable, high-bandwidth, feature-rich networks
- HEP was a pioneer in network-intensive science and international research networks, and continues to lead
  - Dedicated networks optimized for massive data flows
  - US LHC now testing the first 100Gb transatlantic production link
  - Investment and innovation pays dividends
- **Making the most of the network translates to more science and lower computing costs, LHC a major example**
  - **Collaborations should design workflows around this fact**
- Next generation networks allow applications (such as workflow and dataflow managers) to interact with the network, reacting to conditions and proactively controlling it
  - HEP computing projects to exploit this are underway



In general it's much cheaper to transport data than to store it

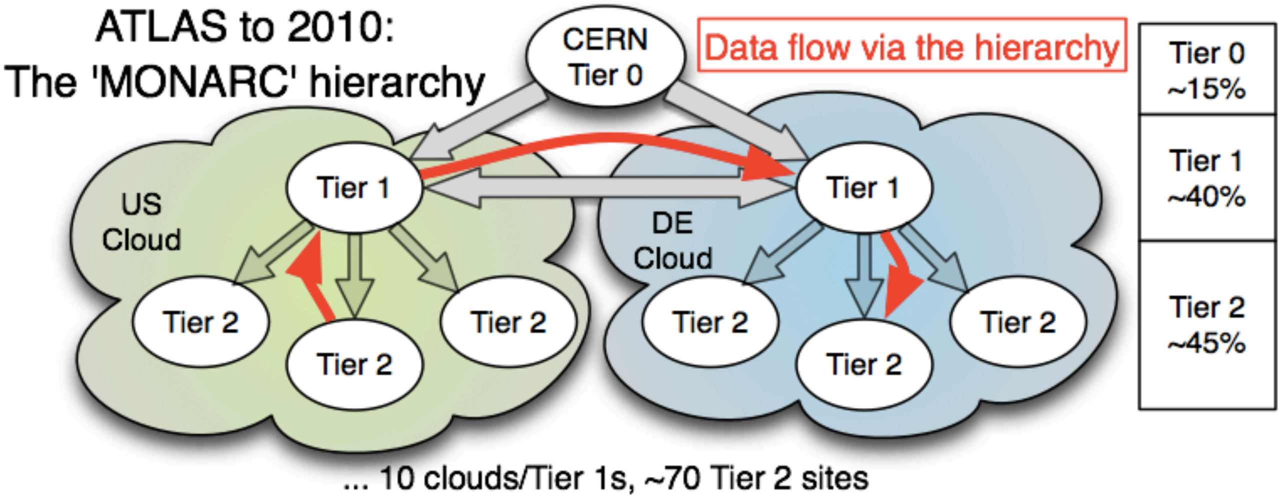
Networking growth has been dramatic

ESnet as an example



Networking has been a critical enabler for LHC computing

Computing Model Evolution in ATLAS



Original model:

Static, strict hierarchy

Multi-hop data flows

Lesser demands on Tier 2 networking

Virtue of simplicity

Designed for <~2.5 Gb/s within the hierarchy

Today:

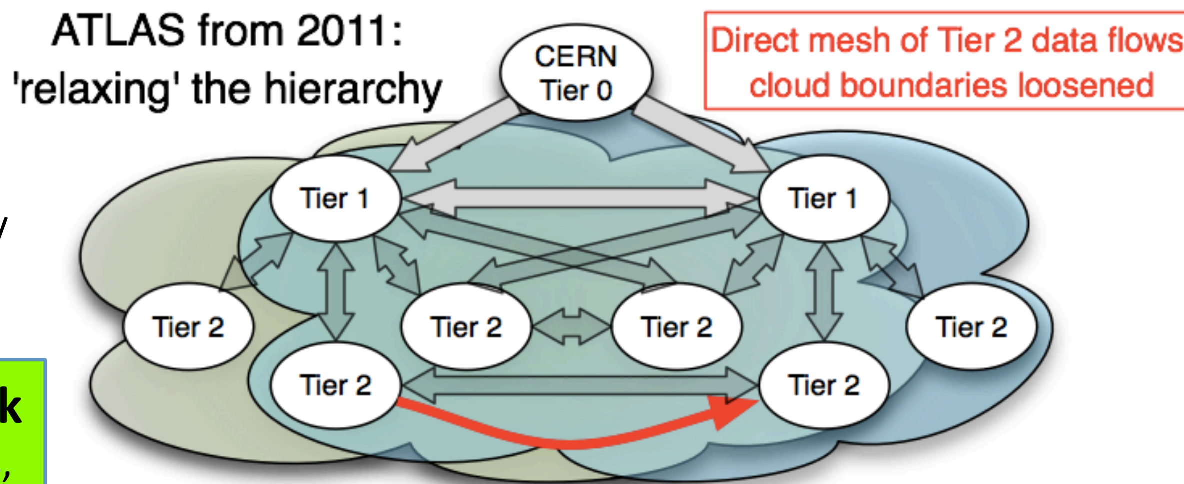
Bandwidths 10-100 Gb/s, not limited to the hierarchy

Flatter, mostly a mesh

Sites contribute according to capability

Greater flexibility and efficiency

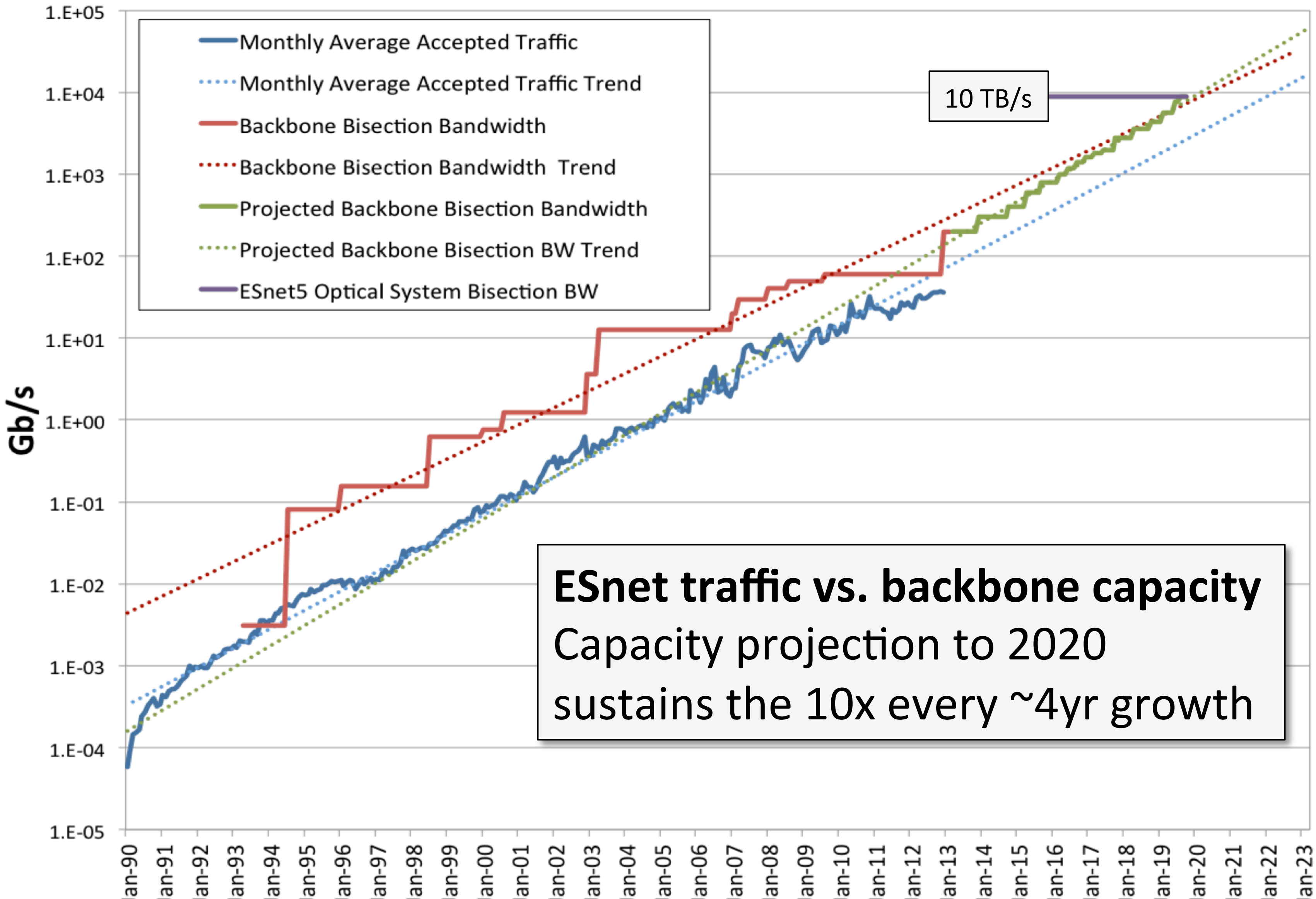
More fully utilize available resources



The enabler is the network

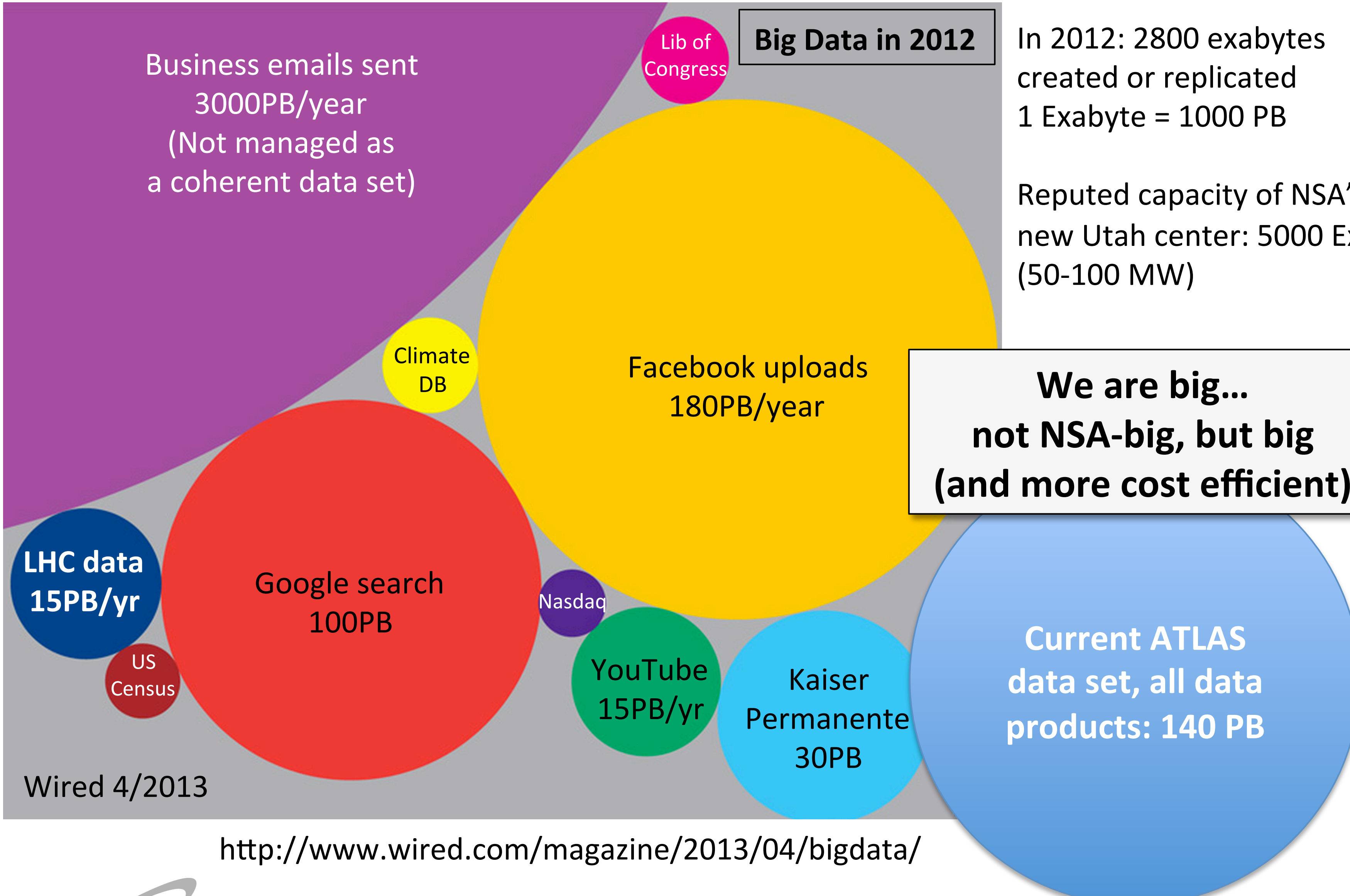
Excellent bandwidth, robustness, reliability, affordability

Planned capacity growth sustains the trend



Data Management

Where is HEP in Big Data Terms?



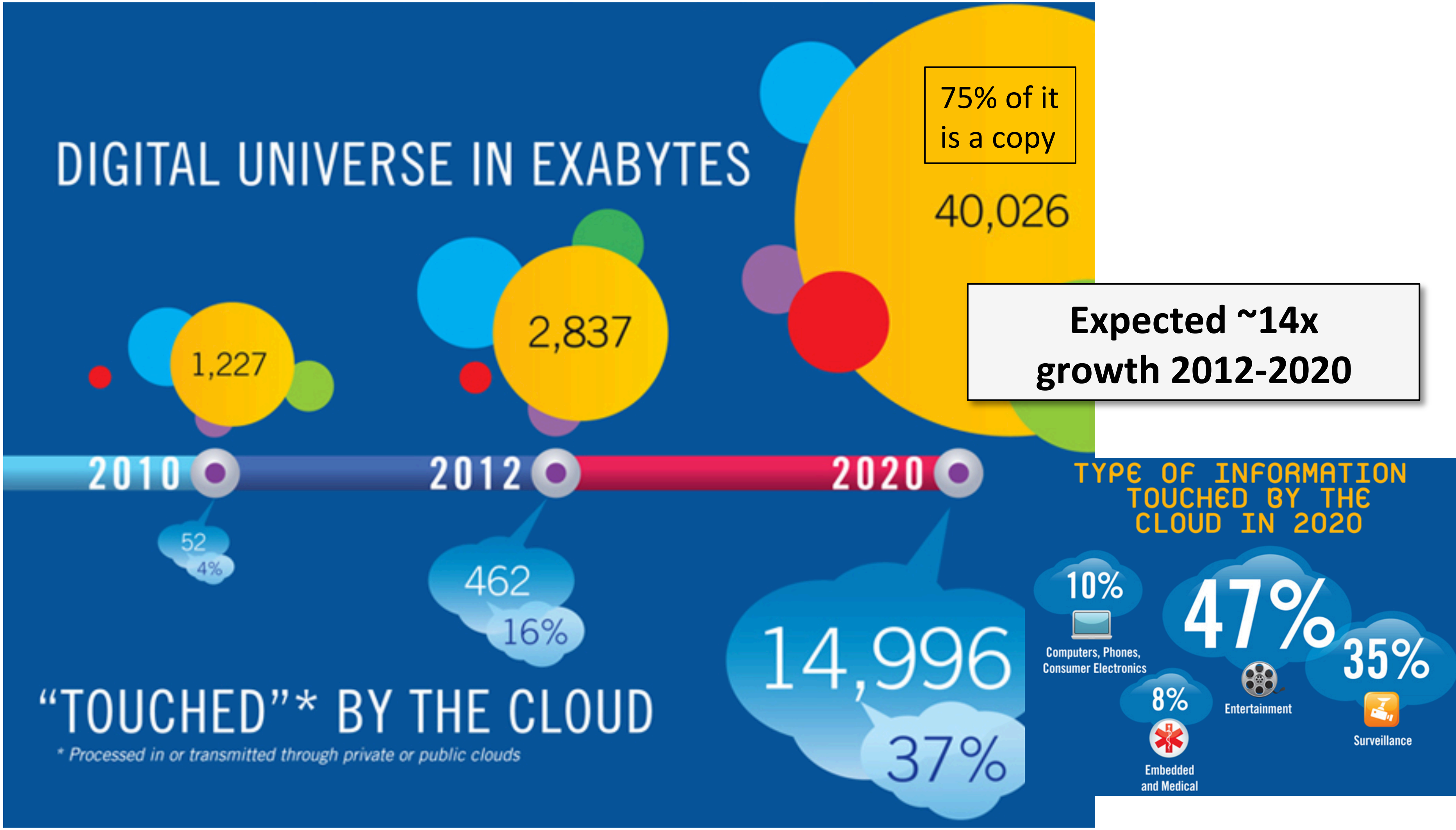
http://www.wired.com/magazine/2013/04/bigdata/

In 2012: 2800 exabytes created or replicated
1 Exabyte = 1000 PB

Reputed capacity of NSA's new Utah center: 5000 ExaB (50-100 MW)

We are big…
not NSA-big, but big
(and more cost efficient)

Predicted Data Growth to 2020



75% of it is a copy

Expected ~14x growth 2012-2020

IDC Digital Universe study 12/2012

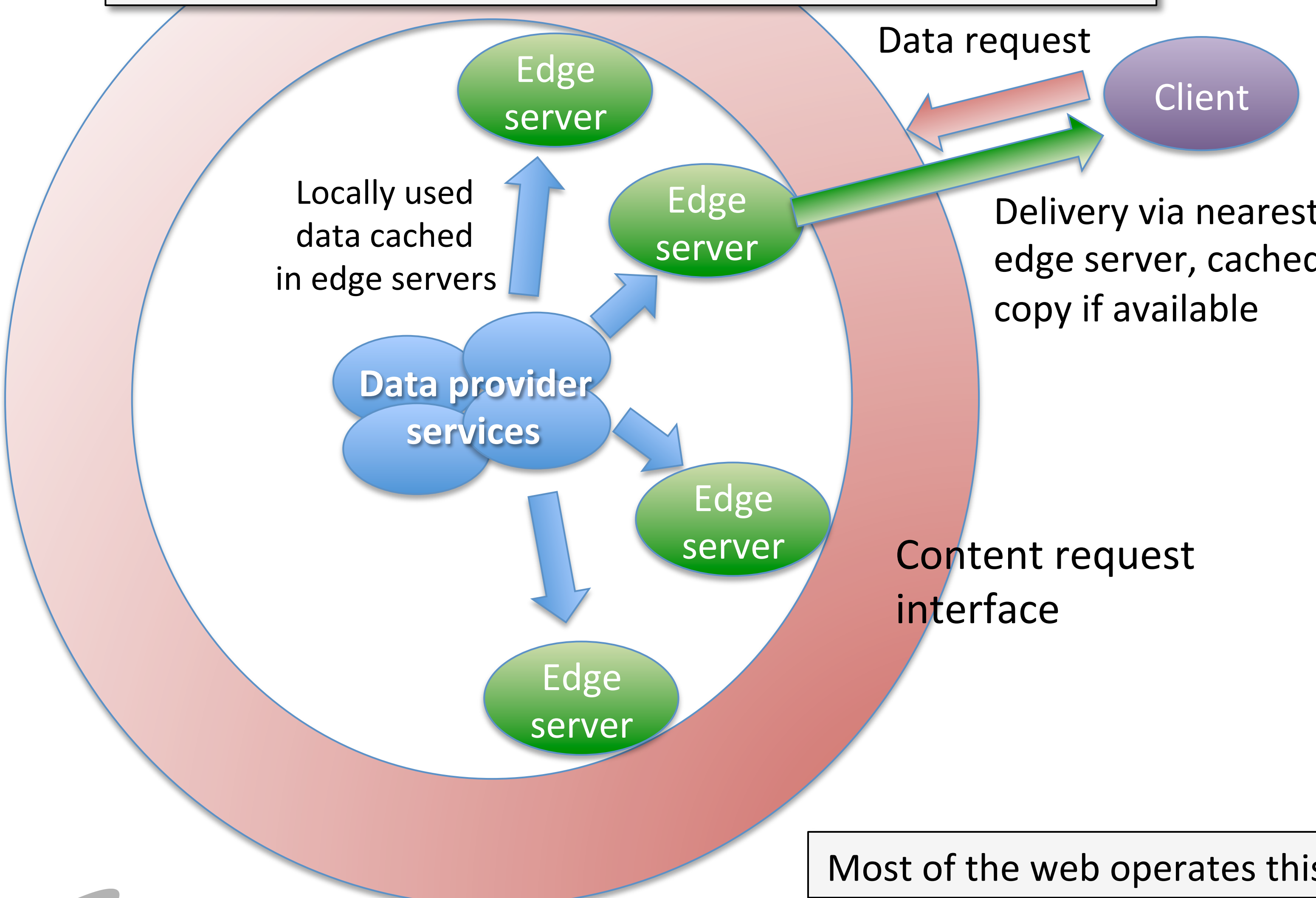
Evolution of HEP Data Management

- Data management challenges extend to all frontiers
  - Traditional data-intensive areas like EF and also data intensive computing at high performance computing (HPC, aka supercomputing) centers
- Surfing the Big Data wave helps greatly – e.g. powerful new technologies such as **Hadoop, large scale DBs** – but isn't enough, we can't afford to simply scale up
- We need **more efficient distributed data handling, lower disk storage demands, lower operational load** (storage is highly labor intensive for operations)
- (Aspire to) ***send only the data you need, only where you need it***
- One successful approach: building **intelligent data placement** into workflows
- Another in (HEP) development now reaching production: **Federated data storage**
  - Transparent distributed access, caching, robust against missing/lost data
- Further steps underway *transform our traditional approach altogether*
  - Dispensing with file-based management and delivering events – **event service**
  - Once you're delivering events you can trade off retrieval against on-demand generation – **virtual data**
- Industry has been at this approach for years, in **content delivery networks**

Everything relies on the network, including campus last mile

The Content Delivery Network Model

Content delivery network: deliver data quickly and efficiently by placing data of interest close to its clients



Most of the web operates this way

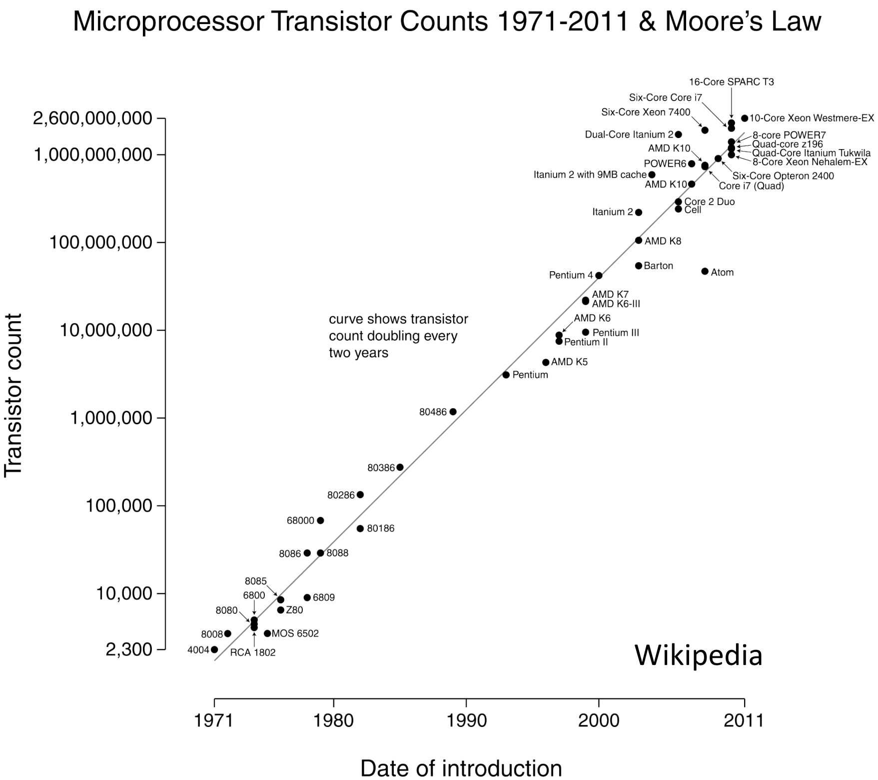
The Content Delivery Network Model

A growing number of HEP services are designed to operate broadly on the CDN model

Service	Implementation	In production
Frontier conditions DB	Central DB + web service cached by http proxies	~10 years (CDF, CMS, ATLAS, …)
CERNVM File System (CVMFS)	Central file repo + web service cached by http proxies and accessible as local file system	Few years (LHC expts, OSG, …)
Xrootd based federated distributed storage	Global namespace with local xrootd acting as edge service for the federated store	Xrootd 10+ years Federations ~now (CMS AAA, ATLAS FAX, …)
Event service	Requested events delivered to a client agnostic as to event origin (cache, remote file, on-demand generation)	First ATLAS implementation coming in next 6-12 months
Virtual data service	The ultimate event service backed by data provenance, regeneration infrastructure	Few years?

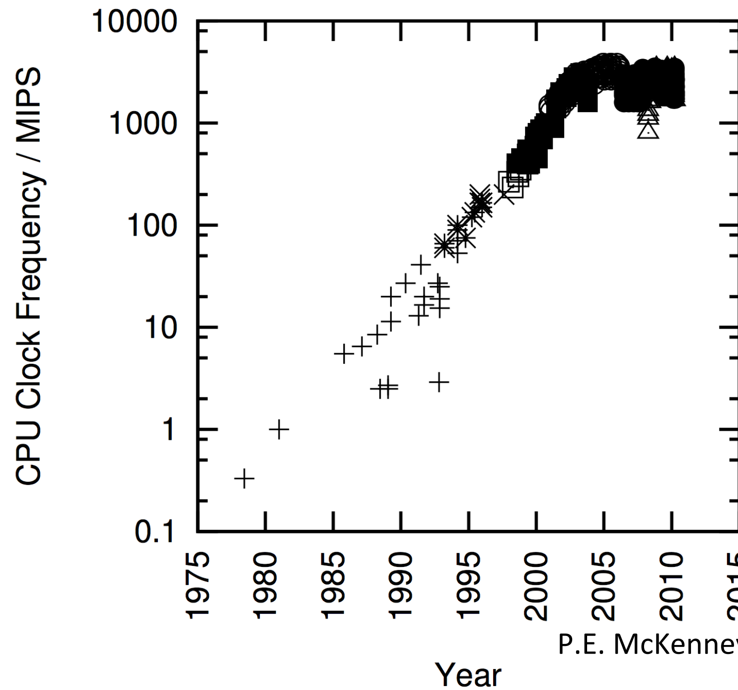
Processing

Transistor count growth is holding up…



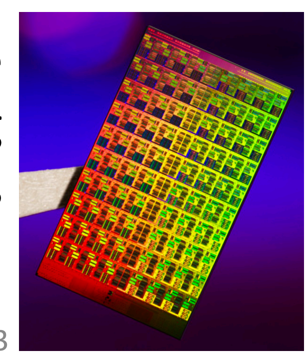
Wikipedia

…but clock speed growth died a heat death…



P.E. McKenney, IBM 2013

… replacing the free lunch of ever faster processors with the necessity of sustaining throughput growth by leveraging growth in core count, co-processors, concurrency features



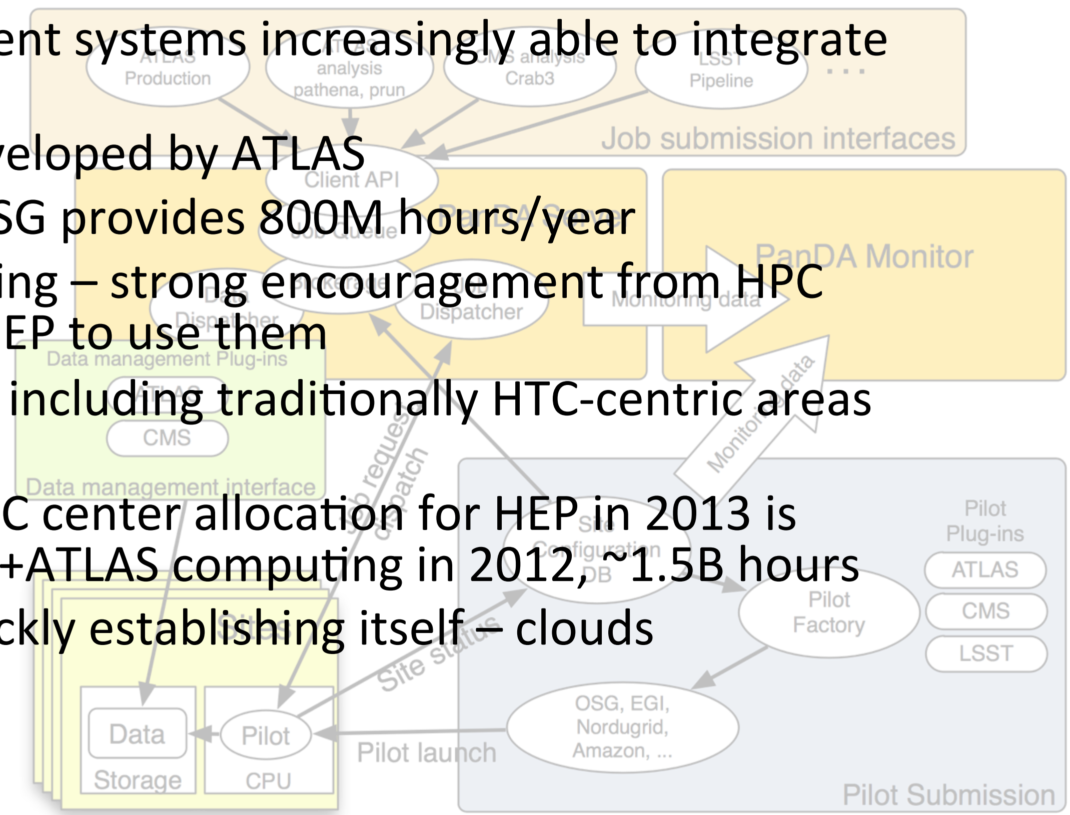
Intel 80-core chip

Adapting Software to New Processors

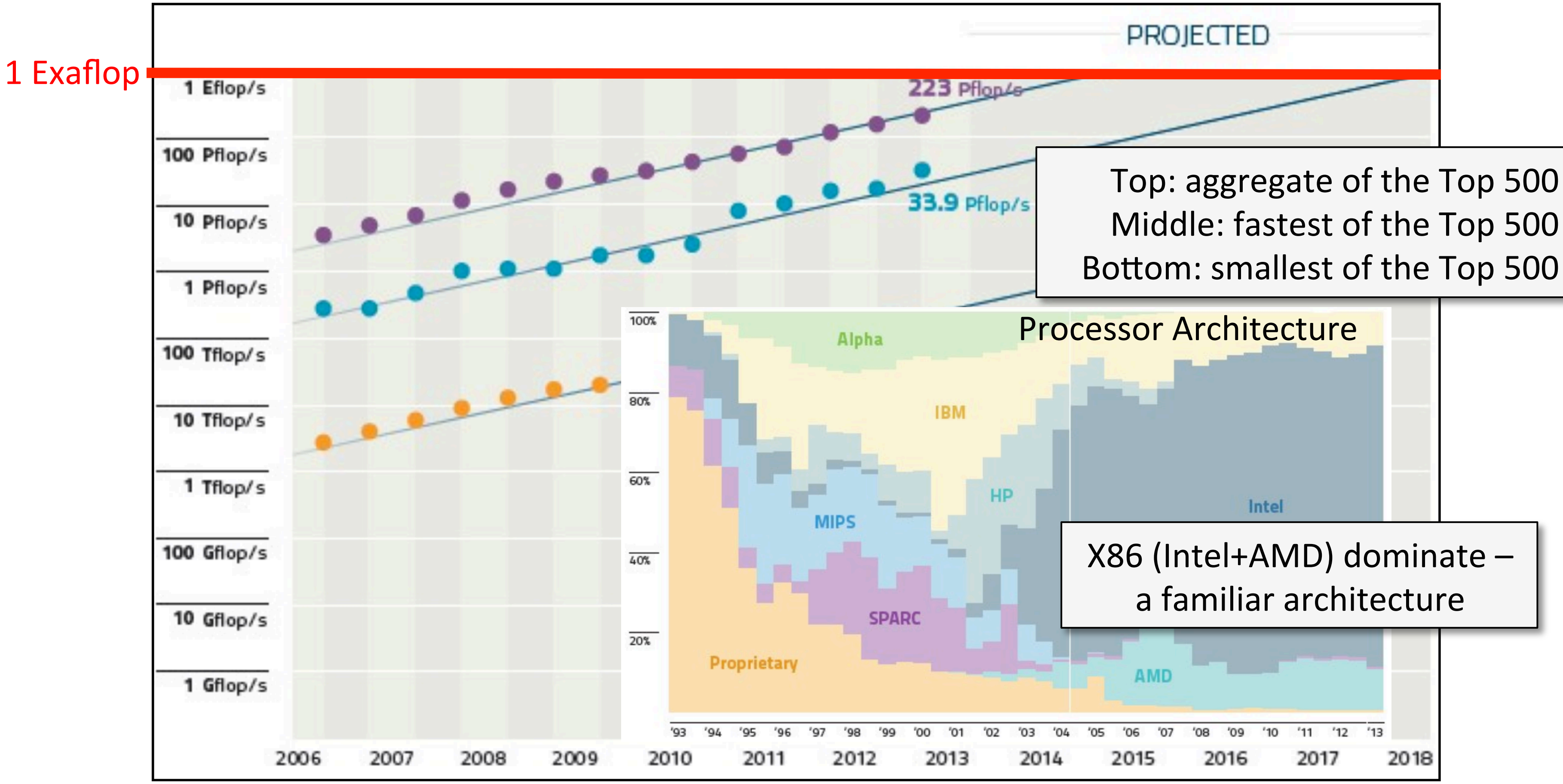
- High concurrency, modest memory/core, GPUs, … : the new environment
  - Multi-core now ➔ many-core soon ➔ finer grained parallelism needed
  - GPUs present challenges in programmability and data bottlenecks
  - *Many or most of our codes will require extensive overhauls*
- Impacts all large scale processing across the frontiers
- *The whole world faces it, which is good news* – tools, libraries, compilers are emerging that can help the migration from serial software
- Our common tools are being adapted, to the benefit of all
  - Geant4, ROOT, math libraries, reconstruction tools, …
- But it's a large effort on the part of scarce software experts
  - The software investment is necessary – *living with inefficient software is much more expensive*
  - Objective should be products widely usable in the HEP community
  - The open source world has the tools to facilitate code sharing
  - Software development should be engaged with the science & planning

Exploiting Diverse Resources

- HTC (data intensive, loosely coupled) and HPC (supercomputing) are both essential to HEP computing
- There is increasing convergence and overlap
  - HPC computing has a growing number of data intensive use cases
  - **More concurrency in software is making HTC-targeted applications more suited to HPC**
- Workflow and data management systems increasingly able to integrate diverse facilities
  - e.g. the PanDA system developed by ATLAS
- HTC the mainstay for EF, IF; OSG provides 800M hours/year
- HPC resources large and growing – strong encouragement from HPC centers and funders to *all* of HEP to use them
  - Community is responding, including traditionally HTC-centric areas such as LHC computing
  - US national DOE & NSF HPC center allocation for HEP in 2013 is comparable to global CMS+ATLAS computing in 2012, ~1.5B hours
- A relatively new entrant is quickly establishing itself – clouds



HPC growth remains rapid… an Exaflop system by 2020?



100+ PFlop plans: NERSC: NERSC-8 2015-16 ORNL: OLCF-4 2016-17

The Register 6/2013

Utility Computing: From Grids to Clouds

The Grid: 1998 and 2003 (2nd Ed.)

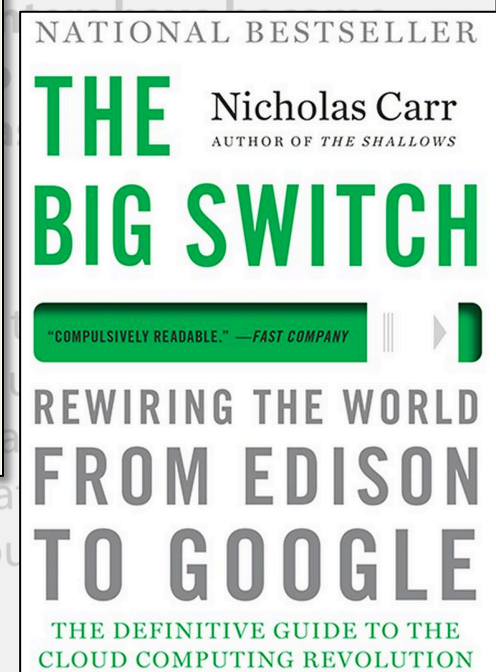
Grid is used by analogy with the electric power grid…
has had a dramatic impact on human capabilities…

The Grid has served us well, but industry yawned… and invented the cloud.
HEP sees merit in complementing grids with clouds

- **Users control their environment**
- **Sites see easier management, scaling**
  - **e.g. CERN is moving towards fully cloud-based processing**
  - **Standardization is an enabler: Openstack, Amazon**
- **Gateway to new resources**
  - **Commercial clouds approaching viability, e.g. Amazon spot market**
  - **Dynamic, good for load peaks**
- ***Data management must be resolved***
  - ***CDN type approaches well suited***

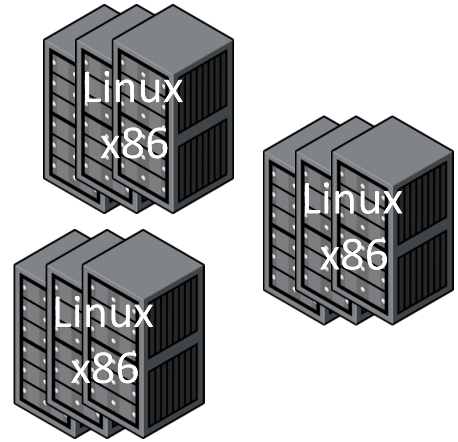
The Big Switch [to the Cloud]: 2009

Computing is turning into a utility…
will ultimately change society as completely as cheap electricity did…



Distributed Computing Evolution

1980s: Plethora of architectures & OSes



1990s: **Uniform OS/architecture**
Linux/x86 standard for commodity cluster computing

2000s: **Uniform fabric and access**
Globally federated resources enabled by network and grid

2010s: **Uniform environment**
VMs and clouds put the user in control of the environment – take it with you anywhere and everywhere

CDN
http
xrootd
AAA
FAX
Virtual data
Event service

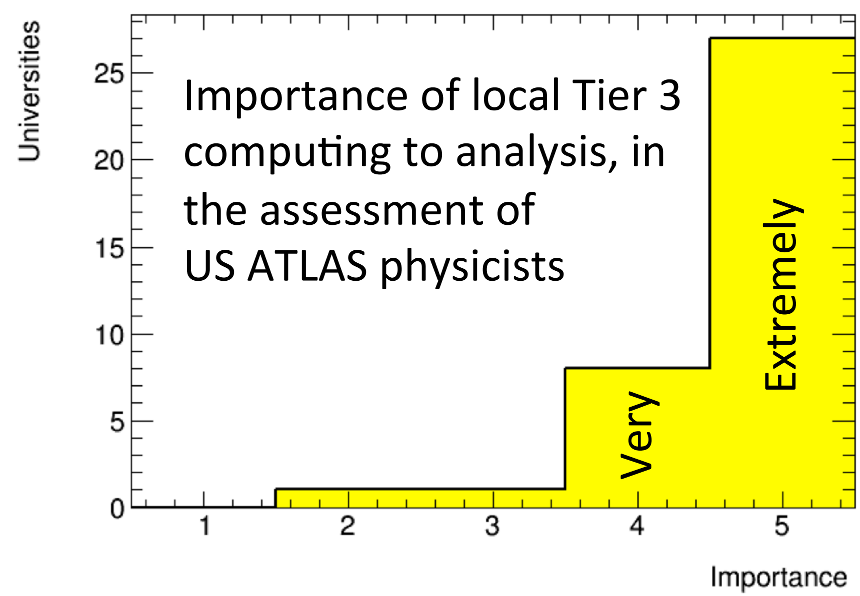
2010s: **Uniform data access**
Working towards transparent distributed data access enabled by the network

Access and Usability

- Developer and user training essential in increasingly complex software & computing environments
- With distributed computing becoming more pervasive, access and usability must be improved…
  - Easy to acquire rights and authenticate
  - Easy to access, move, manage data
  - Easy to use processing resources
- Promising initiatives underway, Open Science Grid an important driver

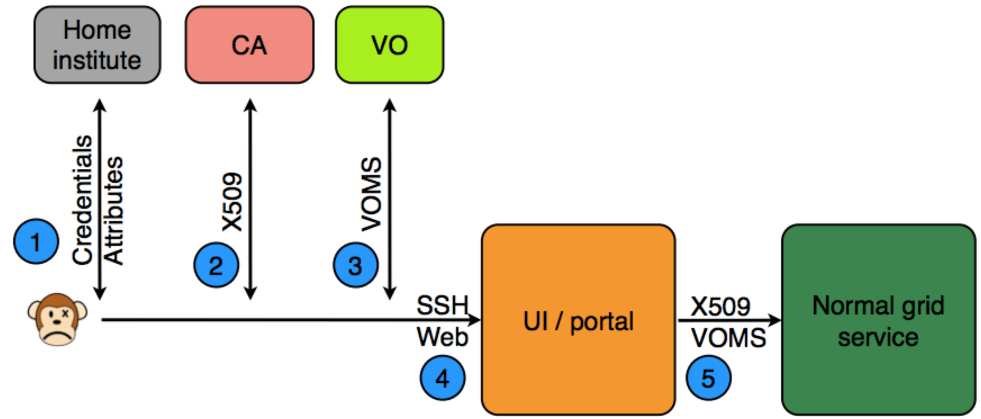
And yet, even in the LHC community accustomed to grids, local computing remains an extremely important complement to large scale resources…

Scientific computing begins and ends at home, on campus…

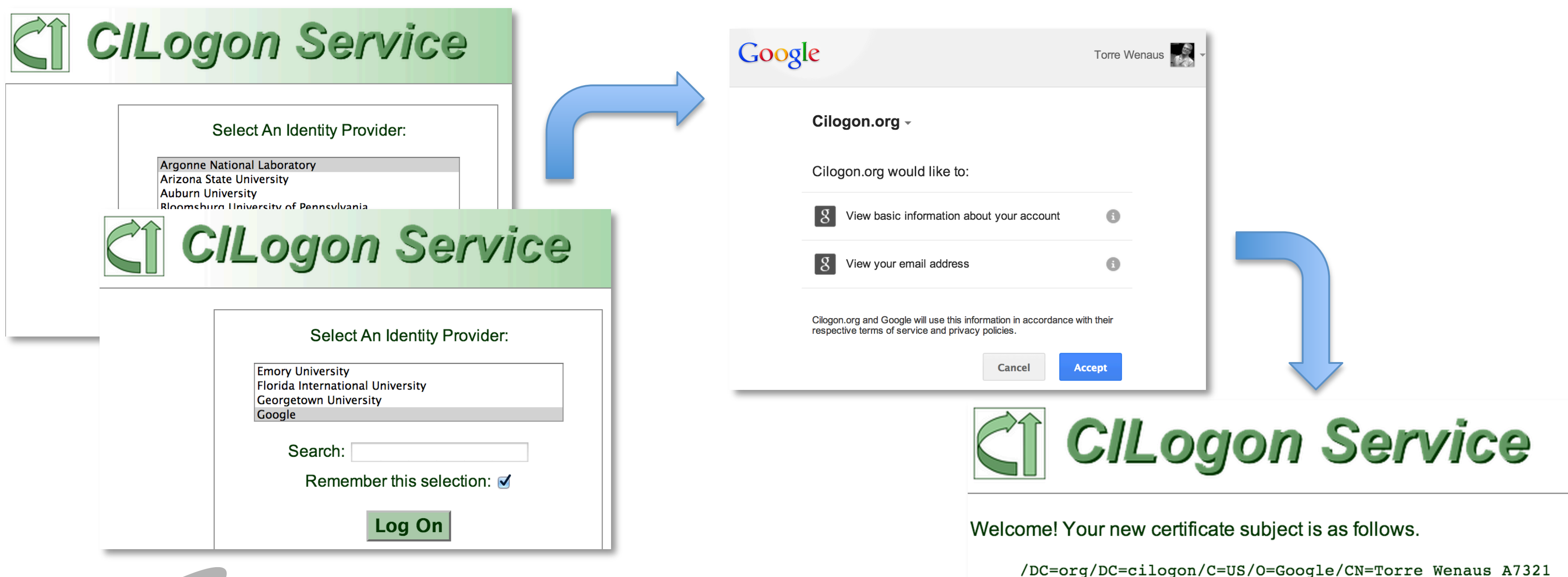


Improving on Grid Certificates

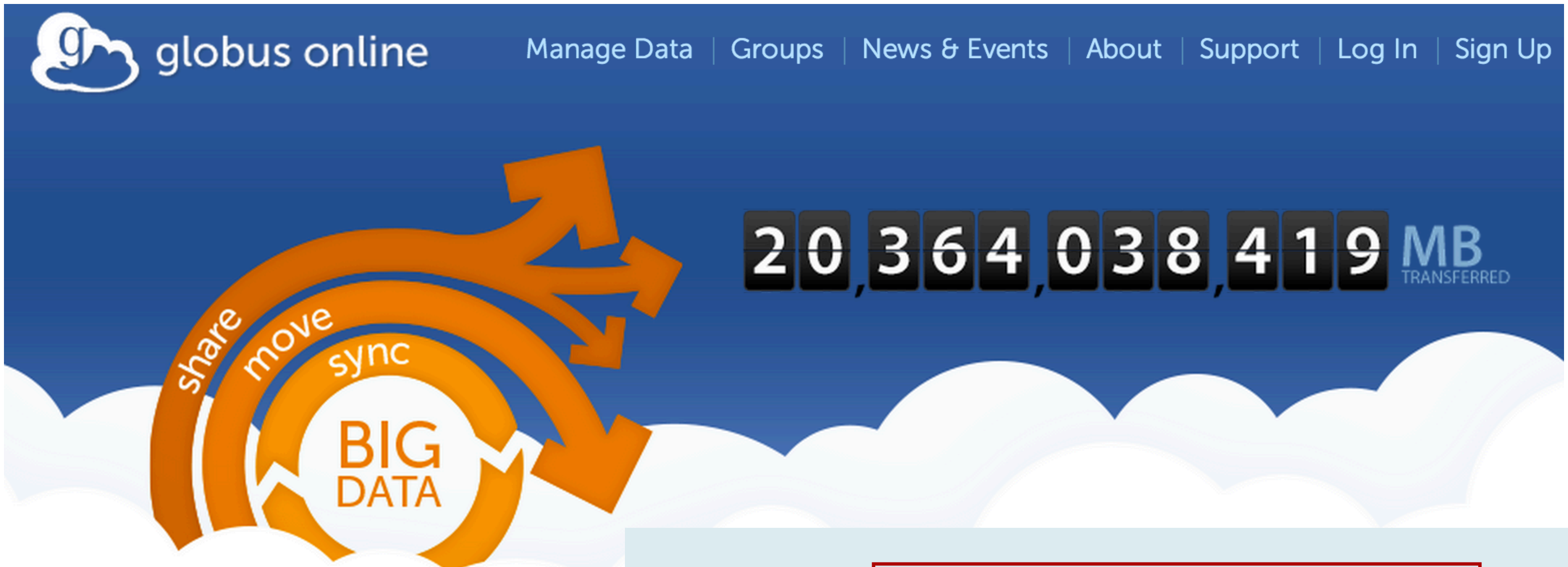
The current bad old days…



OSG and WLCG pursuing an easy to use (and manage) CILogon.com based service. Objective: A certificate-less grid



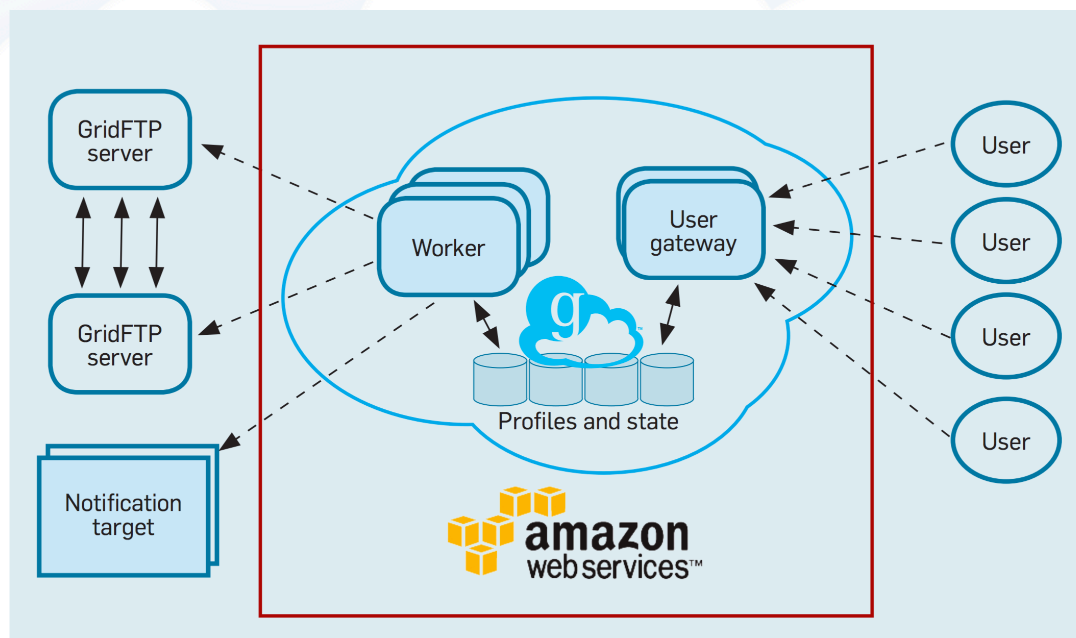
Easy Data Access and Movement



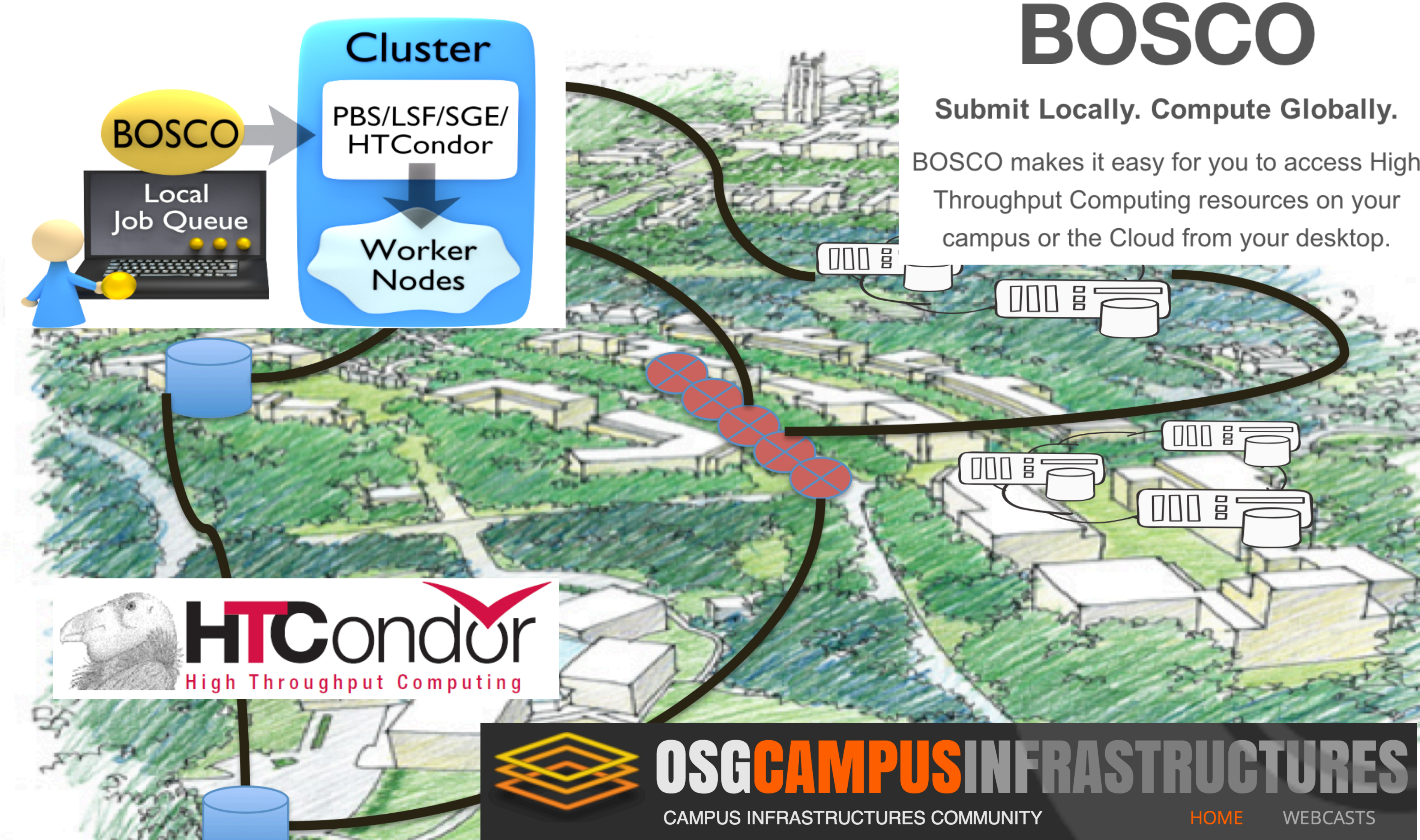
Easy and (generally) free tool to access and move data

Example of software service utilizing a commercial cloud (transparently)

https://www.globusonline.org
Training webinars



Easy Access to Processing



http://www.campusgrids.org/

Training webinars

Conclusions

- HEP science demands **substantial growth** in data and computing in coming years
- We're resource limited – big challenges met with small budgets – so we must make efficient and flexible use of **all available resources**
  - We get better at it as networks and distributed systems become more capable and software becomes less serial
- **Data management and access** must see large gains in efficiency, and prospects of achieving that are good, with continued R&D and development
- **Networks** are a powerful foundation and enabler, a resource requiring investment and paying big dividends
  - Likewise the distributed data and processing management services above that *unify resources into a usable system*
  - Emerging network technologies will powerfully link these layers
- We are entering an unavoidable, challenging **re-architecting of our serial software**
  - We can't track Moore's Law otherwise – *cost effectiveness would shrivel*
  - **Investments in common development to solve common problems are important**
    - Several already happening with agency support and sponsorship
    - Virtual Center for HEP Computing being explored in DOE
- Ubiquitous Big Data is good news, clouds are good news, industry clearing the path is good news… *there is much for us to leverage*

More information

- Detailed Snowmass presentations from the computing frontier subgroups
  - https://indico.fnal.gov/sessionDisplay.py?sessionId=44&confId=6890#20130730
- Snowmass computing frontier contacts
  - computingfrontier@denali.physics.indiana.edu for Bauerdick & Gottlieb
  - allcomputingfrontier@denali.physics.indiana.edu for all subgroup conveners

Webinar: Big Data Management for Science

Presented by Globus Online and ESnet
Thursday, August 22, 2 pm EDT/11 am PDT

Target audience: Research scientists dealing with big data
Length: one hour
Register: https://www.globusonline.org/events/big-data-management-for-science-joint-esnetglobus-online-webinar/

Meeting on HEP Computing Dec 9-11 2013

Chairs: Paul Avery, Salman Habib

Advisory committee: A. Boehnlein, R. Roser, H.Schellman, S. Sharpe, C. Tull, T. Wenaus

Objectives:

- Identify cross cutting aspects of computing and simulations with common features that could benefit from common solutions
- Identify opportunities for computational R&D with high impact, including potential for international leadership
- Survey widely used HEP software packages and identify
    - Which need continued maintenance and updating
    - Gaps where a common tool is needed but missing
    - Areas we can benefit most from partnerships beyond HEP (use tools from others)
    - How to manage software lifecycle
- Survey computing and data management practices across HEP and determine whether more and improved program-wide structure would accelerate science
- Survey computing hardware usage across HEP to identify improvements in efficiency, cost effectiveness, and technology selection
- Explore the value of a Virtual Center for HEP Computing Excellence
    - Distributed experts in different aspects of computing to promote/facilitate common solutions

Organized by Lali Chatterjee & Larry Price, DOE HEP Computing

Data Preservation

- Irreplaceable resource of 100s of PBs to preserve for decades, meet open access requirements
- Cosmic Frontier in good shape – images and catalogs are accessible and intelligible by community and public
- Policies being developed in EF, acting on them will require new funding
  - Common projects underway, international and cross-disciplinary

DASPOS Data and Software Preservation for Open Science

https://daspos.crc.nd.edu/

DPHEP Study Group for Data Preservation and Long Term Analysis in High Energy Physics

http://www.dphep.org/

From the European HEP strategy statement on computing…

A HEP-wide forum is needed where strategic issues for computing for the next decade can be discussed and the common work coordinated.

Many particle physics experiments have a lifecycle that is beyond the lifecycle of the computing technology used and as a consequence data preservation is a significant concern.

The study group for Data Preservation and long-term analysis in High Energy Physics (DPHEP) has taken the lead in this important area.

The experimental collaborations in particle physics are aware of the need for data preservation and open access to the data and are developing clear policies and plans.