

2020-2025

Scientific Computing Environments (Distributed Computing in an Exascale era)

August 7 2013

Geoffrey Fox

gcf@indiana.edu

<http://www.infomall.org> <http://www.futuregrid.org>

School of Informatics and Computing
Community Grids Laboratory
Indiana University Bloomington



<https://portal.futuregrid.org>

2020-2025 Scientific Computing Environments (Distributed Computing in an Exascale era)

- 1) The components of national research computing in exascale era with mix of high end machines, clouds (whatever commercial companies offer broadly or publicly), university centers, high throughput systems and with growing amounts of distributed and "repositorized" data serving High End and Long Tail researchers.
- 2) The nature of an environment like XSEDE in the Exascale era; i.e. the nature of a distributed system of facilities including one or more exascale machines. Should it be relatively tightly coupled like XSEDE or more loosely coupled like DoE leadership systems (or both!)



Considerations

- Both of these major topics can be considered with attention to
- A) What services do 2025 science projects need from cyberinfrastructure; examples are -- Collaboration; On-demand computing; Digital observatory; High-speed scratch and persistent storage; Data preservation; Identity, profile, group management, reproducibility of results, versioning, and documentation of results
- B) What are requirements in 2025 -- are there changes in distributed system requirements outside details of exascale machines with their novel architecture e.g.
 - a) Will big data lead to new requirements
 - b) Will feeding/supporting exascale machine lead to new requirements
 - c) Will supporting long tail of science lead to new requirements
 - d) Can we do more to make people use central services rather than building their own



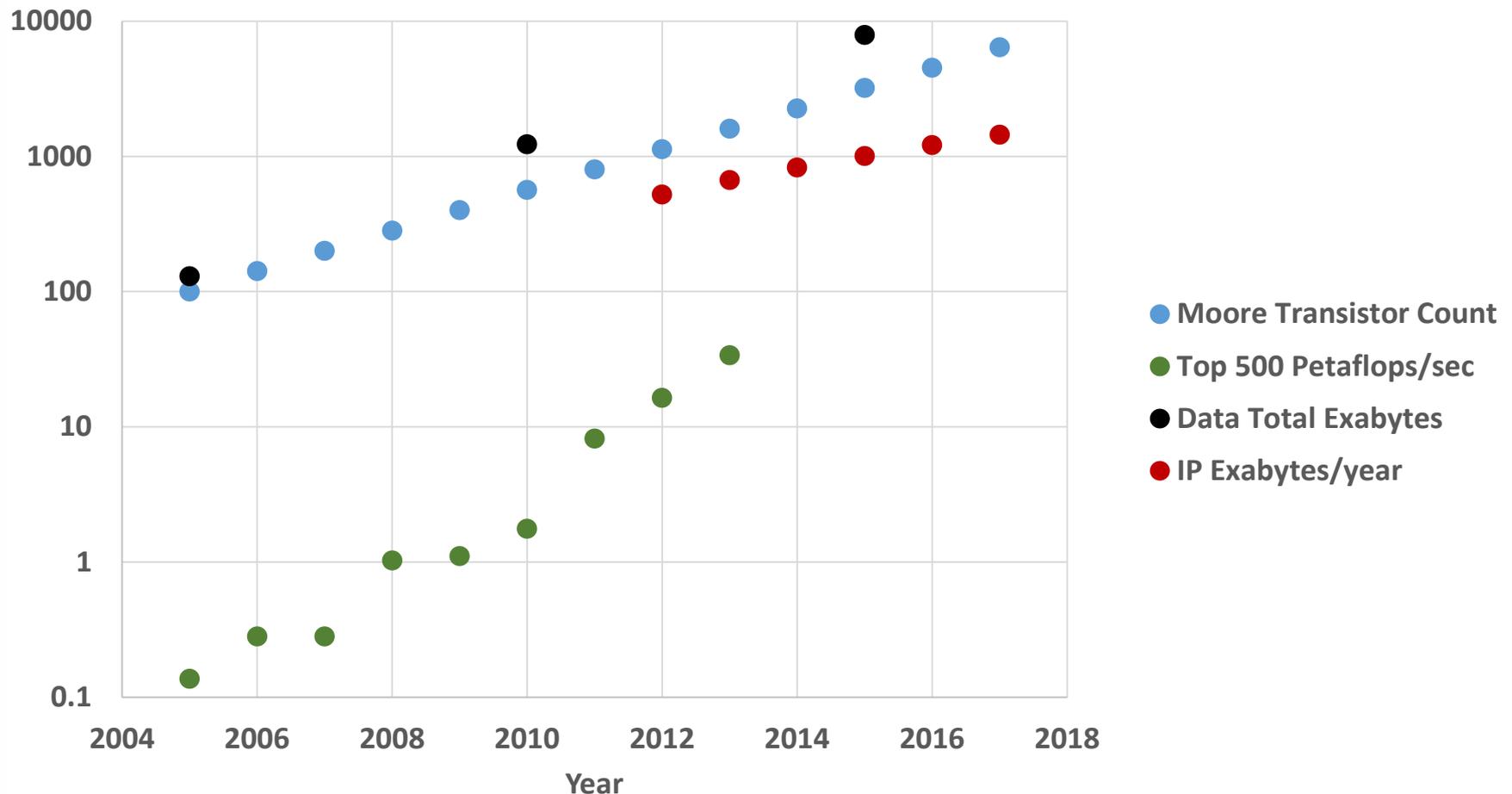
Speakers

- Miron Livny, Wisconsin
- Shantenu Jha, Rutgers
- Dennis Gannon, Microsoft
- Ioan Raicu, IIT
- Jim Pepin, Clemson



Network v Data v Compute Growth

- Moore's Law Unnormalized
- Slope of #1 Top 500 > Total Data > Moore > IP



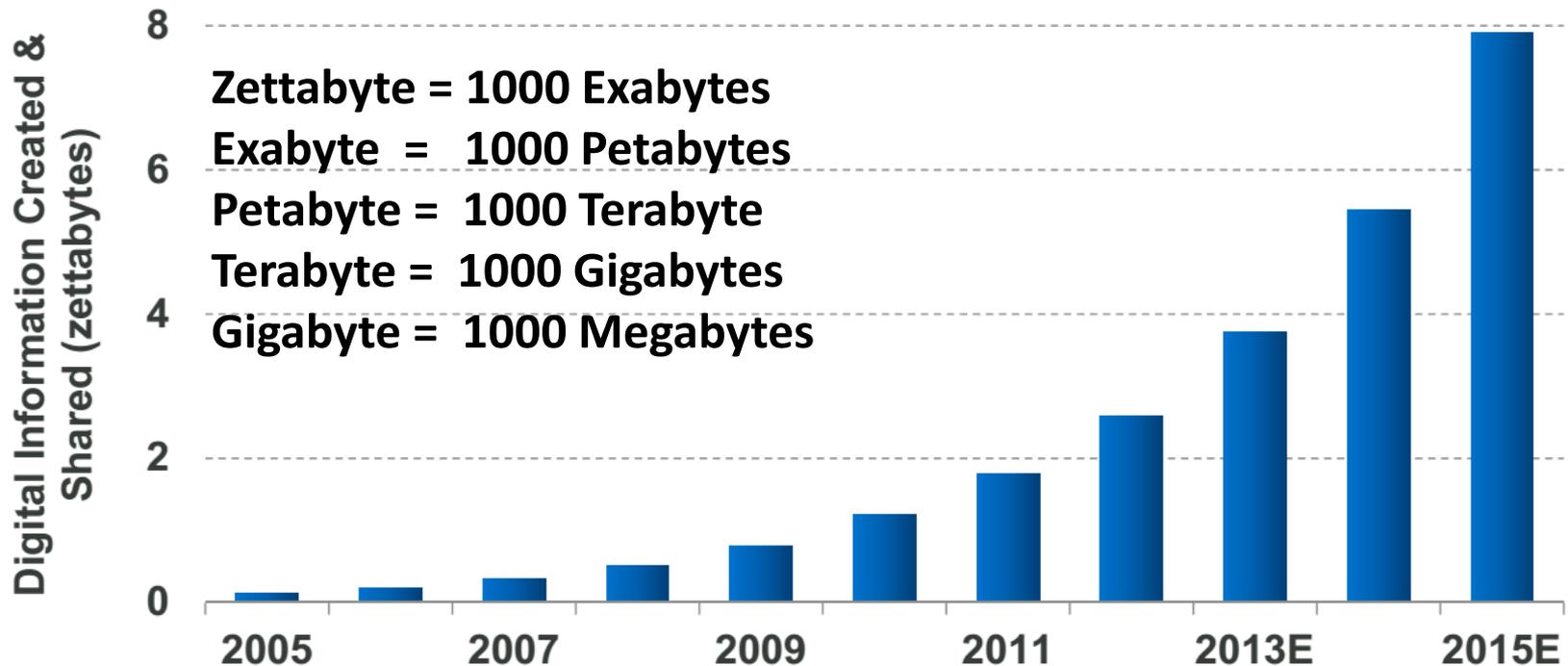
Trends Network, Data, Computing

- Data likely to get larger and produced all over the world i.e. stay distributed
- Network rise underlies MOOC's and Cloud computing
- Not obvious that data/network increase any larger than computing
- Cisco network traffic < Moore's Law
- IDC total data > (little bit) Moore's Law
- Some areas of data like genomics and social images have seen huge (one time?) increases

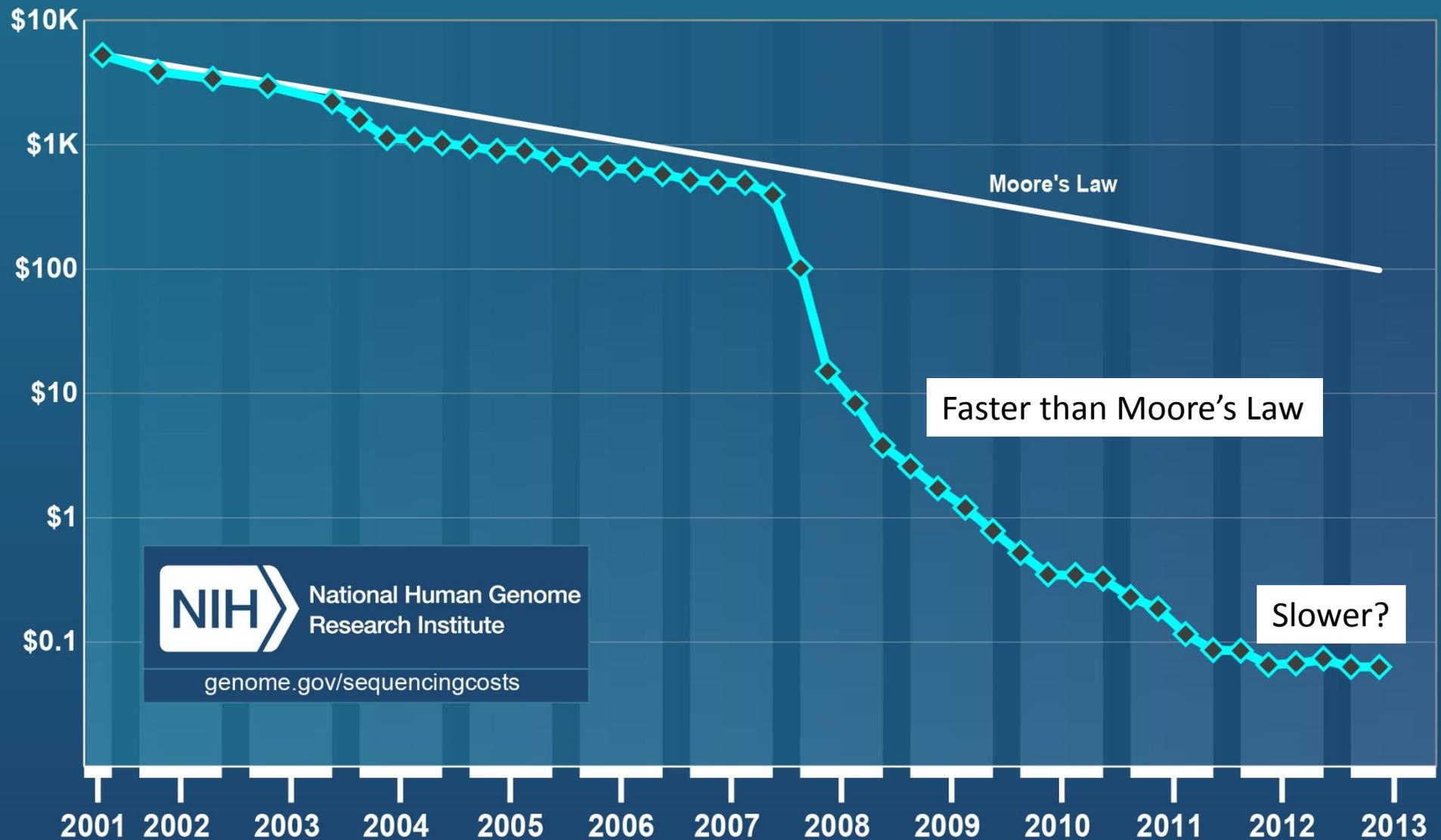
World's Content is Increasingly Findable + Shared + Tagged - Digital Info Created + Shared up 9x in Five Years

Amount of global digital information created & shared – from documents to pictures to tweets – grew 9x in five years to nearly 2 zettabytes in 2011, per IDC.*

Global Digital Information Created & Shared, 2005 – 2015E



Cost per Raw Megabase of DNA Sequence



National Human Genome
Research Institute

genome.gov/sequencingcosts

<http://www.genome.gov/sequencingcosts/>

Photos = 500MM+ Uploaded & Shared Per Day, Growth Accelerating, on Trend to Rise 2x Y/Y...

Daily Number of Photos Uploaded & Shared on Select Platforms, 2005-2013YTD

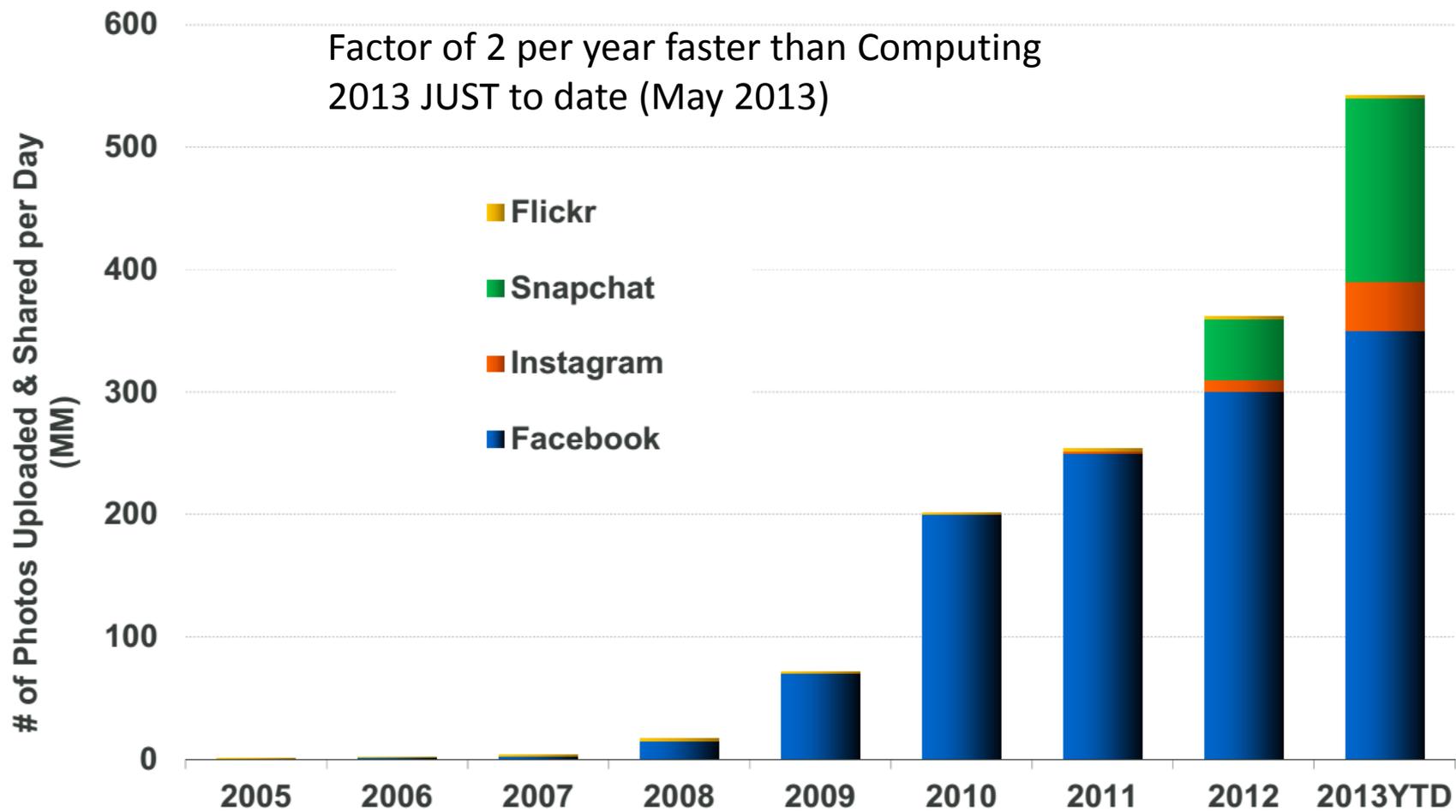


Image based Computations

- **Deep Learning** with COTS HPC, Adam Coates, Brody Huval, Tao Wang, David J. Wu, Andrew Y. Ng and Bryan Catanzaro ICML 2013 (Stanford AI group) http://www.stanford.edu/~acoates/papers/CoatesHuvalWangWuNgCatanzaro_icml2013.pdf
- 64 GPU's on 16 nodes; MPI Speed up of 32; GPU parallelism “perfect”
- Train 11 BILLION parameters in 3 days on just 10 million 200 by 200 images from YouTube (note 500 million per day on FaceBook etc.)
- MPI Parallelism over pixels; GPU uses optimized Matrix-Matrix multiplication with Parallelism over Neuron banks and Images
- Earlier paper NIPS2012 using MapReduce variant with Google (Dean) had MUCH poorer performance on 16000 Intel style cores
- Next: Neural networks for driving: 100 million ~1000 by 1000 images

Object	Random guess	Best in random net	Best in 1.8B param. net	Best in 11B param. net.
Human faces	64.7%	64.8%	88.2%	86.5%
Upper body	64.7%	64.8%	80.2%	74.5%
Cats	64.7%	64.8%	73.0%	69.4%

Is Big Data Changing Requirements?

- Will Compute/Data/Network ratios change?
 - Not obvious but needs more study
- I expect “Data Science” to grow and increased use of large scale data analytics as in deep learning and image clustering (100 million images, 10 million clusters)
 - Richer set of data areas and new users like AI/Image processing
- Compute requirements unclear for data analytics
 - Status of bringing data to computing still unclear
 - NIST BigData effort defining use cases and associated reference architecture
- So changes due to Big Data just because we haven’t got it right now
- However applications like LHC analysis and Long Tail Science will keep high throughput computing structure