Data Centers in a Decade

Eric Lançon

Scientific Data and Computing Center

BROOKHAVEN NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

BROOKHAVEN SCIENCE ASSOCIATES

Content

- Science landscape
- Hardware landscape
- Challenges and path forward for a Data Center

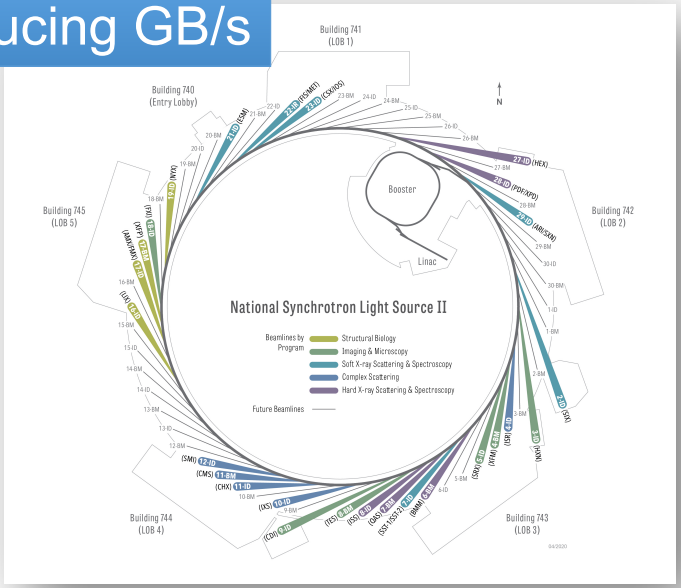
Do you remember 2010….

Science landscape

BNL as an example

X-Ray Light Sources & Nanoscale Science Research Centers

Multiple apparatus per beam line
Some beam line producing GB/s

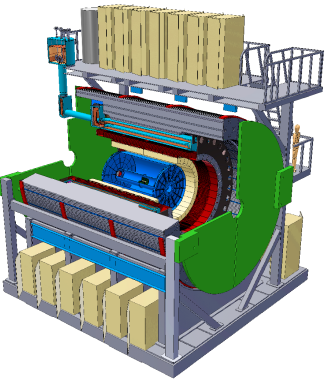


- Entering the Big Data era
- Science programs expanding over next decades
- Need to maximize usage of the instrument

BNL's future programs are Data Intensive

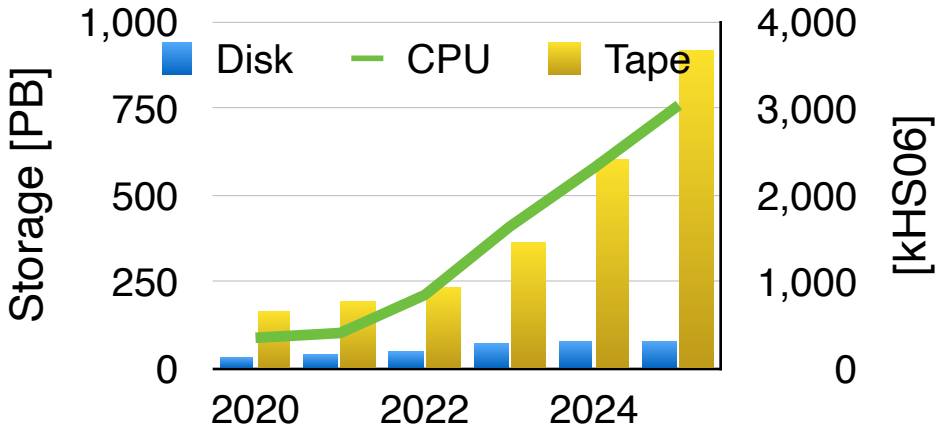
135Gbps streaming into HPSS

sPHENIX



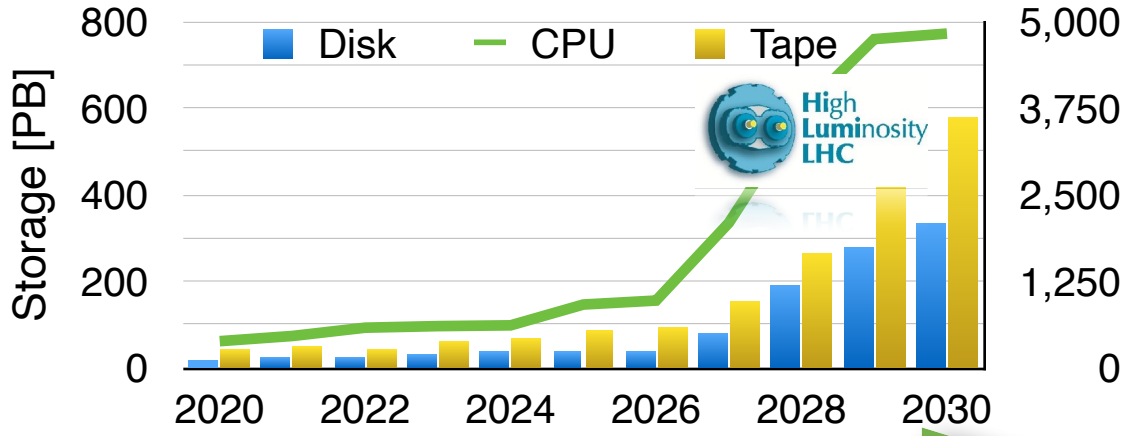
2023 - 2025

RHIC Tier-0



2027 - …

US ATLAS Tier-1



2021- …

NSLS-II

Year	NSLS-II
2021	2.5 PFLOPS
2028	45 PFLOPS

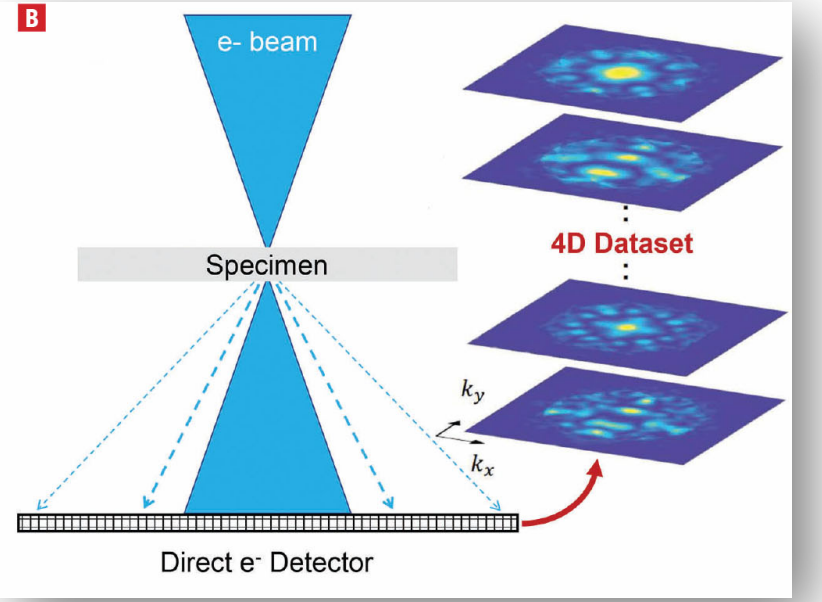
These are estimates….

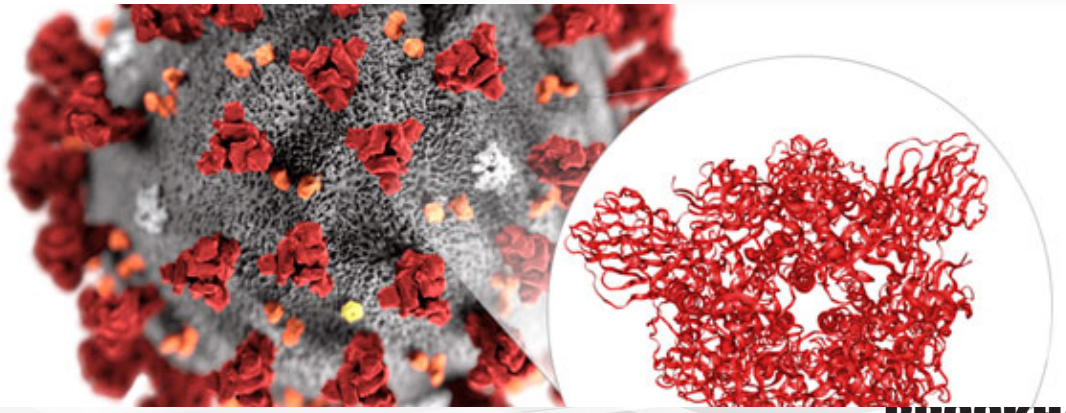
U.S. DEPARTMENT OF ENERGY

BROOKHAVEN NATIONAL LABORATORY

Proliferation of data intensive science programs

- Science of very small scale generates large amount of data
- Data to be processed and analyzed quasi-online
  - Dynamic data taking
- Most applications will exploit co-processor capabilities and AI/ML technics





AI down the road

"AI won't replace the scientist, but scientists who use AI will replace those who don't."

Adapted from a Microsoft report, "The Future Computed"

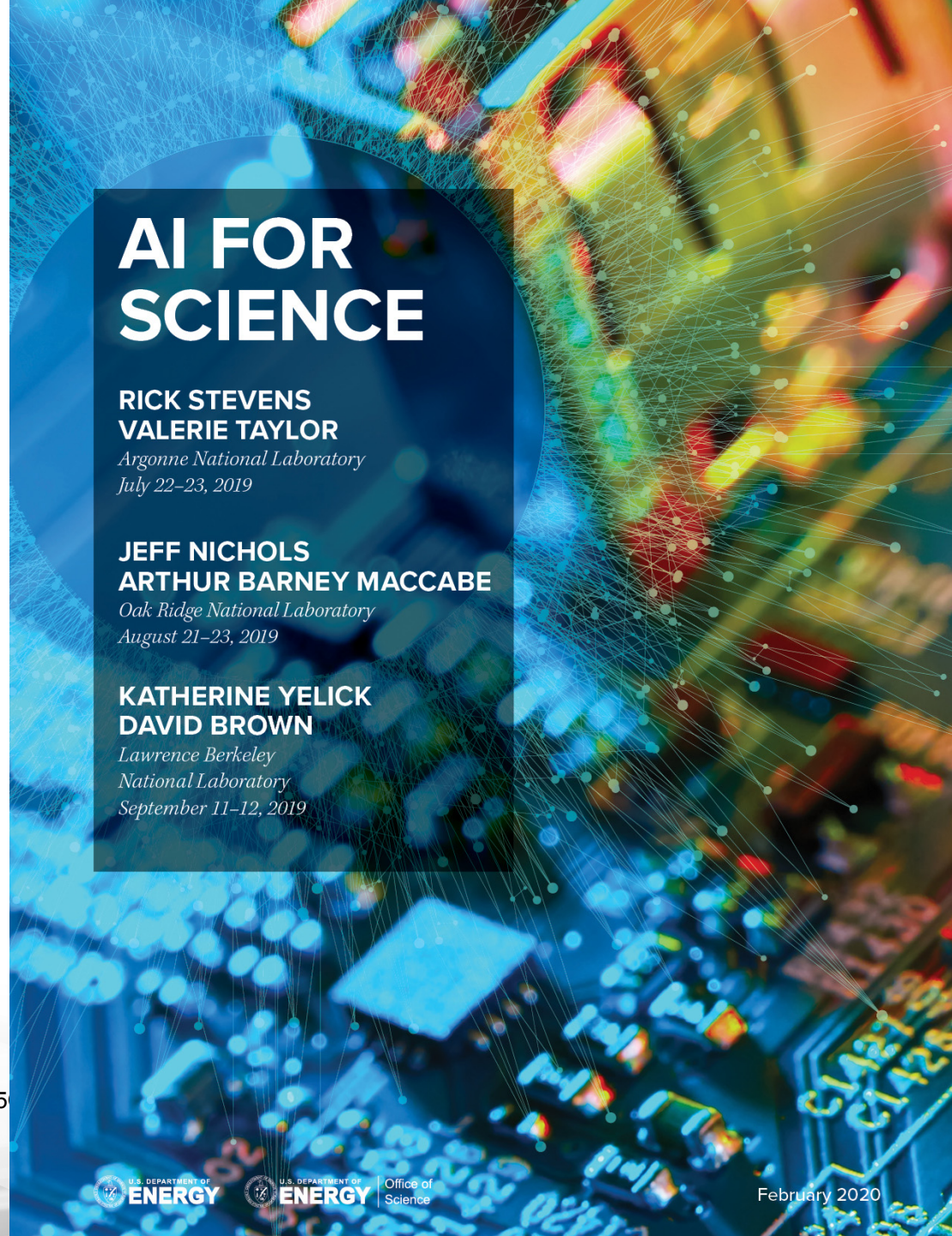
US Department of Energy

Advanced Scientific Computing Advisory Committee (ASCAC)

Subcommittee on AI/ML, Data-intensive Science and High-Performance Computing

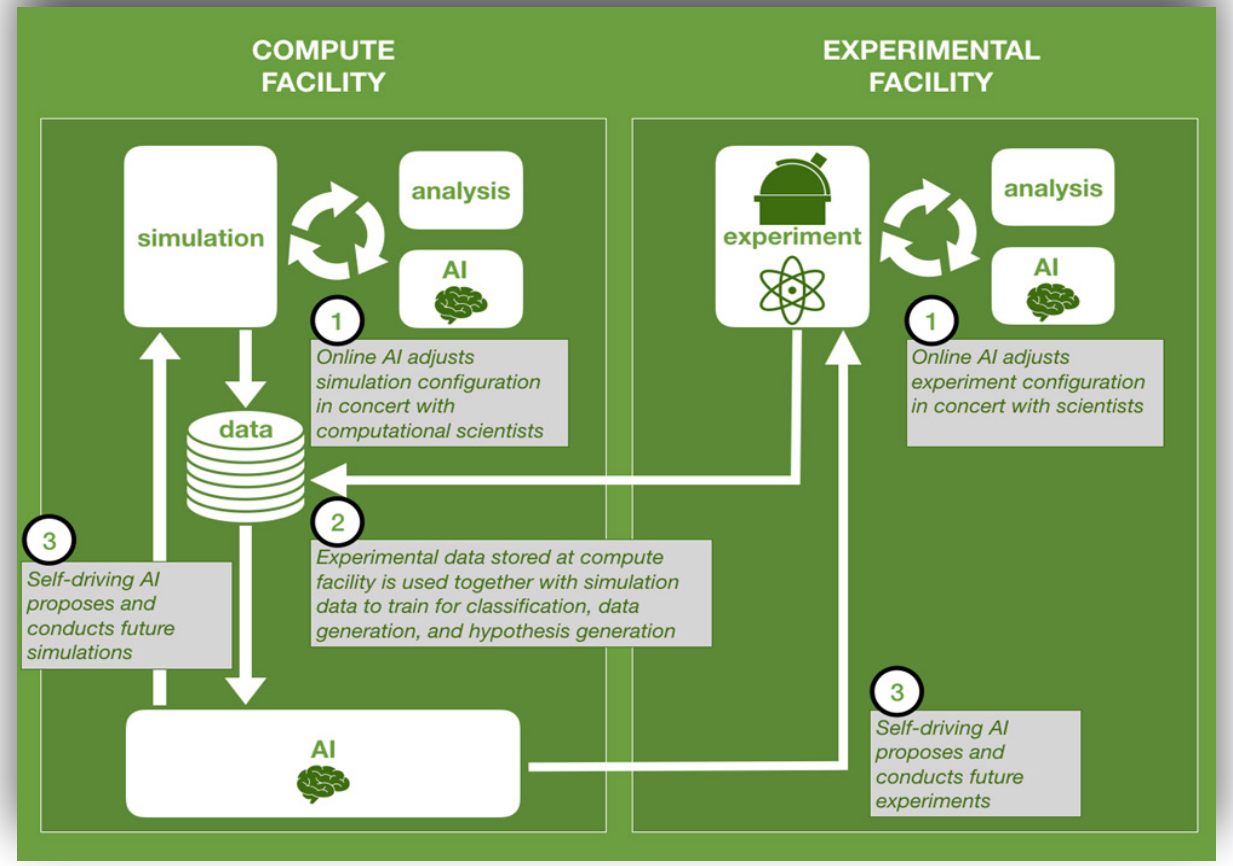
Final Draft of Report to the Committee, September 2020

https://science.osti.gov/-/media/ascr/ascac/pdf/meetings/202009/AI4Sci-ASCAC_202009.pdf?la=en&hash=3BE63EF95



AI at the Data Center (and everywhere)

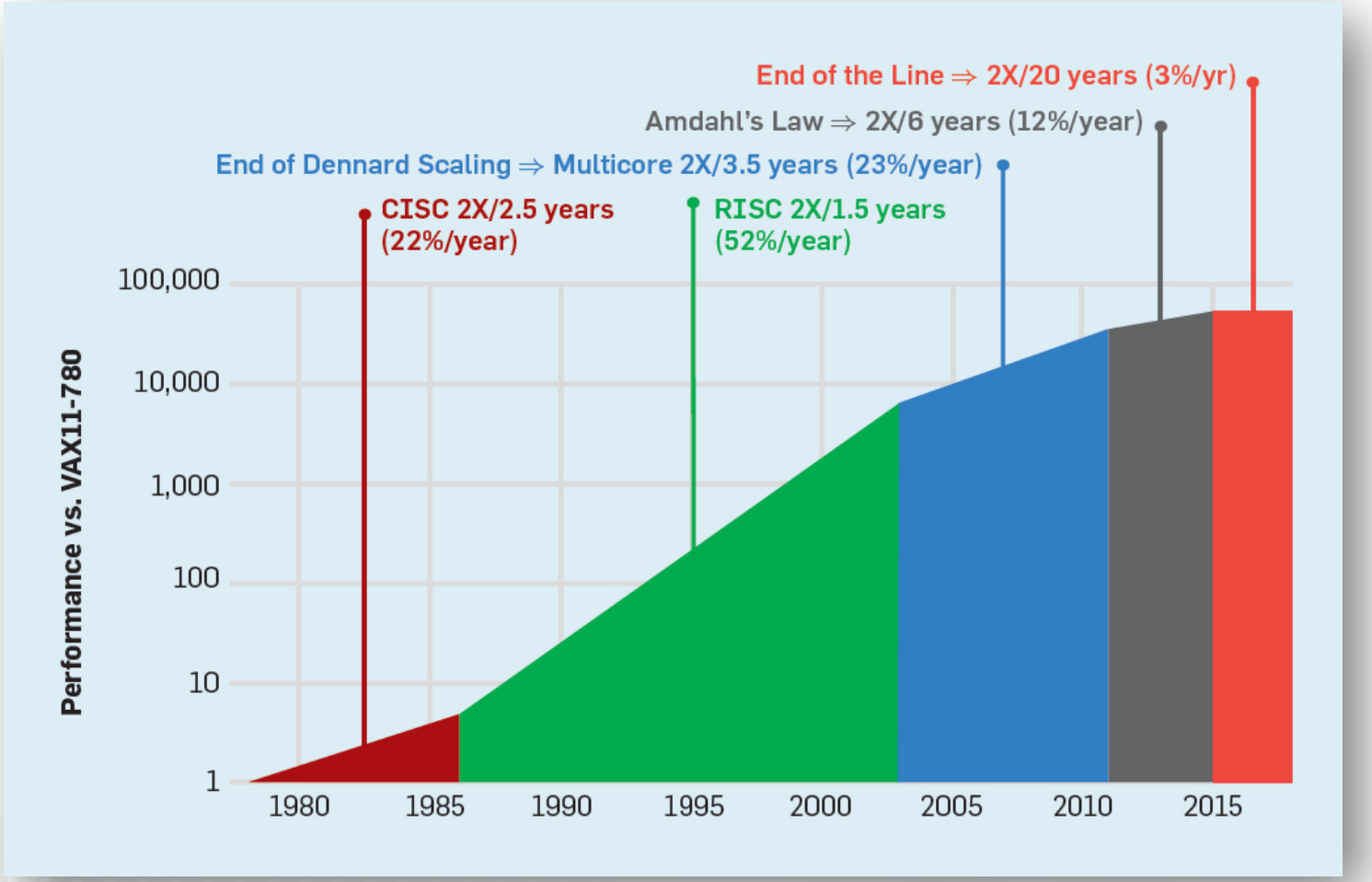
- Complexity and scale of data-intensive science + growing demand for rapid data analysis -> researchers adopts AI methodologies
- These require highly performant computing systems, increasingly combining networks, and fast storage with specialized hardware (accelerators)
- Data center need to provide suitable hardware for AI applications
- Reliable applications require large data samples for training (storage…)



Hardware landscape

2010: Tianhe 1A became #1 of top500 in November

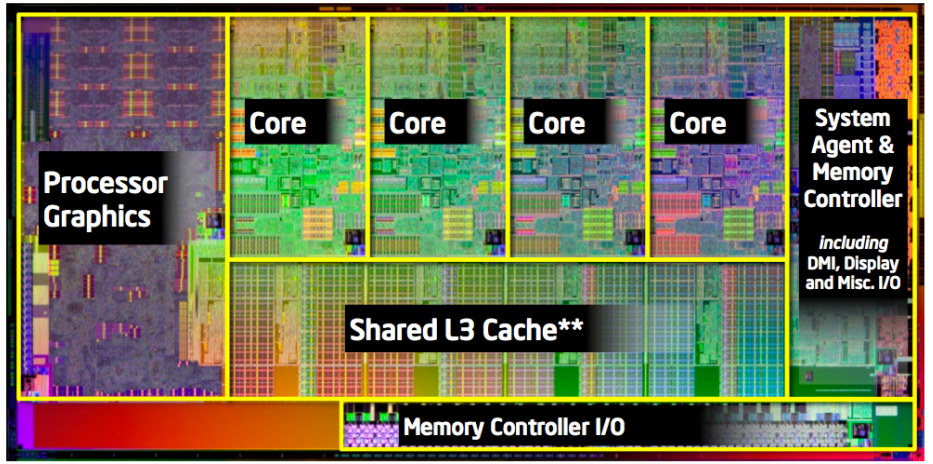
Golden age of easy computing is over



https://cacm.acm.org/magazines/2019/2/234352-a-new-golden-age-for-computer-architecture/fulltext#R1

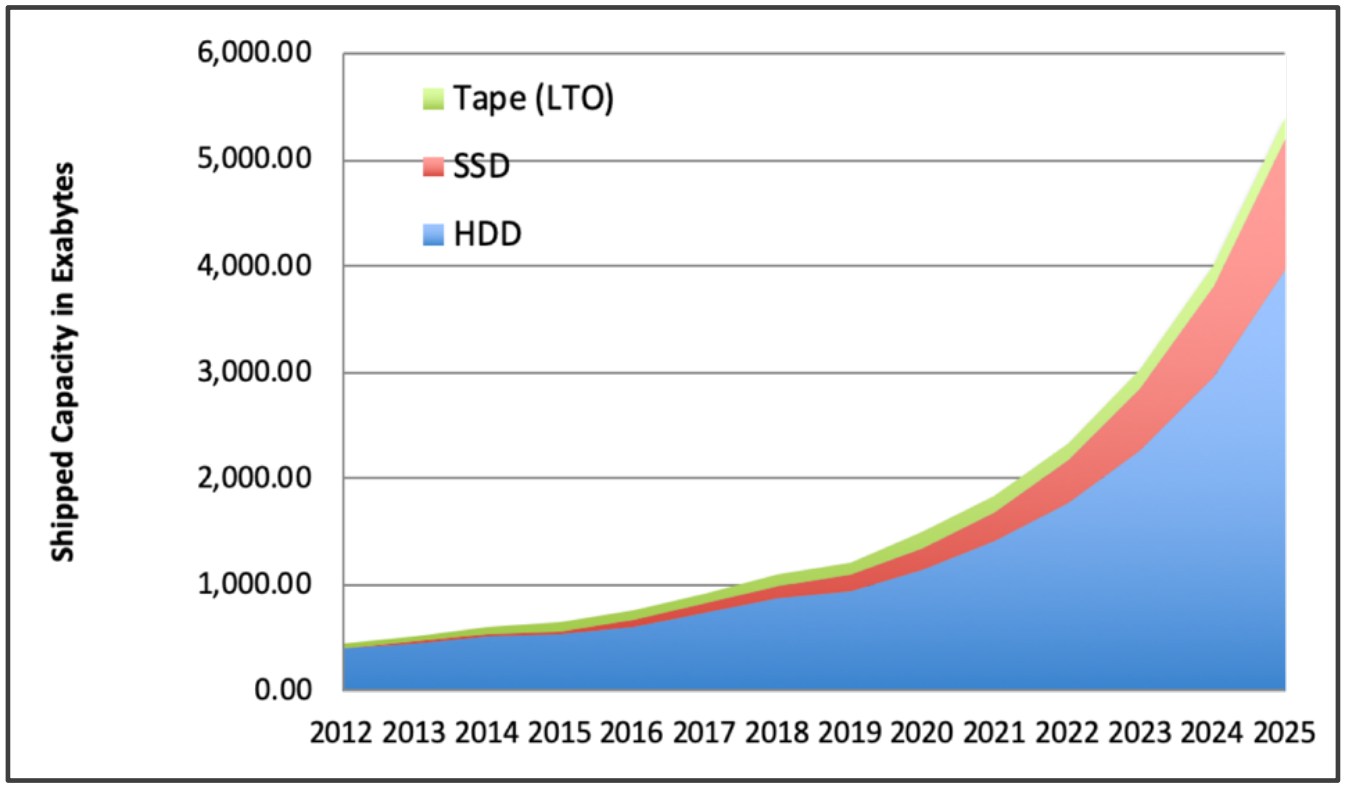
Beyond Moore's : Heterogeneity for compute

- Multiple approaches to future hardware
  - New semiconductor technologies
  - Domain specific accelerators
  - New approaches (quantum computing)
- Algorithms and software customized to match application specific accelerators
- Accelerators and low precision hardware is our immediate future in complement of traditional multi-core CPUs
- 'Gold age for computing architecture'



Storage: shipped volume over time

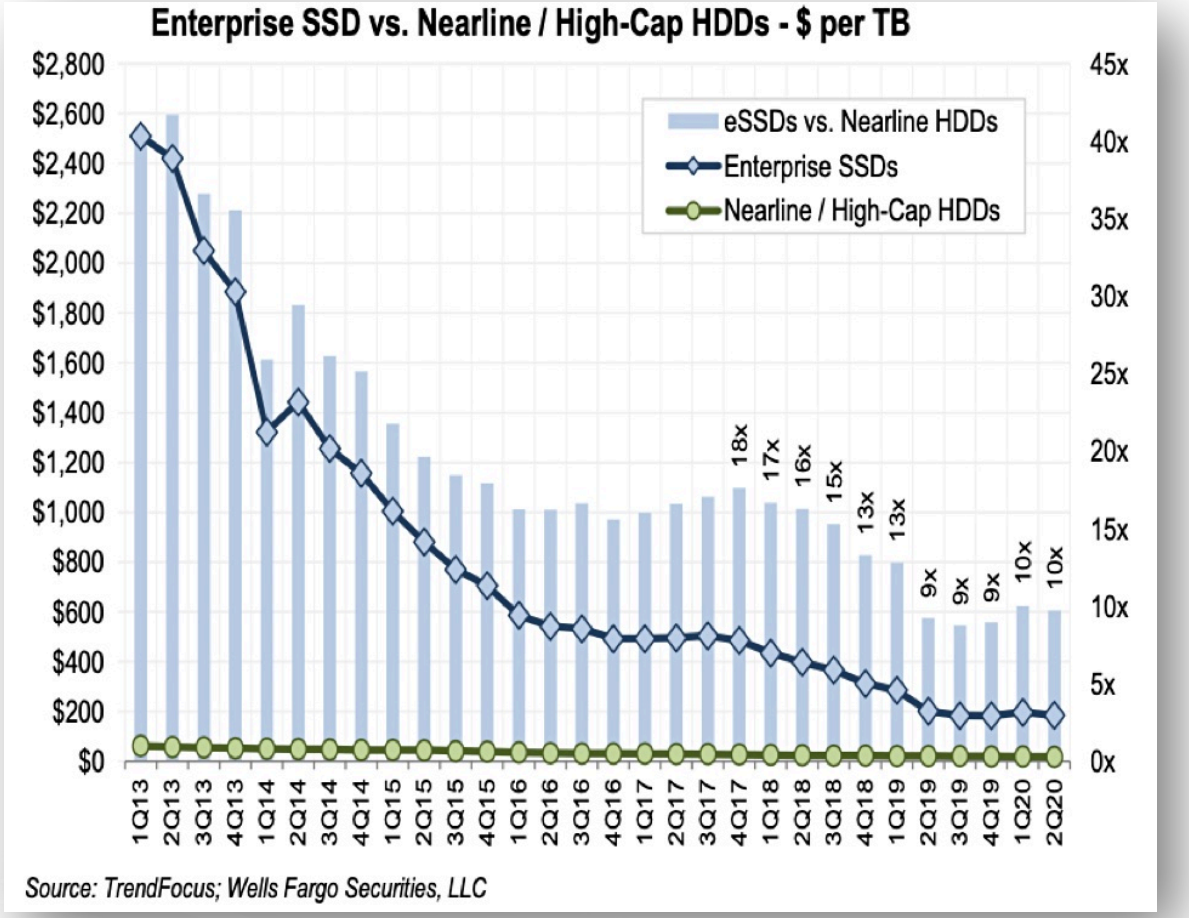
- This is the ‘shipped capacity’ in ExaBytes:
  - This is not the volume ‘on the floor’
- SSD still a small fraction
- Disk continues to dominate
- Tape volume is small



https://www.forbes.com/sites/tomcoughlin/2020/05/29/hdd-market-history-and-projections/#28ef75686682

SSDs

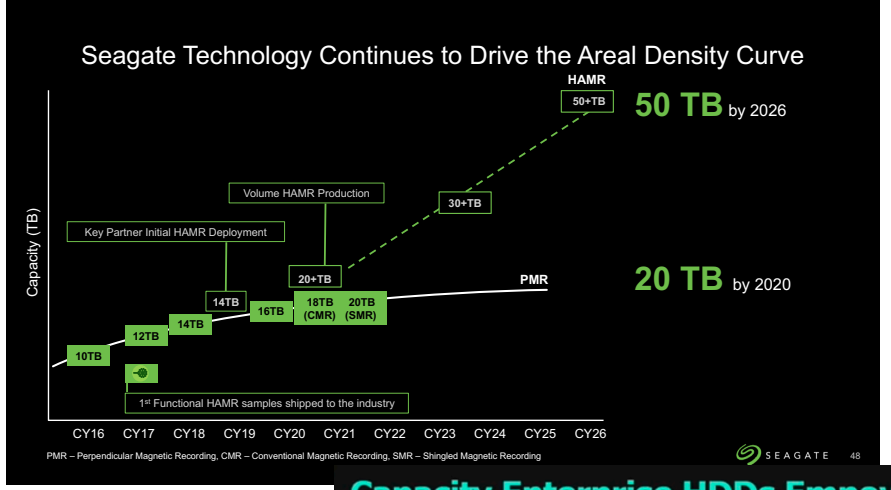
- SSDs have many advantages:
  - Cheaper to operate than disk,
  - Need less power and cooling,
  - Much faster to access
- However they are expensive
  - $/GB for current SSD vs HDD has recently stabilized around 10
- Less performant (cheaper) SDDs are coming
- SDD will not take over HDD for some time (in Data Centers)

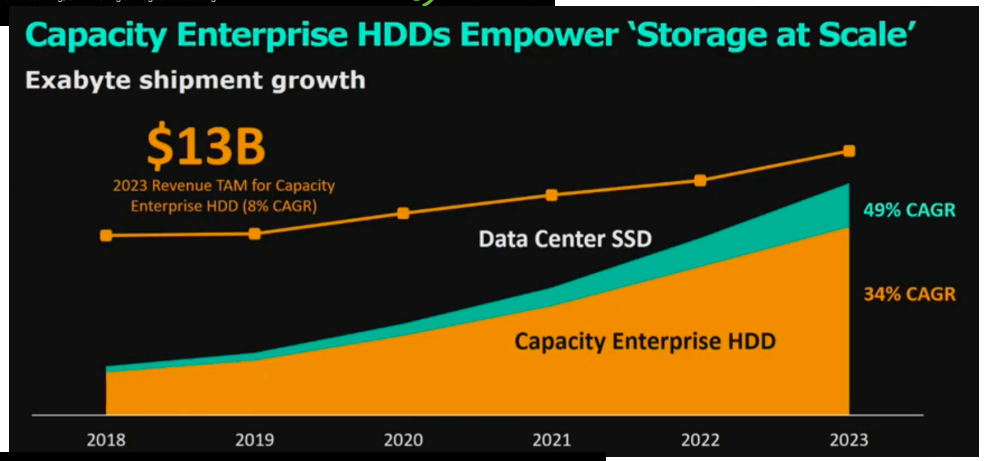


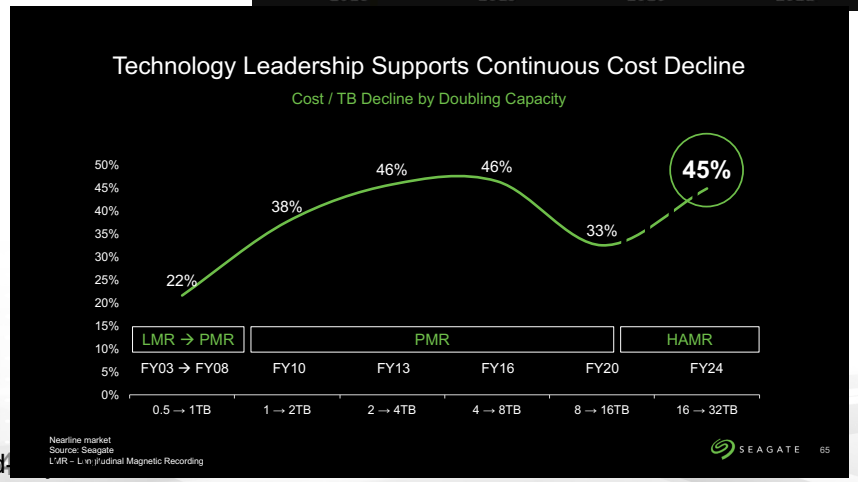
https://blocksandfiles.com/2020/08/24/10x-enterprise-ssd-price-premium-over-nearline-disk-drives/

Disks are still around

- HDDs have been saved by '**hyperscalers**' (Amazon, Apple, Facebook, Google, Microsoft, Baidu, Alibaba,…) who buy ~½ of all disk drives
  - For their use cases, tape is cheaper but too slow & SSD is too fast for the need and too expensive
- Hardware companies (few left) have convincing roadmaps with high capacity low cost disk







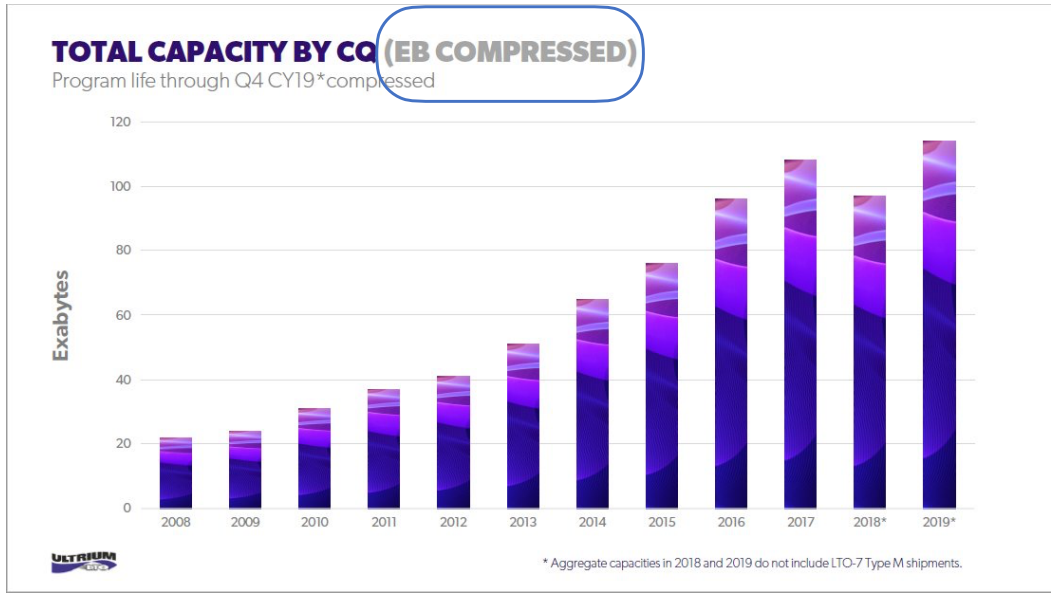
https://s24.q4cdn.com/101481333/files/doc_downloads/presentation/09/2019-Seagate-Analyst-Day.pdf

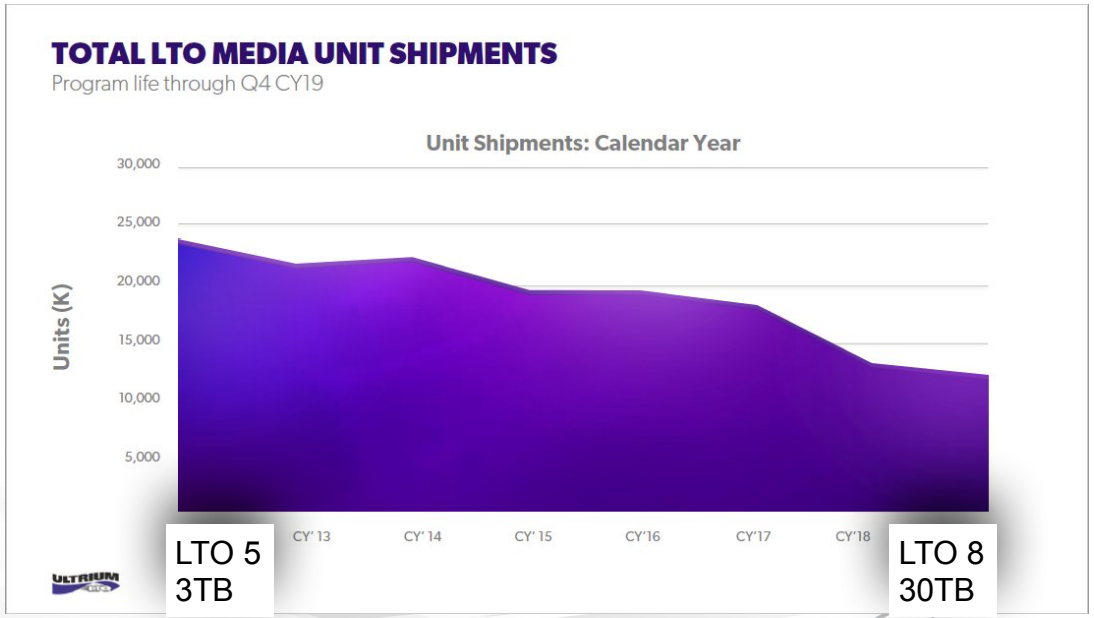
https://techfieldday.com/appearance/western-digital-presents-at-storage-field-

https://www.anandtech.com/show/15457/western-digital-roadmap-updates-energy-assisted-recording-multistage-actuators-zoned-storage

Tape are around as well

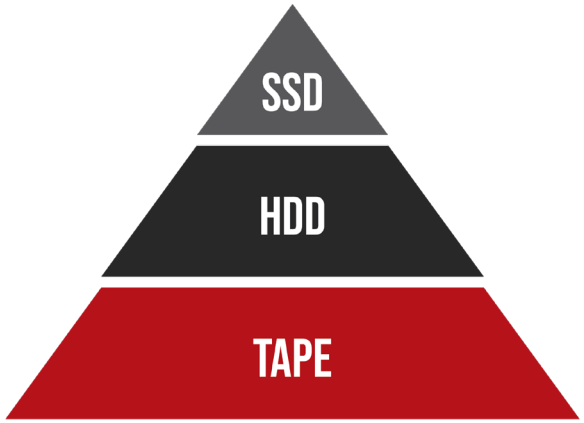
- After the LTO 8 dispute, shipments have started to rise again
- Number of media unit shipped per year continues to drop
  - Tape capacity/unit increases
  - Tape have long lifetime compared to Disk
- ‘Hyperscalers’ are the main consumers of tape systems
- Concern: IBM is the only company in the drive business



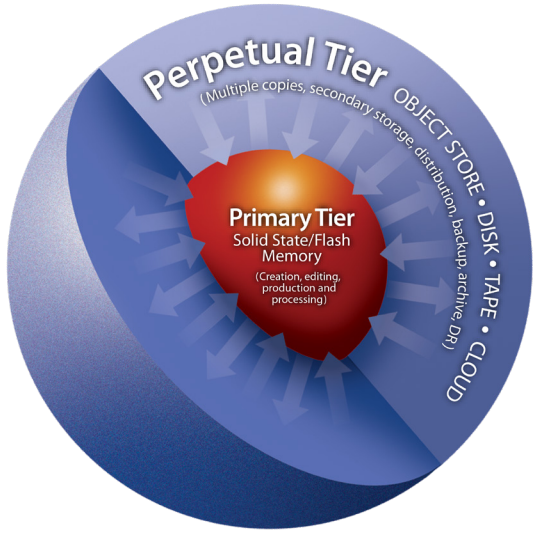


https://blocksandfiles.com/2020/07/10/lto-tape-shipments-2019/

Tape will stay for some times



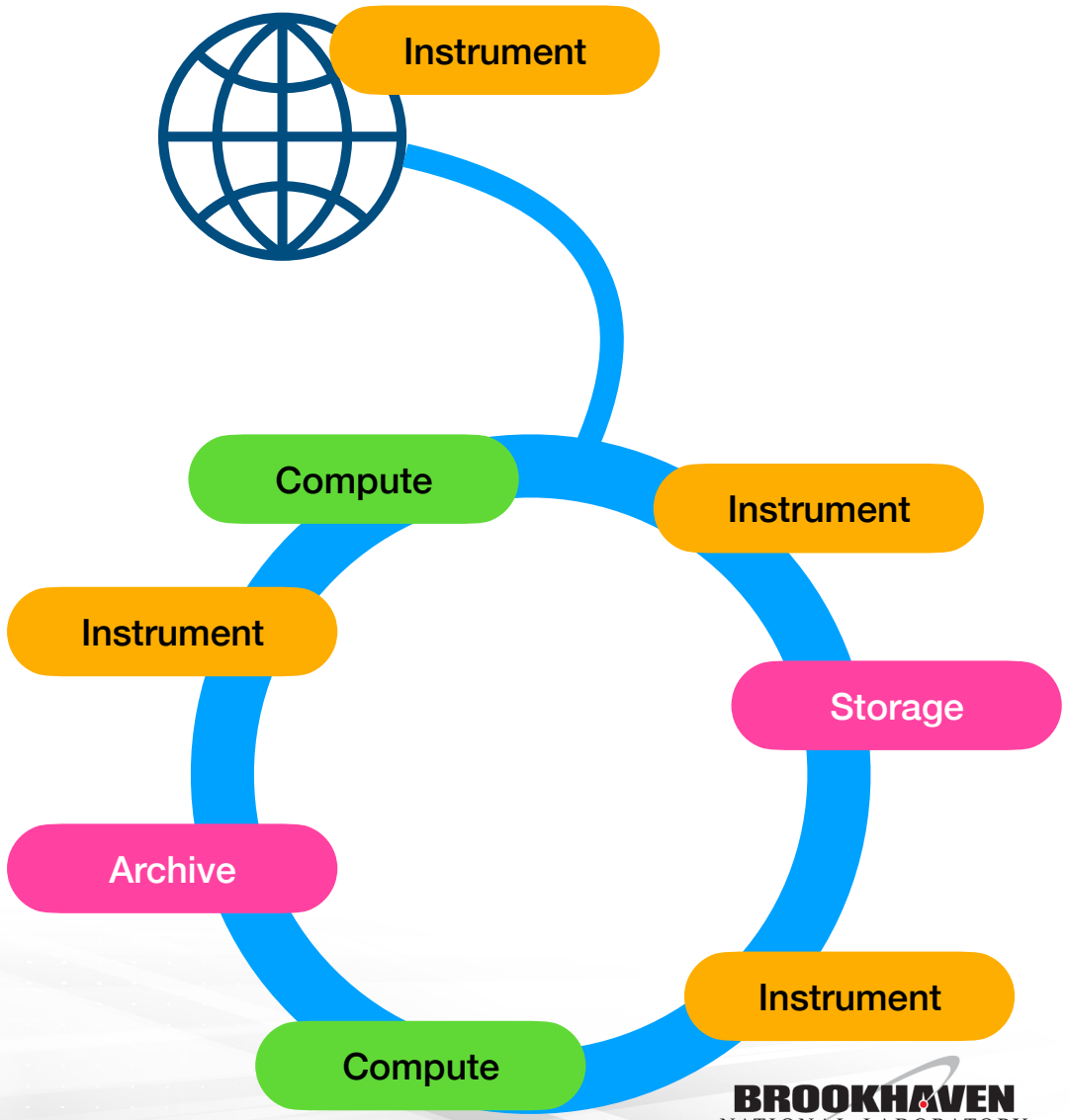
- Tape will remain the most competitive mean to store large amount of infrequently used data
- Secure and reliable long-term archival storage, cost substantially lower than disk or cloud storage when considering factors such as power, cooling and retrieval
- Tape is part of every solution addressing cost effective storage



Storage solutions will evolve towards improved hierarchical architectures involving various technologies

Data intensive computing needs network

- No efficient solution without high performance, reliable network
- Speed of science delivery goes with the quality and performance of the network
- LAN: internal multi Tbps network connecting instruments and hardware
  - Data streaming
  - Data processing / analysis
- WAN: external multi 100s Gbps network
  - Data exchange



In summary

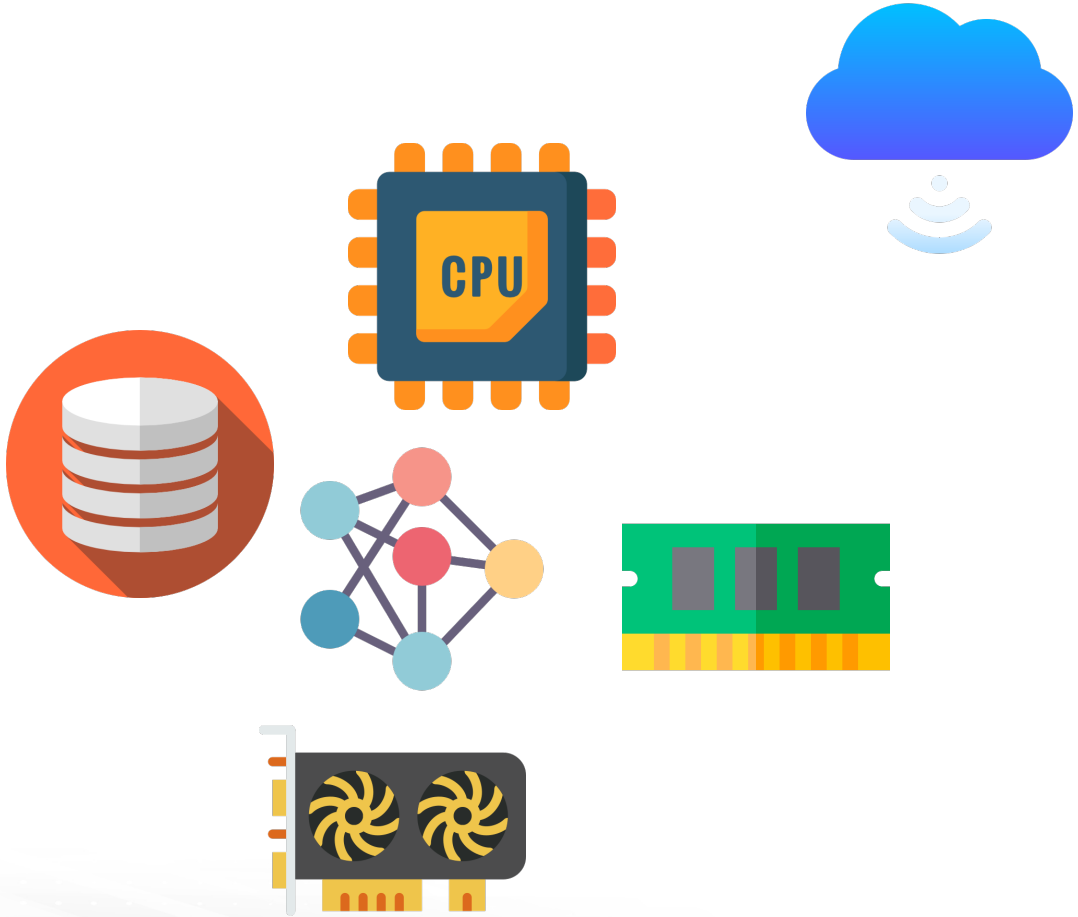
- Tape and HDDs will remain the building blocks for large storage solutions (with some software improvements)
- Accelerators will progressively expand
- Dedicated (fast) storage will be required to match hardware and AI applications needs
- Highly performant reliable LAN as the foundation of a Data Center

One solution for all will not work

And critical size is required …

Data Center: an evolving composable system

- Foundation will remains network, long term repository and central storage other components will evolve with time
- Evolution required to address the evolving needs to stakeholders and changing technologies
- May includes incorporating cloud-based services as needed

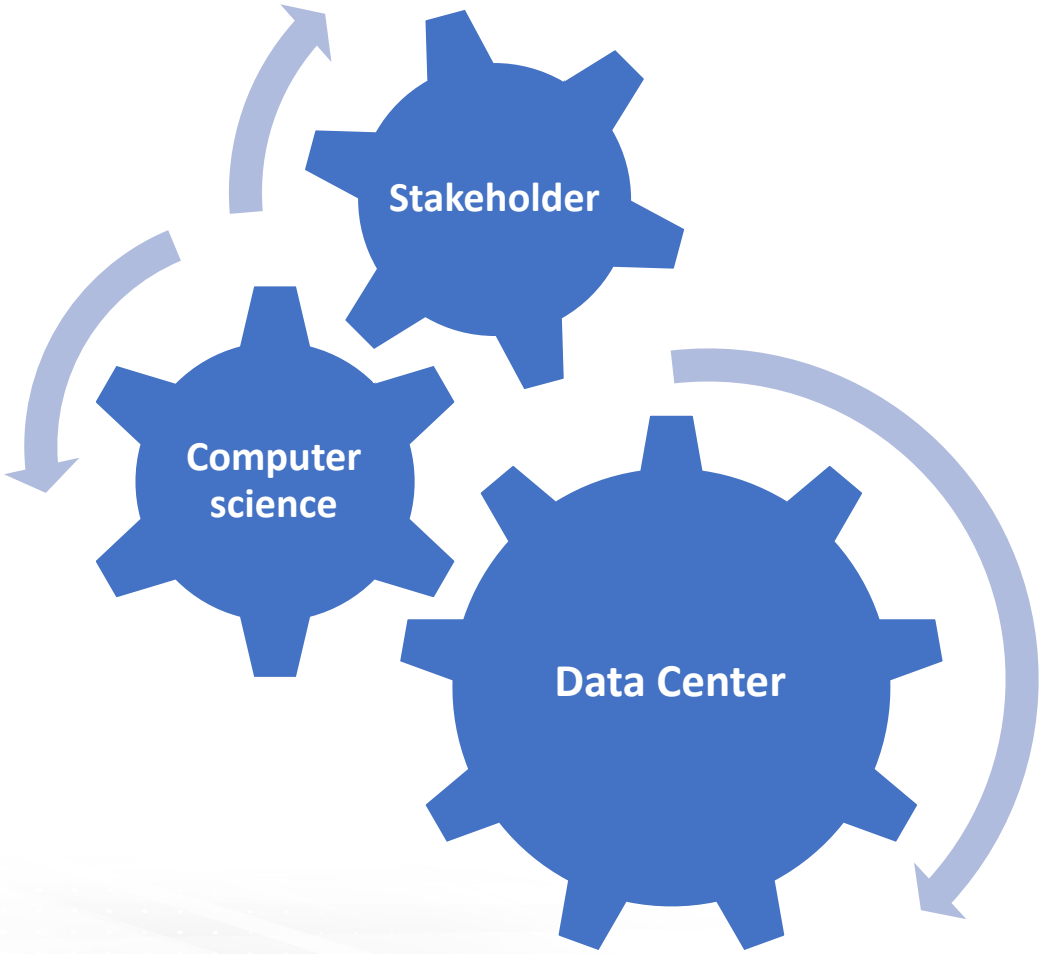


Heterogeneity of resources & User communities

- Data Center need to provide **scalable heterogenous** resources (today: HPC & ML hardware)
  - In support for new computing paradigms
  - For development, code upscaling, production
- **Agility** to adapt to new user requirements and use cases will be critical
- A strong **science engagement** is also required
  - To gather and anticipate evolving requirements (planning)
  - To guide and propose solutions (optimisation of resources)

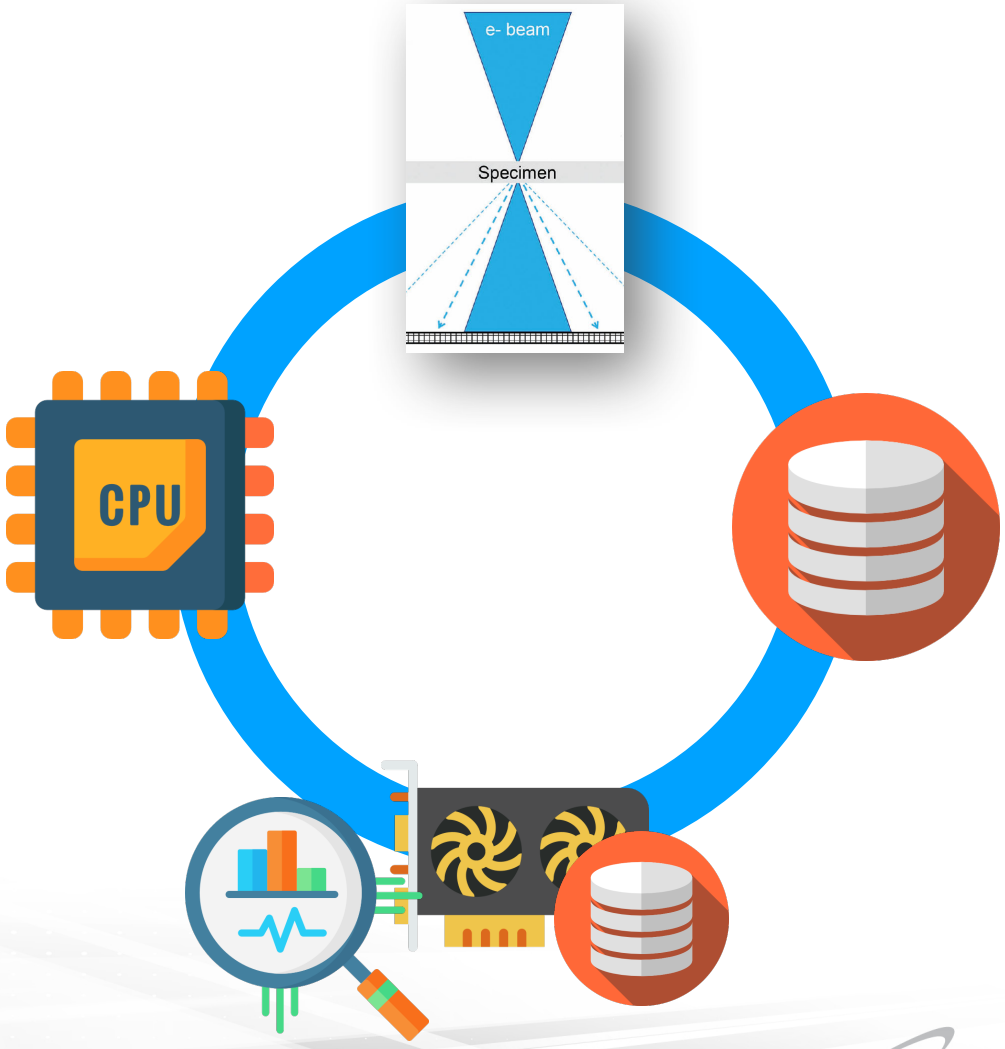
Data Center: an embedded system

- The Data Center cannot be an isolated entity
- The Data Center needs close relationship with Stakeholders and Computer Scientists
  - Ideally Computer Scientists should belong to same organization
  - Data Center staffs need to be embedded into major projects
- Success comes from collaboration between these 3



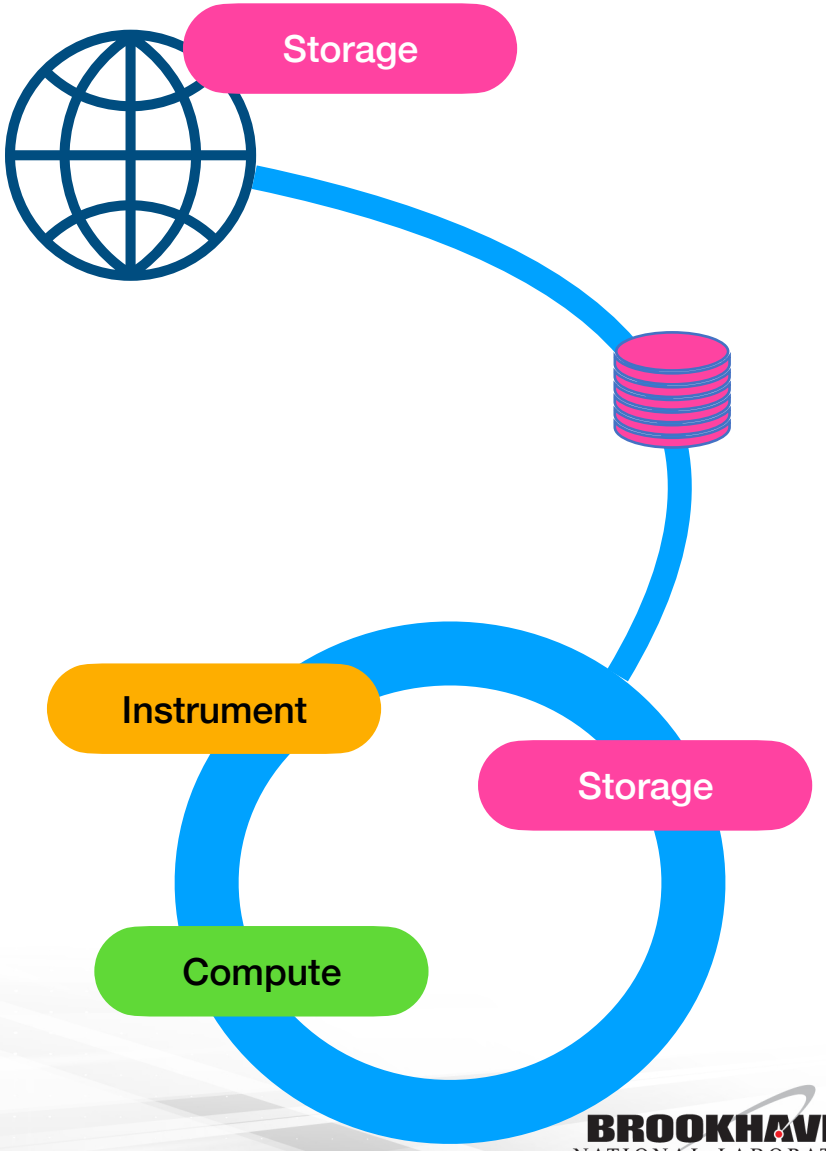
Streaming & Edge computing

- Integration of Data Center and instruments through the network
  - Instruments generating GB/s of data to timely process and analyse
- Advantages for projects/instruments:
  - Better hardware
  - Reliability & robustness
  - Expertise
  - Scalability
  - Low latency
  - Cost



Possible storage evolutions

- Need high performant (low cost) WAN capabilities
- Geographically distributed storage across Data Centers
  - ‘users’ don’t need to know where data are stored
- Caching over the network
  - Faster access, highly dependent on the data model of projects
- Gains need to be assessed



The Superfacility model

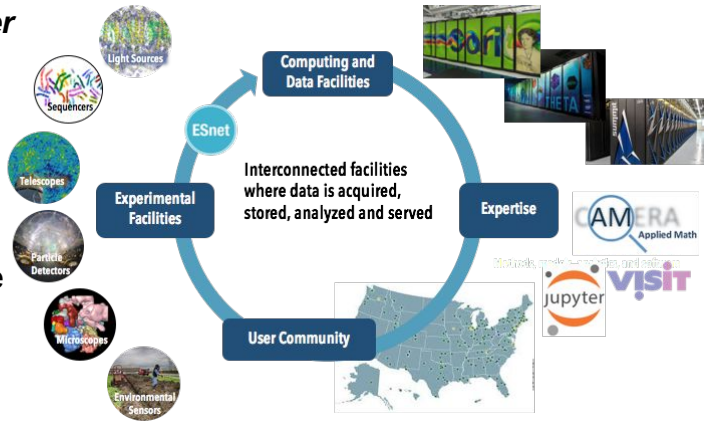
The superfacility concept developed at LBNL, is a blueprint for seamlessly integrating experimental, computational and networking resources to support reproducible science

- Instruments (experimental facilities) can also be remote
  - Collaboration with ESnet is KEY
- Somehow implemented in HEP
  - LHC distributed computing
  - Belle II remote RAW data center
- Advantages:
  - Shared infrastructure at Data center
  - Shared expertise and solutions
  - Easy access to new programs without deploying their own infrastructure

Superfacility: an ecosystem of connected facilities, software and expertise to enable new modes of discovery

Superfacility@ LBNL: *NERSC, ESnet and CRD working together*

- A model to integrate experimental, computational and networking facilities for reproducible science
- Enabling new discoveries by coupling experimental science with large scale data analysis and simulations

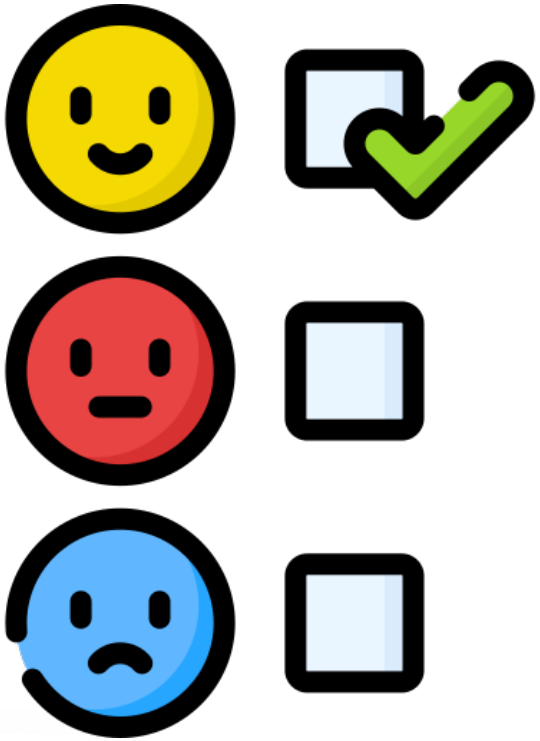


NERSC 2 BERKELEY LAB Bringing Science Solutions to the World U.S. DEPARTMENT OF ENERGY Office of Science

https://www.nersc.gov/research-and-development/superfacility/

Data Center: a User Facility

- Need to provide an easy access to resources to 1,000s of users from a large variety of origins (computing moving to web based access)
  - Federated Identity
- Management of 1,000s of users requires
  - Authorization management and related tools
  - User support
- Support for projects
  - Digital repositories
  - Collaborative tools
- And most of all user feedback



The data center of the future - I

- An energy efficient multi-program facility
  - To acquire a ‘critical’ mass to be able to follow technological evolutions and
  - Leverage economies of scale and reduce operation costs
- Need to propose high level services
  - The added value to hardware operation
  - Cannot compete directly with large HPC centers, Cloud providers…
- Closer to instruments
  - Literally: integrated with instruments to meet streaming requirements
  - Virtually: serving remote instruments

The data center of the future - II

- A user facility
  - Provide services and support for 1,000s of users
  - Services for dat and knowledge preservation
- A hub for developments and investigations
  - In close contact with computer scientists & researchers
- Need to establish collaboration with other data centers
  - To complement offered list of services and
  - Leverage on collaborations for developments and knowledge sharing

This will not be the data center of the future

Summary

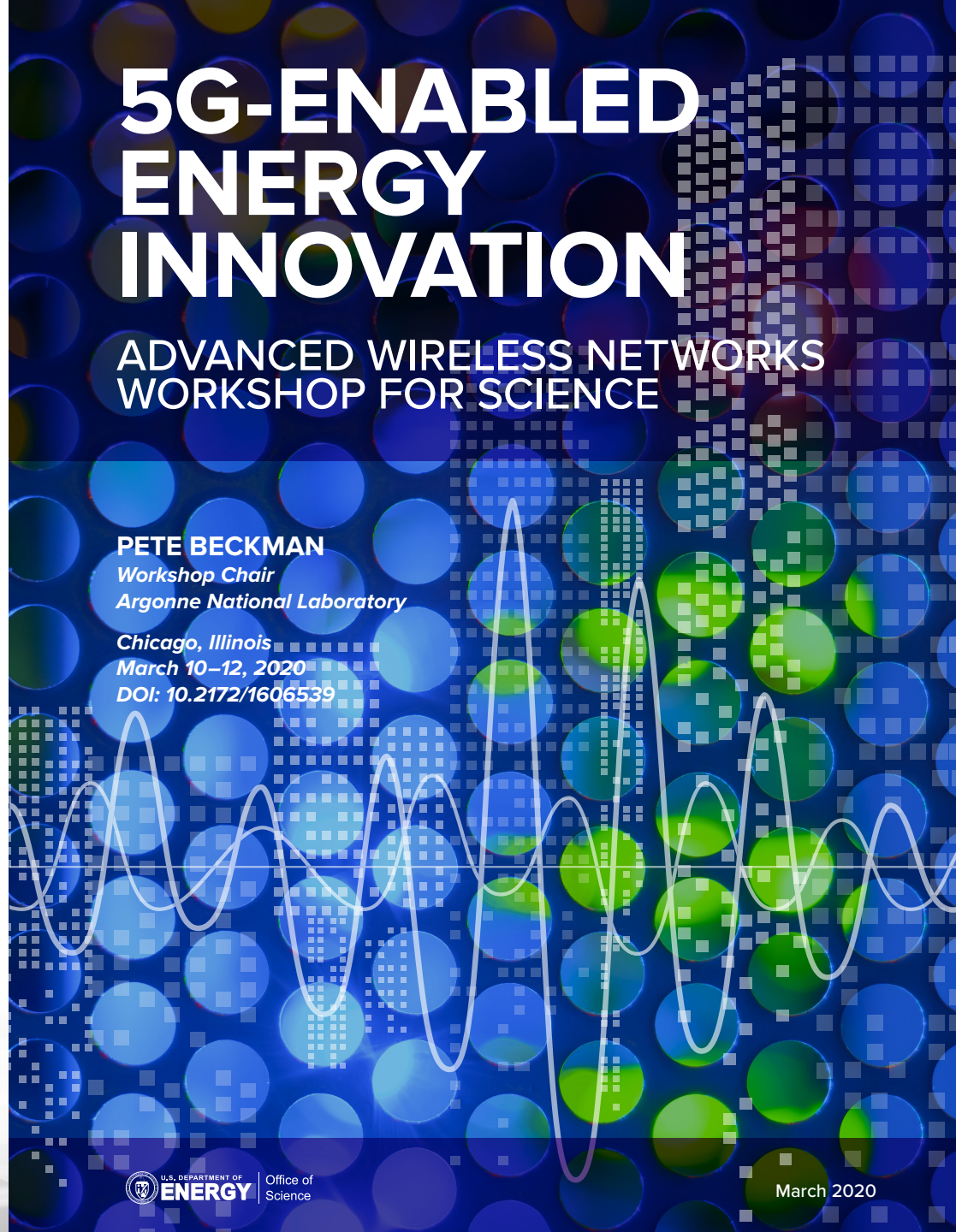
- A challenging area…
- Evolving hardware and requirements driven by compute evolution and emergence of AI
- The path for a Data Center is multi-programs and agility

Acknowledgments

Jérôme Lauret Shigeki Mizawa

Topics not addressed

- IoT
- 5G
- …



Support for the ‘historical’ NP & HEP programs

24/7 availability

- **The RHIC Tier 0**
  - Store and process RAW data from RHIC experiments
  - Provide analysis means for 1’500 users
  - Long term data preservation
- **The US ATLAS Tier 1**
  - ~25% of ATLAS Tier 1 computing capacity worldwide
  - Analysis center for US physicists (41 institutes)
- **The Belle II data center outside Japan**
  - Store 100% of RAW data
  - Operate remotely key services : databases and data management