Fermilab

U.S. DEPARTMENT OF ENERGY | Office of Science

Challenges for DOE Networking in 2025

(Fermilab perspective…)

Phil DeMar; Wenji Wu; Liang Zhang

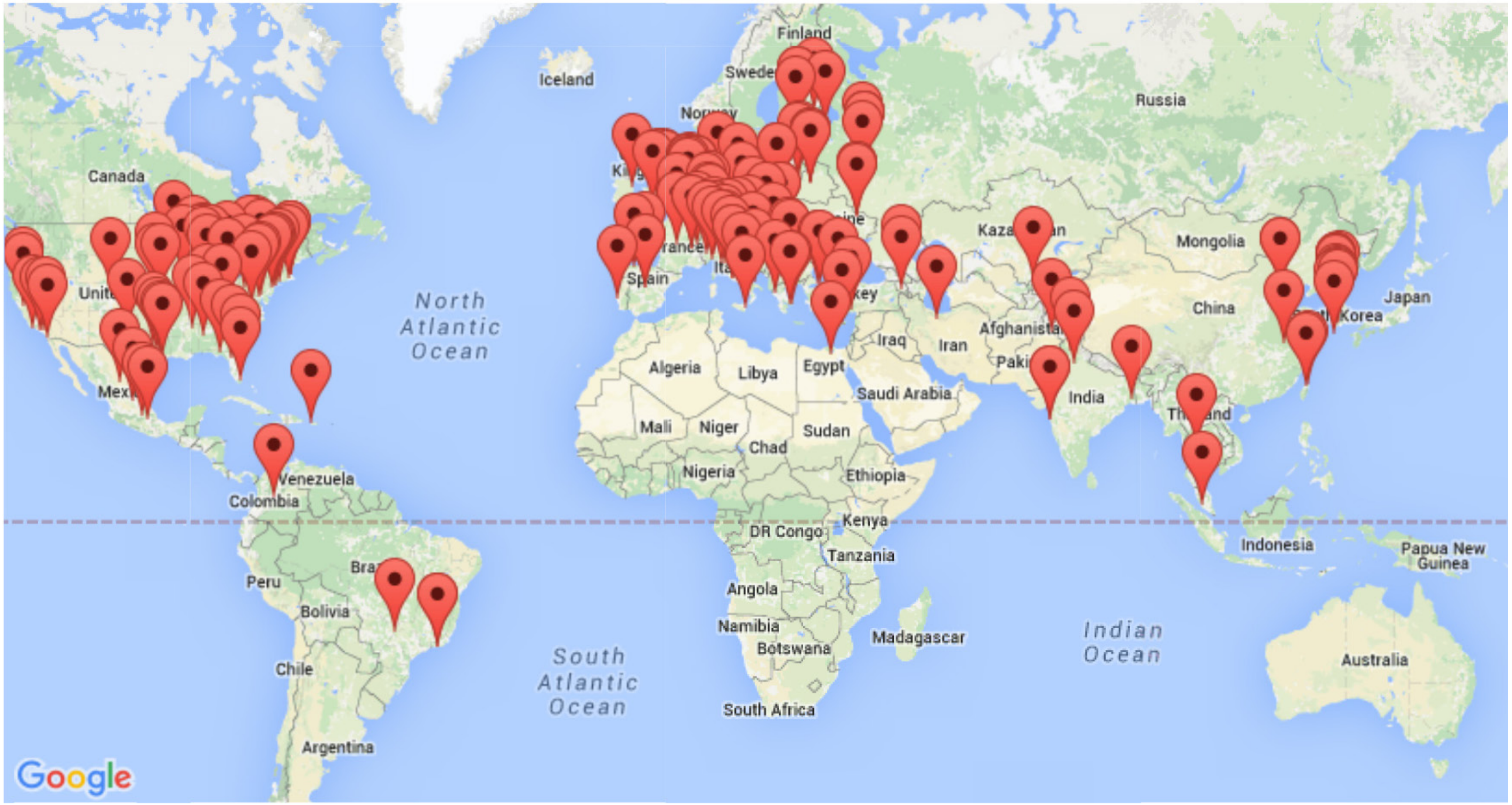
Feb. 1, 2016

Our World is Viewed thru LHC-colored Glasses

- Extreme data volumes & velocities
- Large collaborations of global scale
- Highly distributed computing environment(s):
  - Federated
- Long-lived experiments
- **(Flat budgets…)**

Distributed Computing Collaborations (CMS)

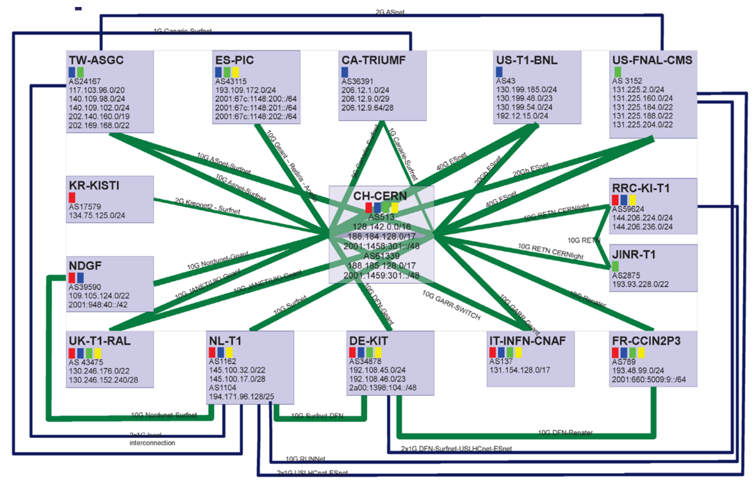
- 186 institutions (globally distributed)
  - Computing resources distributed across collaboration
  - High b/w R&E networks support experiment data movement

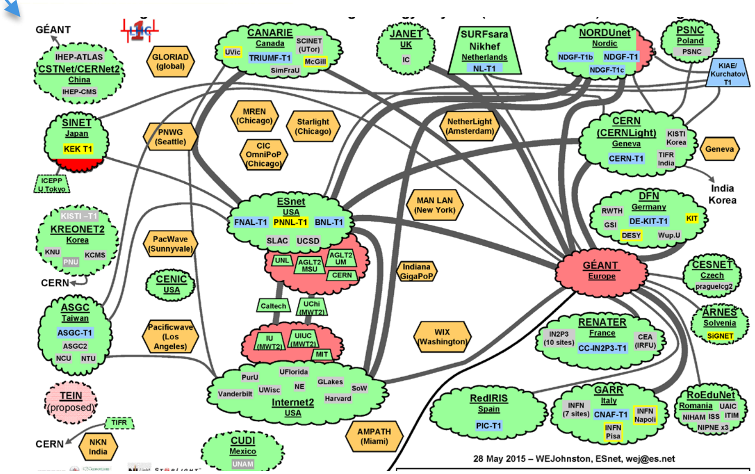


Fermilab

LHC Networking Capabilities (current & future)

- Evolving to “special” networks for LHC data movement:
  - LHCOPN (2006):
    - For raw experiment data
  - LHCONE (2013):
    - For derived/user data
  - Ensures adequate bandwidth
  - Better security risk profiles





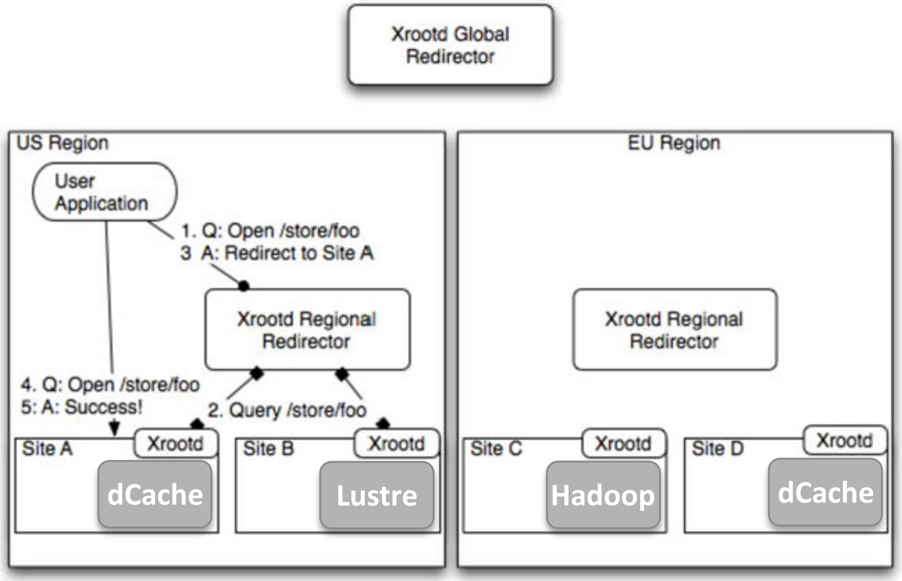
- Network technology trends:
  - Bandwidth technology step-ups
  - Overlay networks
  - Pt-2-Pt network services
  - Software-defined networks (SDN)
  - Content-defined networks (NDN?)

Fermilab

LHC Data Federation(s)

- Federated data storage, based on:
  - High bandwidth WAN connectivity across all tiers
  - Global data namespace(s)

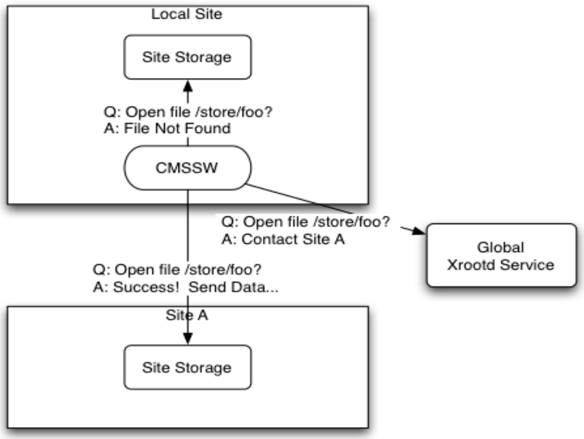
- Implemented with XrootD:
  - "Hides" local file storage systems
  - Maintains catalog of known file location



- High-level philosophy: remote storage ~= local storage

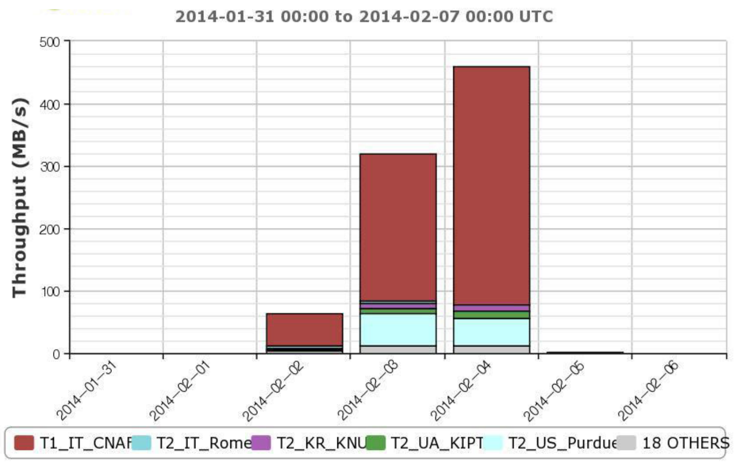
Federated Data Enables More Robust & Dynamic Distributed Computing Environment

- Job unable to access local data:
  - Remote copy of data retrieved
  - Job is able to complete…



- Useful in redirecting jobs to other sites in overflow situations

- Real life example:
  - DB error results in "missing" local data at FNAL
  - Job failover capability locates replica at CNAF (Italy)
  - Jobs run for 2 days using CNAF data, without anyone noticing…



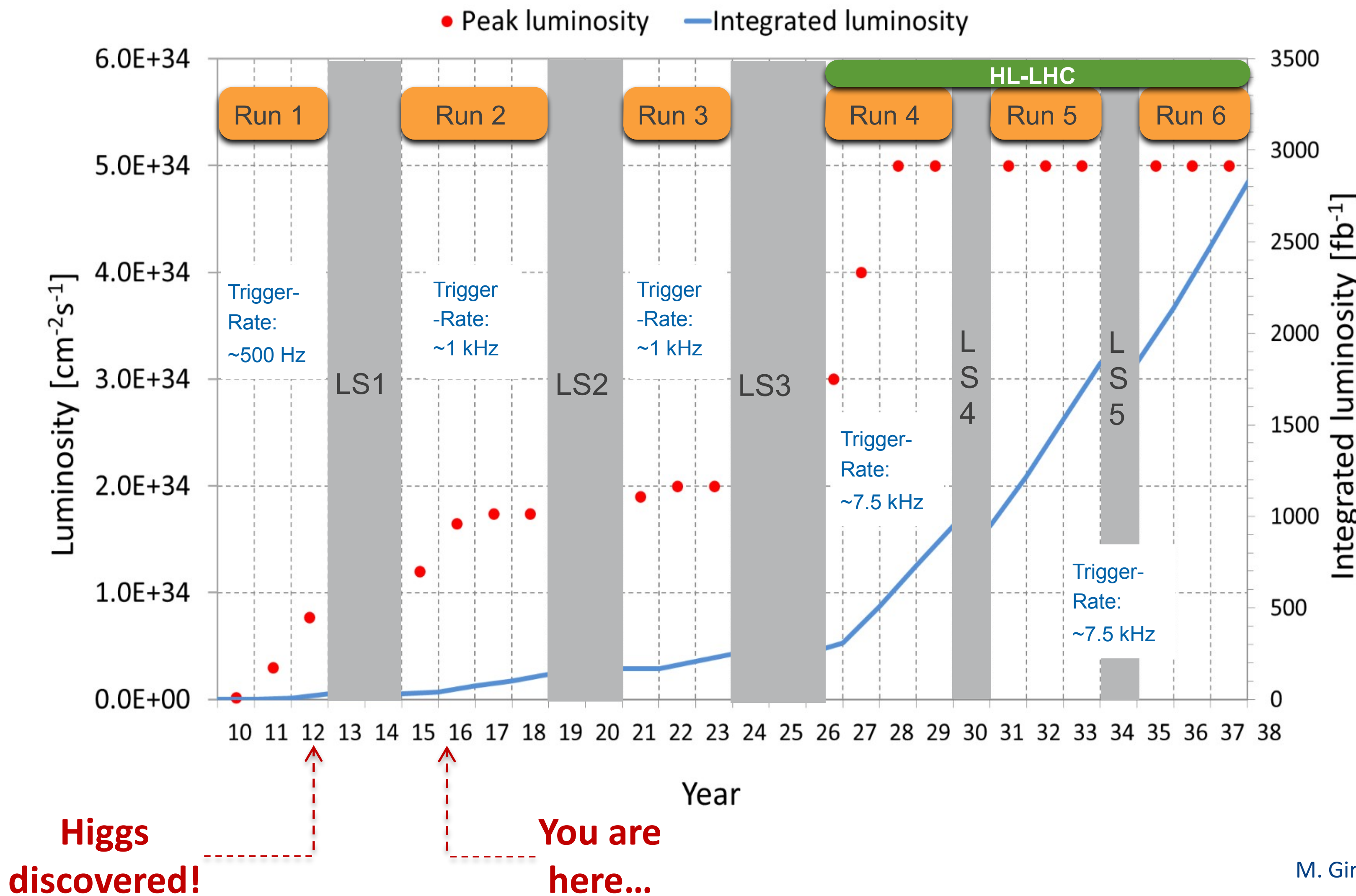
Emerging Trends to Address Computing Challenges

- Dynamic data placement
  - Distributing/redistributing (abbreviated) data sets by popularity
  - Subset of larger trend for dynamic data management in general
- Cloud & High Performance Computing (HPC) cycles:
  - Amazon Web Service spot CPU cycles already highly economic
  - Next gen. super computers will have massive computing power

M. Ernst (BNL)

System attributes	NERSC Now	OLCF Now	ALCF Now	NERSC Upgrade	OLCF Upgrade	ALCF Upgrades	
Name Planned Installation	Edison	TITAN	MIRA	Cori 2016	Summit 2017-2018	Theta 2016	Aurora 2018-2019
System peak (PF)	2.6	27	10	> 30	150	>8.5	180
Peak Power (MW)	2	9	4.8	< 3.7	10	1.7	13
Total system memory	357 TB	710TB	768TB	~1 PB DDR4 + High Bandwidth Memory (HBM) +1.5PB persistent memory	> 1.74 PB DDR4 + HBM + 2.8 PB persistent memory	>480 TB DDR4 + High Bandwidth Memory (HBM)	> 7 PB High Bandwidth On-Package Memory Local Memory and Persistent Memory
Node performance (TF)	0.460	1.452	0.204	> 3	> 40	> 3	> 17 times Mira
Node processors	Intel Ivy Bridge	AMD Opteron Nvidia Kepler	64-bit PowerPC A2	Intel Knights Landing many core CPUs Intel Haswell CPU in data partition	Multiple IBM Power9 CPUs & multiple Nvidia Voltas GPUS	Intel Knights Landing Xeon Phi many core CPUs	Knights Hill Xeon Phi many core CPUs
System size (nodes)	5,600 nodes	18,688 nodes	49,152	9,300 nodes 1,900 nodes in data partition	~3,500 nodes	>2,500 nodes	>50,000 nodes
System Interconnect	Aries	Gemini	5D Torus	Aries	Dual Rail EDR-IB	Aries	2nd Generation Intel Omni-Path Architecture
File System	7.6 PB 168 GB/s, Lustre®	32 PB 1 TB/s, Lustre®	26 PB 300 GB/s GPFS™	28 PB 744 GB/s Lustre®	120 PB 1 TB/s GPFS™	10PB, 210 GB/s Lustre initial	150 PB 1 TB/s Lustre®

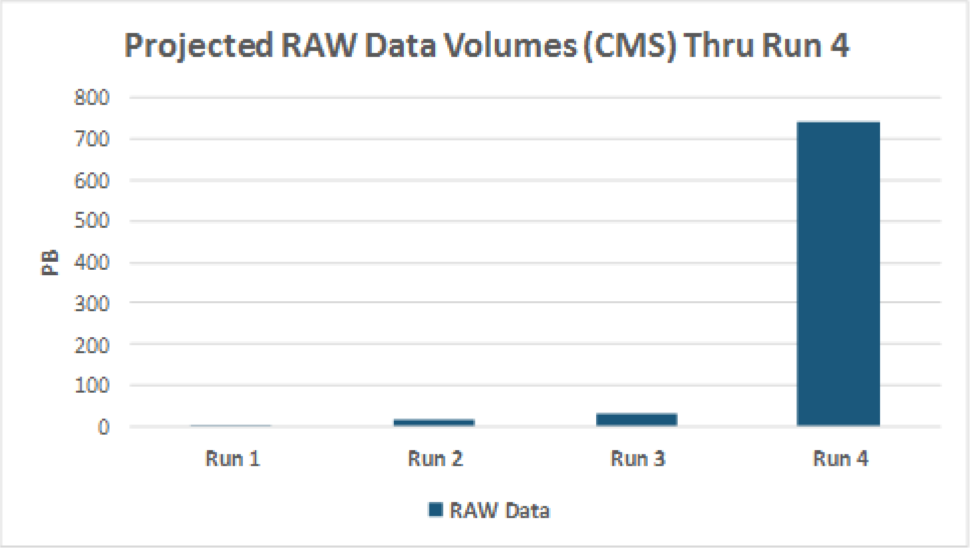
LHC schedule



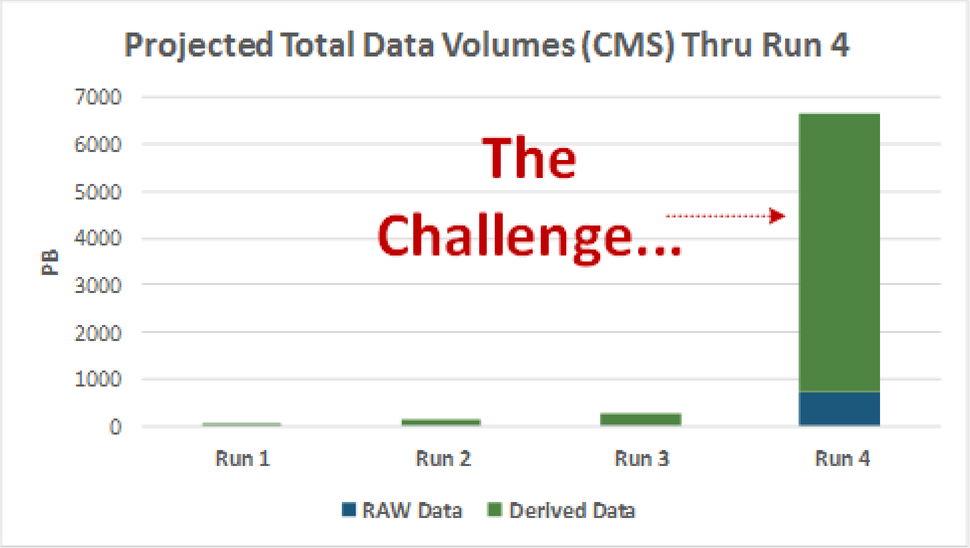
M. Girone (CERN)

Projected LHC data volumes

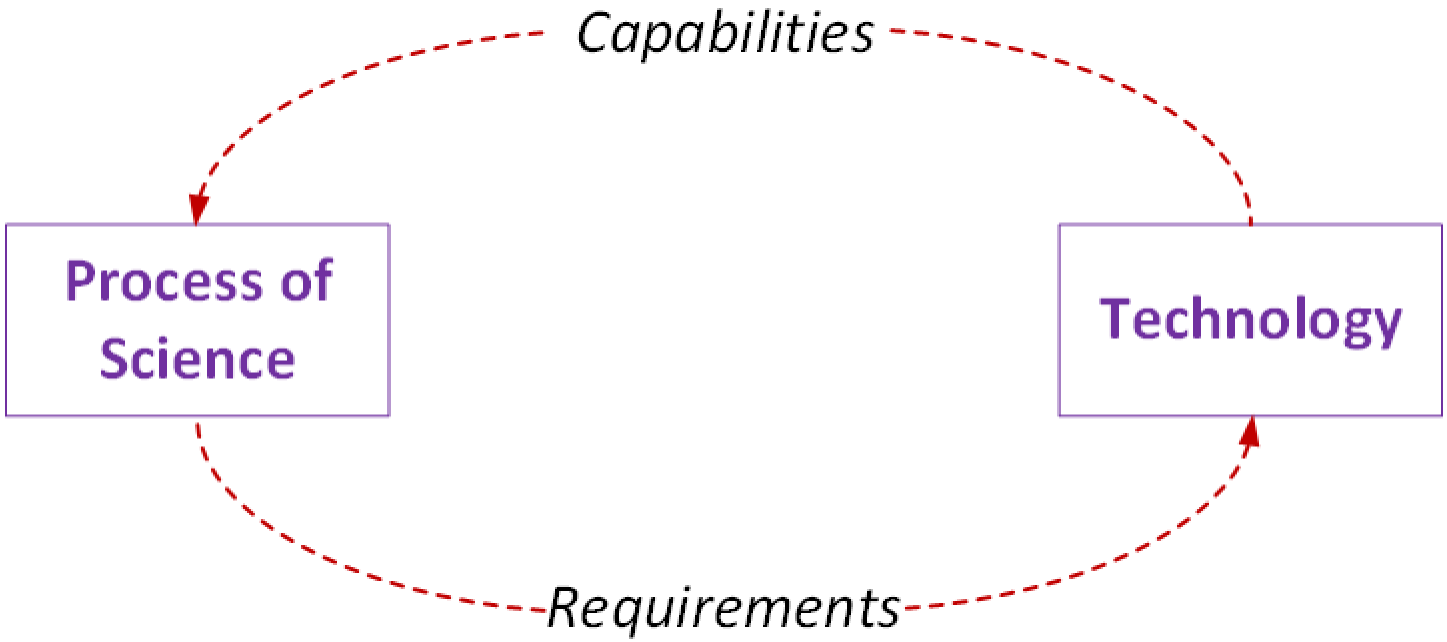
- Raw data = generated by detector(s)



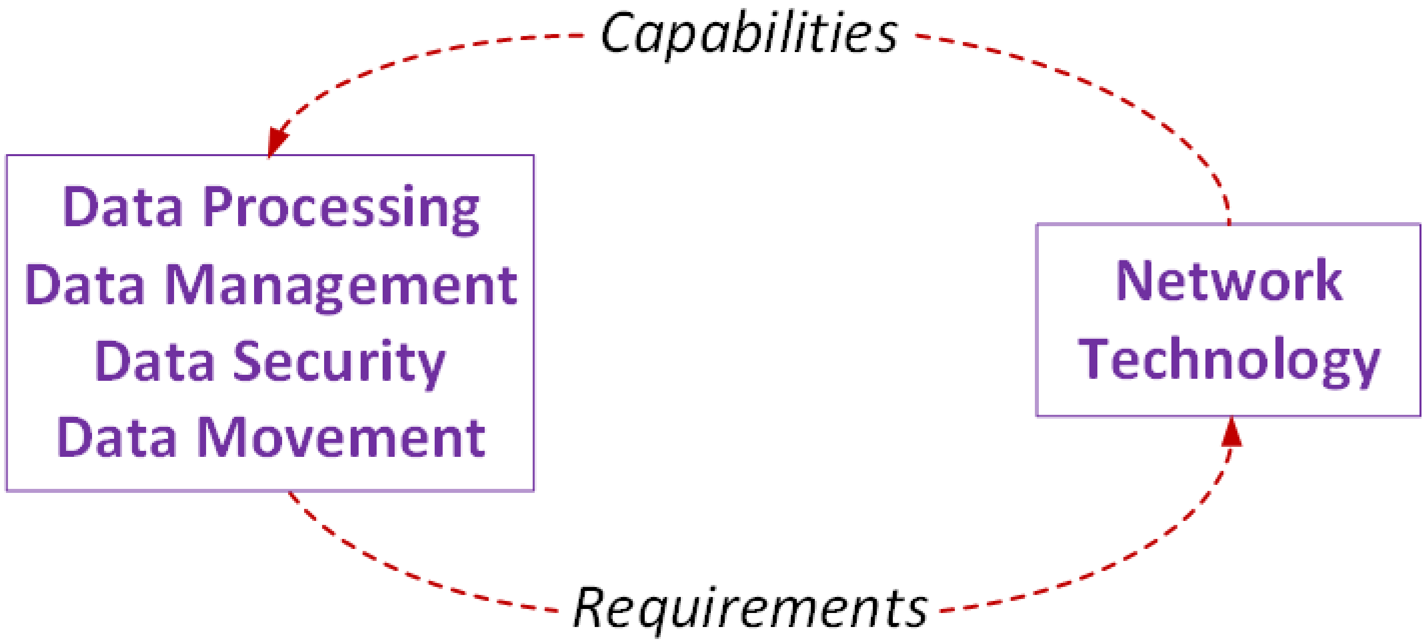
- Derived data = reconstructed data, simulation data, summary data sets, etc…)
  - (derived data) ~= (raw data) x 8



Looking at Things from a High Level



Applying This to DOE Networking 2025…



1] What will large-scale science data demand from the network in 2025?

2] How might emerging network technologies transform large-scale science data handling?

Musing on 2025 Science Requirements vs Network Technology Capabilities [Speculative Opinion]

2025 Large-Scale Science Data Requirements

- Highly dynamic & secure distributed computing environments
  - Including HPC & cloud
- Extreme data movement capabilities across these environments

2025 Network Technology Capabilities (?)

- Very high b/w networks & system NICs
- Customizable, dynamic network capabilities (SDN)
- Content delivery network services (NDN)
- Data movement orchestration services of system, storage, & network resources
- Optimization of multicore / manycore capabilities for data movement on end systems