CSE 5449: Intermediate Studies in Scientific Data Management

Lecture 2: Software and hardware stacks of storage and data management

Dr. Suren Byna

The Ohio State University

E-mail: byna.1@osu.edu

https://sbyna.github.io

01/12/2023

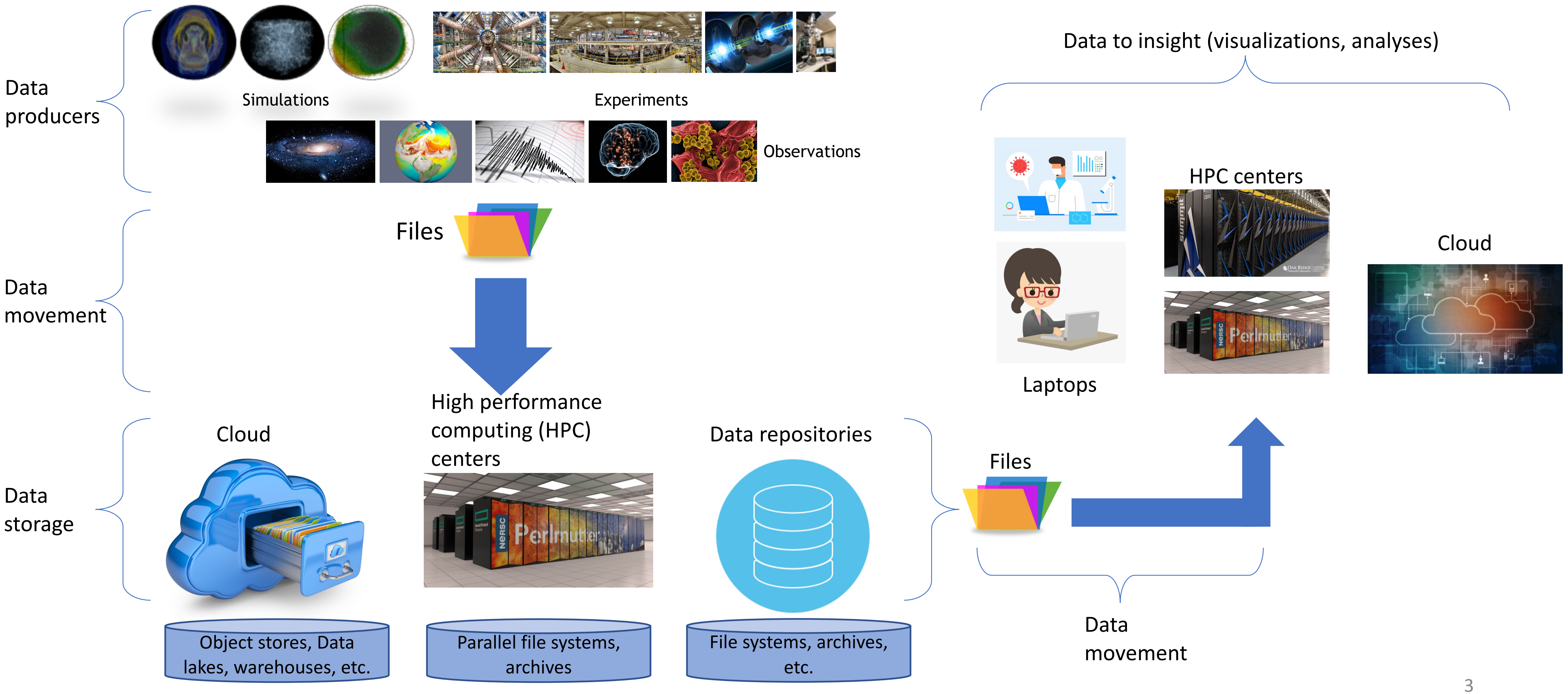
Class logistics - Grading plan

Grading component	%
Attendance, participation in class discussions, and evaluation of class presentations	15%
Class presentations	20%
Final exam (comprehensive, open book)	25%
Class project	40%

Class logistics - Class schedule (tentative)

Week	Topic	Presenter	Notes / Due dates
1 (1/10 & 1/12)	Intros & Data life cycles	Suren Byna	Project topics - Provided by Prof
2 (1/17 & 1/19)	Software and hardware stacks of storage and data management	Suren Byna	Project topics – Discuss and select
3 (1/24 & 1/26)	I/O libraries	Suren Byna	Discuss a project initial plan w/ Prof
4 (1/31 & 2/02)	File and data management systems	Suren Byna	Project proposal due – 1/31
5 (2/07 & 2/09)	Parallel I/O Stack performance tuning	Suren Byna	
6 (2/14 & 2/16)	Performance understanding, bottlenecks, and tuning	Suren Byna & Guest	
7 (2/21 & 2/23)	Knowledge management - metadata and provenance	Suren Byna	Discuss project progress w/ Prof
8 (2/28 & 3/02)	Student presentations - related research, gaps, proposal	Students	Discuss project progress w/ Prof
9 (3/07 & 3/09)	Student presentations - related research, gaps, proposal	Students	
10	**Spring Break - No class**		
11 (3/21 & 3/23)	Designing next-gen data management systems for science	Suren Byna	
12 (3/28 & 3/30)	SDM - Research gaps and challenges	Suren Byna	Discuss project progress w/ Prof
13 (4/04 & 4/06)	Student project presentations - Progress reports	Students	
14 (4/11 & 4/13)	Scientific data discovery, data quality, etc.	Guest	Discuss project progress w/ Prof
15 (4/18 & 4/20)	Student project presentations - Final report outs	Students	
16 (4/25 & 4/27)	Final Exam & Recap / Guest lecture	Suren Byna / Guest	Final Exam on 4/25 (to be confirmed)

Last class - Data life cycle - An overview

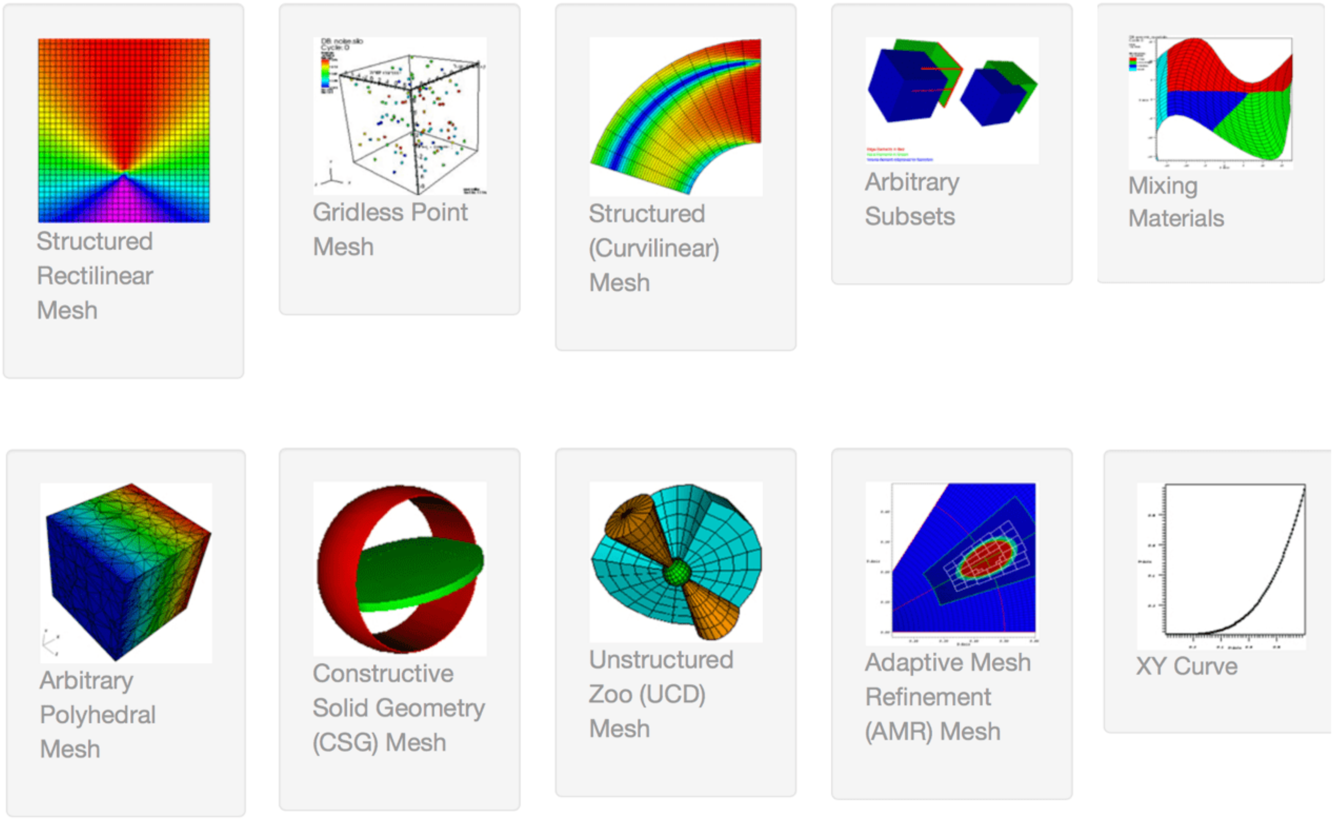


Outline of today's lecture

- Characteristics of scientific data
- Hardware architectures
- Software stacks
- Challenges and opportunities
- Class projects

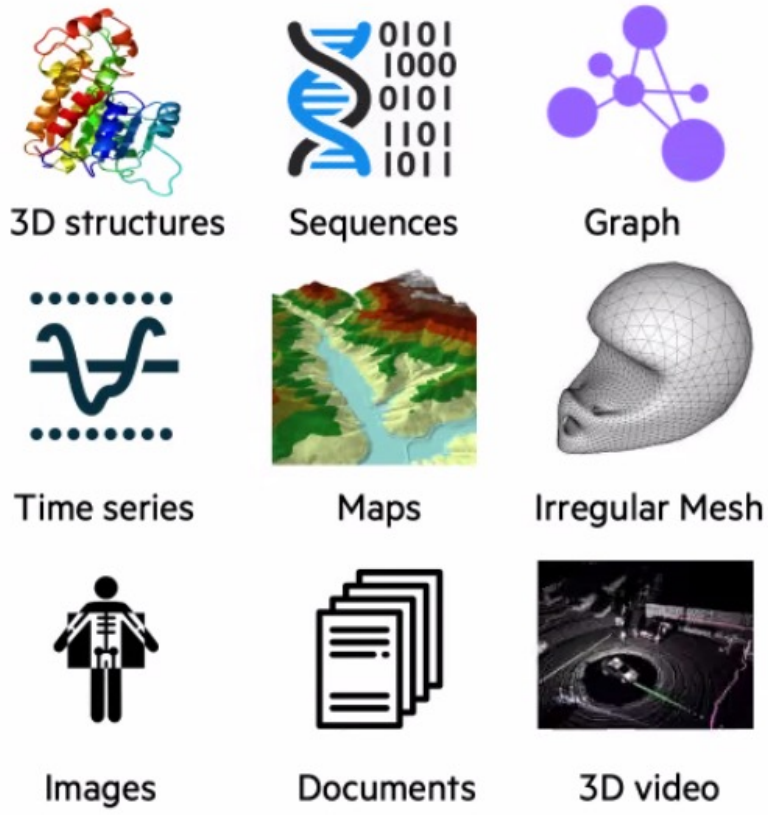
What does scientific data look like

Traditional types of data - modeling and simulation



Source: MACSIO, LLNL
https://github.com/LLNL/MACSio/blob/master/doc/scientific_data_objects.png

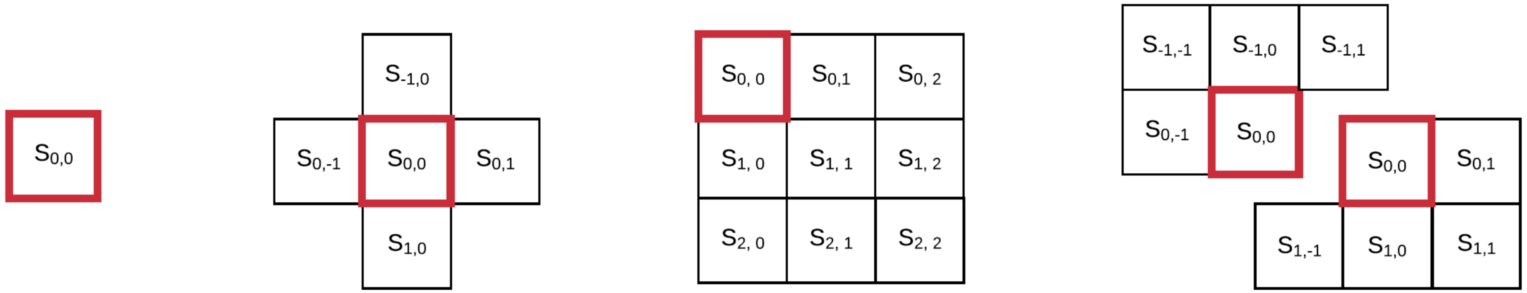
Typical data used for AI / ML



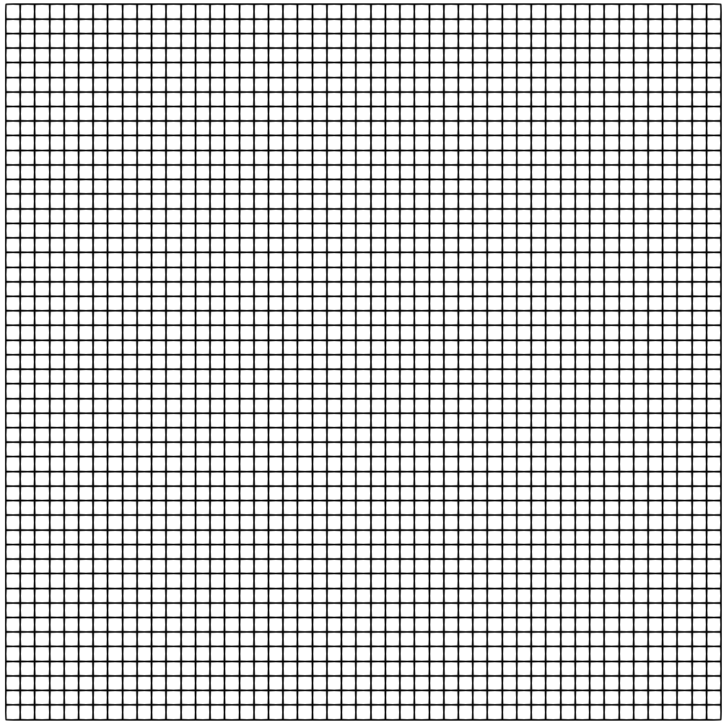
Source: Rangan Sukumar's slides presented at Monterey Data Workshop on 04/21/2022

Structured meshes

- Mesh divides a computational domain into cells
- Cells are logically arranged on a regular grid



(a) (b) (c) (d)

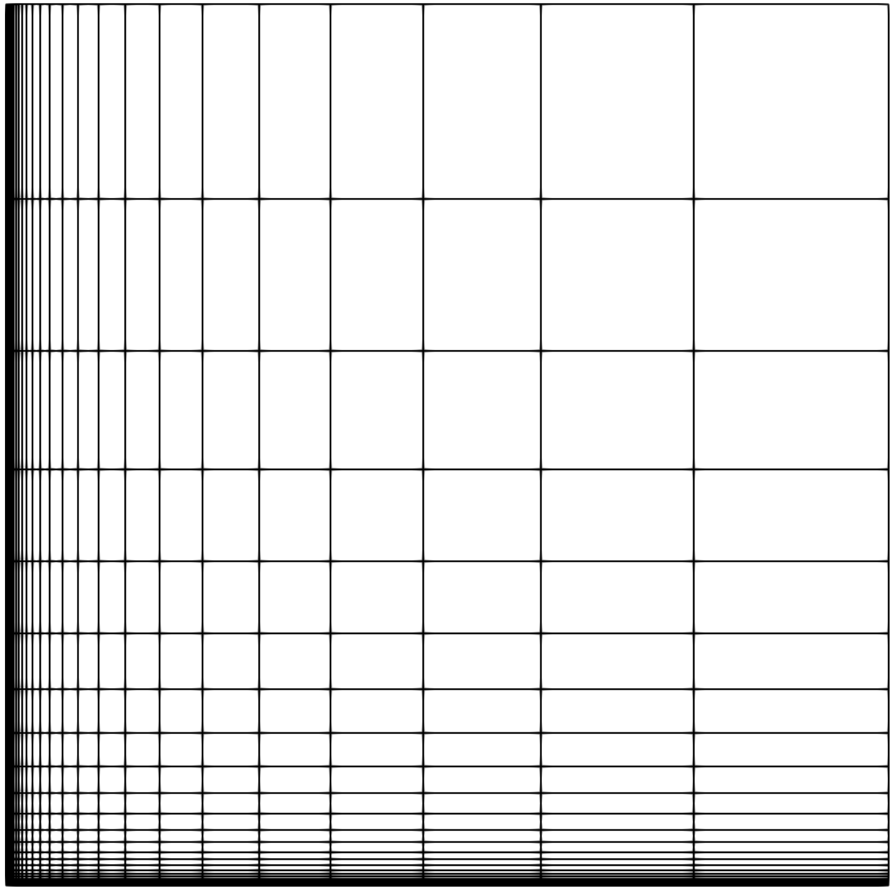


Source: Bin Dong, Kesheng Wu, and Suren Byna, "User-Defined Tensor Data Analysis", SpringerBriefs in Computer Science, DOI: https://doi.org/10.1007/978-3-030-70750-7, ISBN: 978-3-030-70750-7,

Source: https://axom.readthedocs.io/

Rectilinear meshes

- A set of rectangular cells, arranged on a regular lattice
- Nodes are aligned with the Cartesian coordinate axis
- Spacing between adjacent nodes can vary

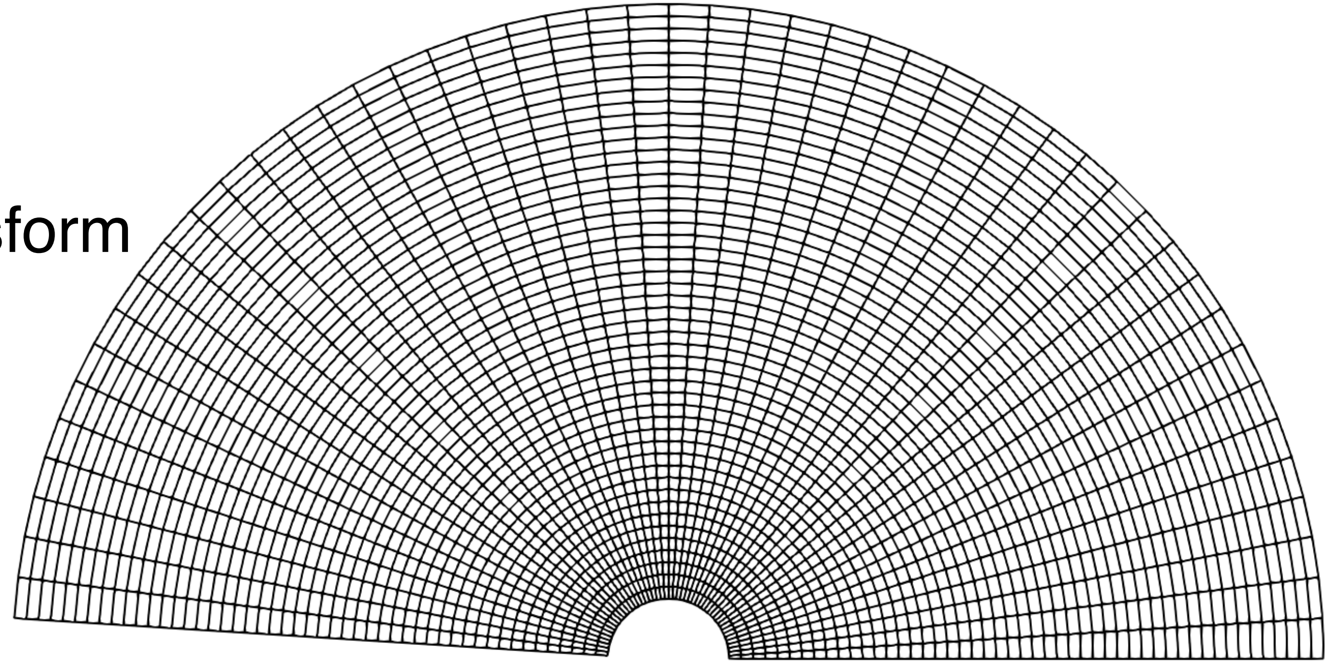


Source: https://axom.readthedocs.io/

Curvilinear meshes

- Logically a regular mesh, but the nodes are not placed along the Cartesian grid lines

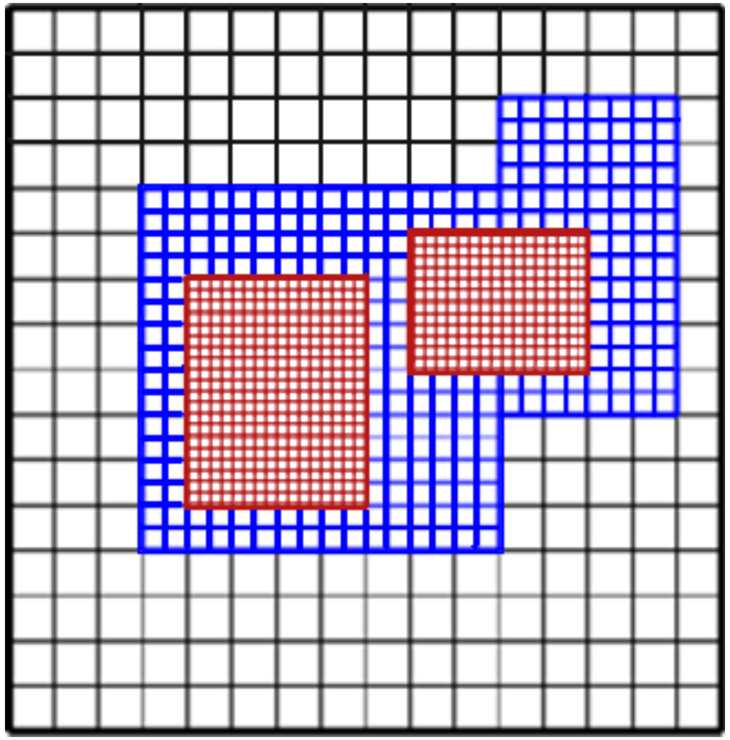
- Partial differential equations (PDE) govern the node locations that transform the Cartesian grid to a curvilinear coordinate system

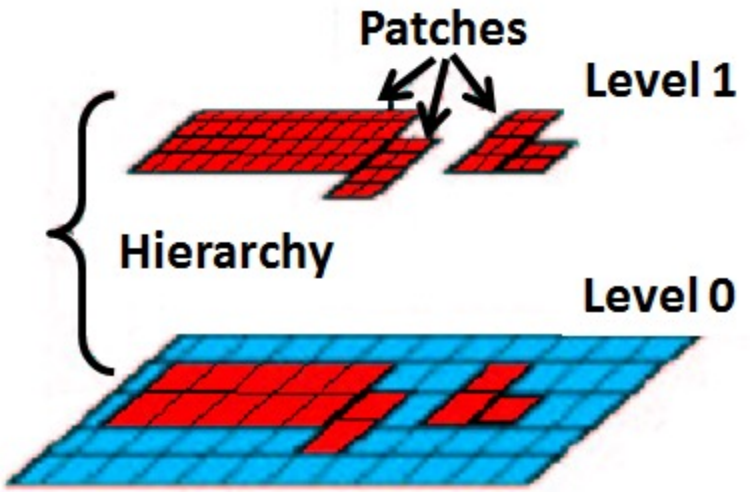


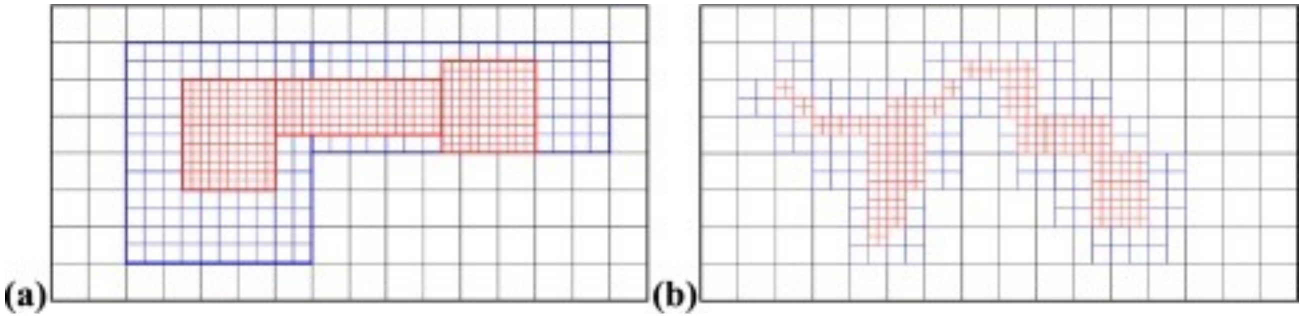
Source: https://axom.readthedocs.io/

Adaptive mesh refinement (AMR)

- AMR assigns different levels of resolution across the computation domain
- Delivers enough precision in highly dynamic areas, where needed, to achieve desired accuracy
- Memory- and computation-efficiency built into algorithms
- Block-structured, point-structured, and unstructured







Homework: Other data structures

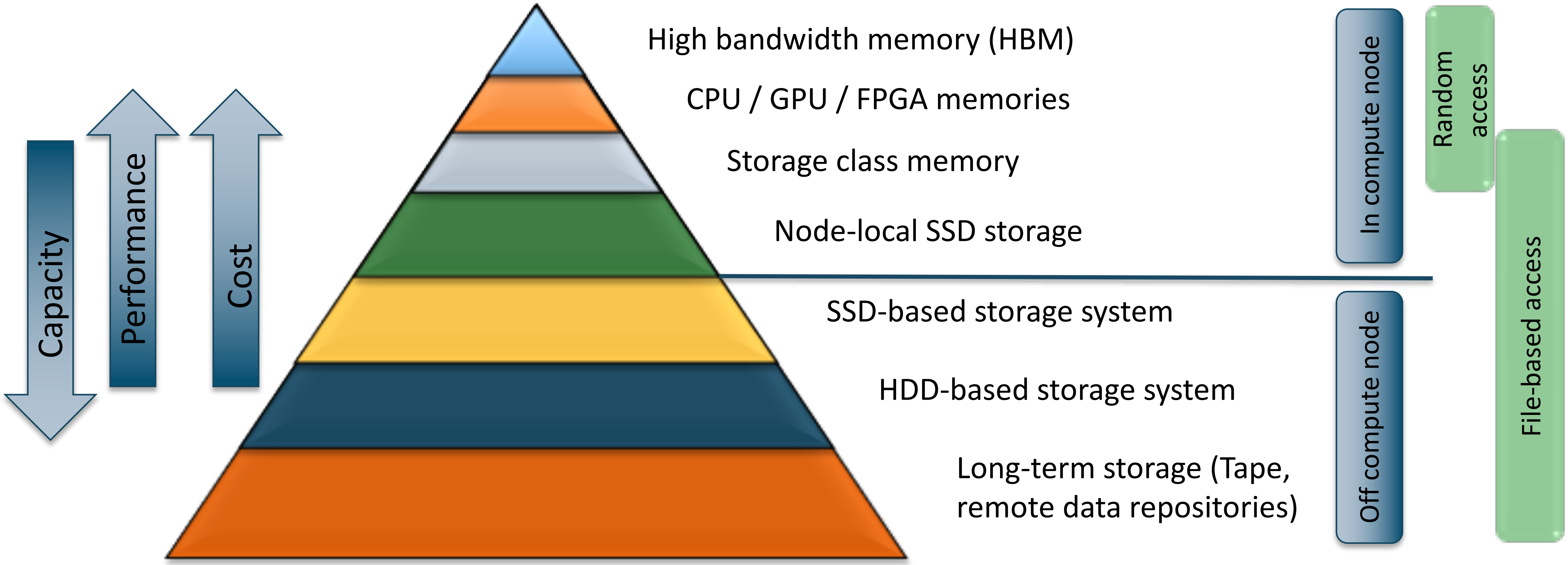
1. Graphs
2. Sequences
3. Time series
4. Irregular mesh
5. Documents
6. Media – images and videos
7. Maps
8. Grid-less point mesh

- 2 min presentation next week
- You may use slides
  - Send the slides to me an hour before the class

Characteristics of scientific data

- **Size** : ranges between a few bytes to petabytes
  - Data generation is often very large, knowledge / information is often a few bytes to megabytes
- **Common data structures**
  - Arrays (regular, irregular, variable-length, sparse, compressed, heaps, lists, etc.)
  - Text that describes the data (metadata), graphs
  - Tabular data
- **File formats**
  - ASCI, binary
  - Self-describing data formats: netCDF, HDF5, ADIOS, ROOT, FITS, etc.
  - Image file formats
  - CSV, Excel, JSON, XML, YAML
- **Cleanliness**
  - Simulations → Mostly clean
  - Experimental / observations → Dirty, requires a lot of processing to clean up to be used
- **Access**
  - Write / produce once,
  - Read a few times or never

Data storage - hardware levels

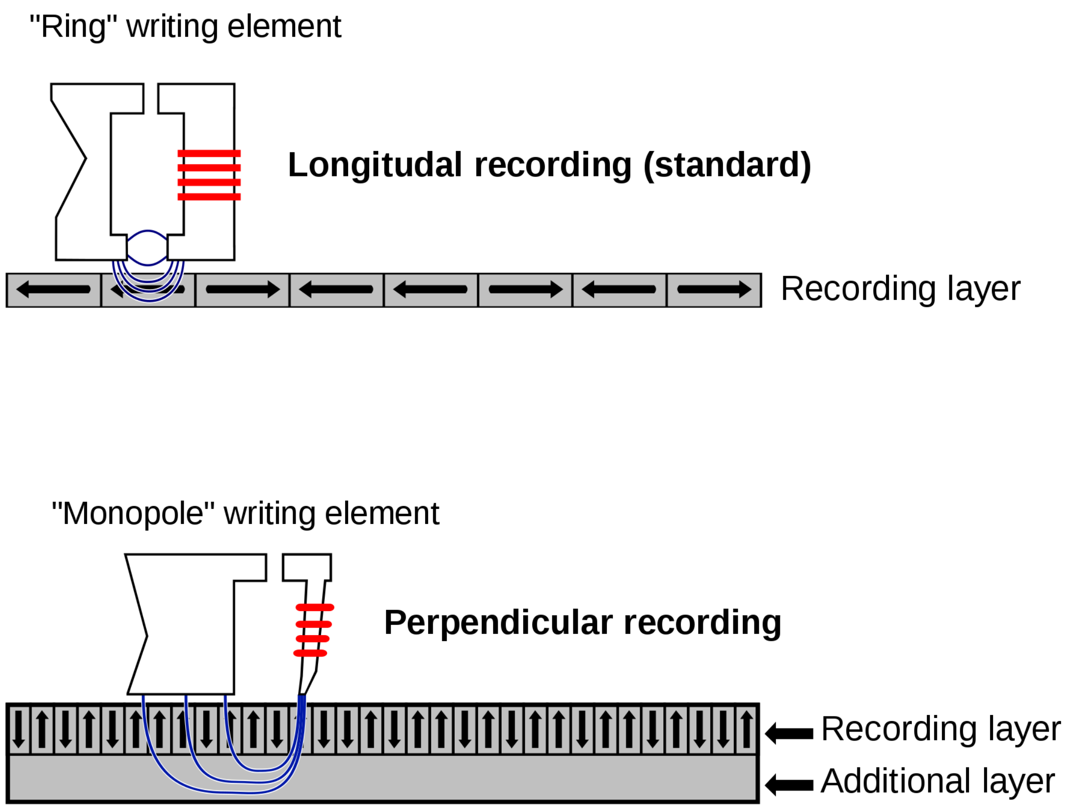


Fast memory and flash storage technologies

- High Bandwidth Memory (HBM)
  - More dense memory by stacking up to 8 DRAM dies
  - Very wide bus compared to DRAM
  - New technology progress to add processing in memory (HBM – PIM)
- Non-volatile memory (NVM)
  - Stores data even after power is removed
  - Often used as secondary storage or long-term persistent storage

Storage technologies – Magnetic disks

- Most used technology
- Arial density is being increased continuously – using superparamagnetic effect *
  - Longitudinal recording → 100 to 200 $Gbit/in^2$
  - Perpendicular recording → 1 $Tbit/in^2$
  - Future: Heat-assisted magnetic recording (HAMR) and microwave-assisted magnetic recording (MAMR)



Storage technologies – Arial density trends

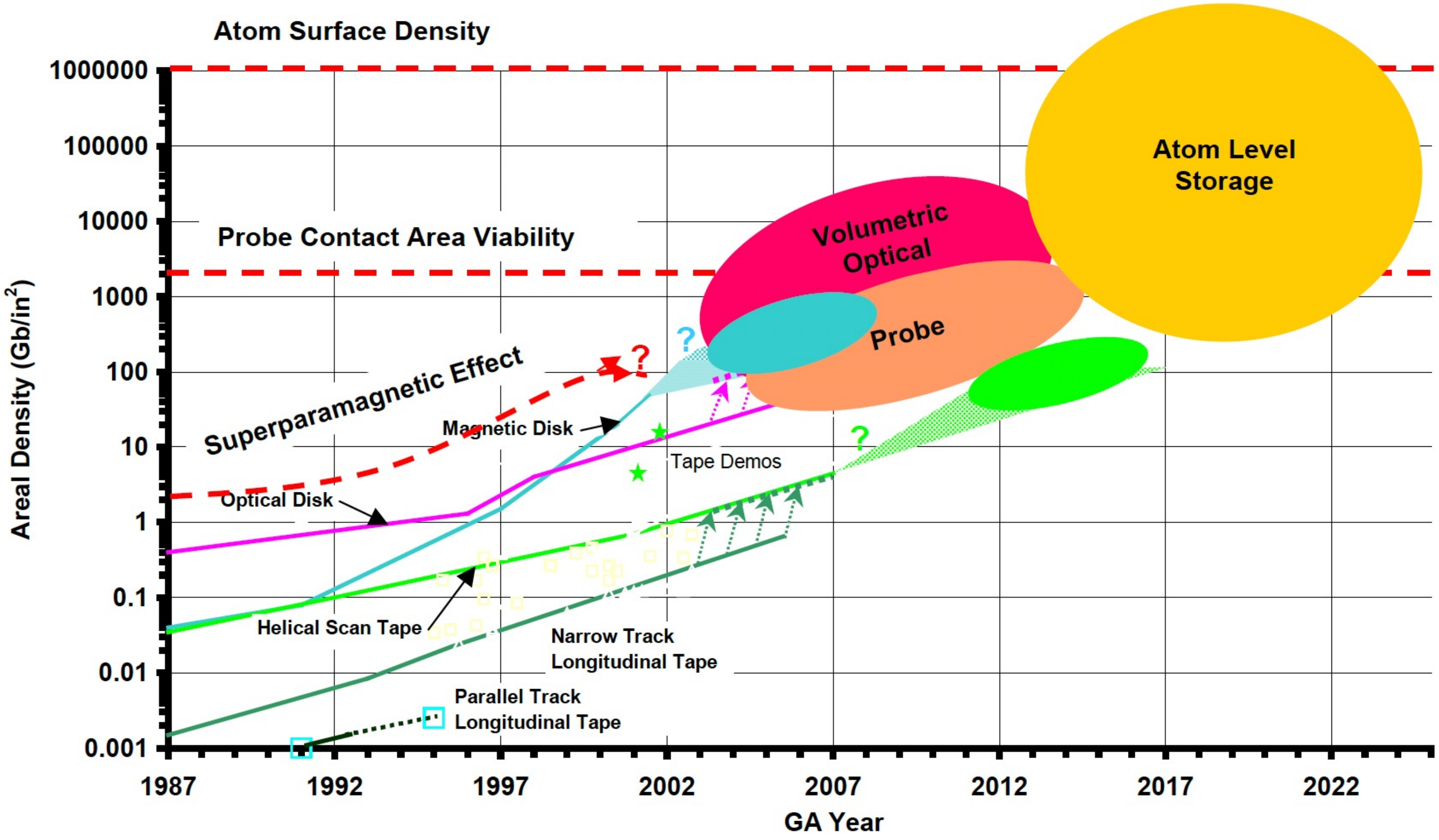
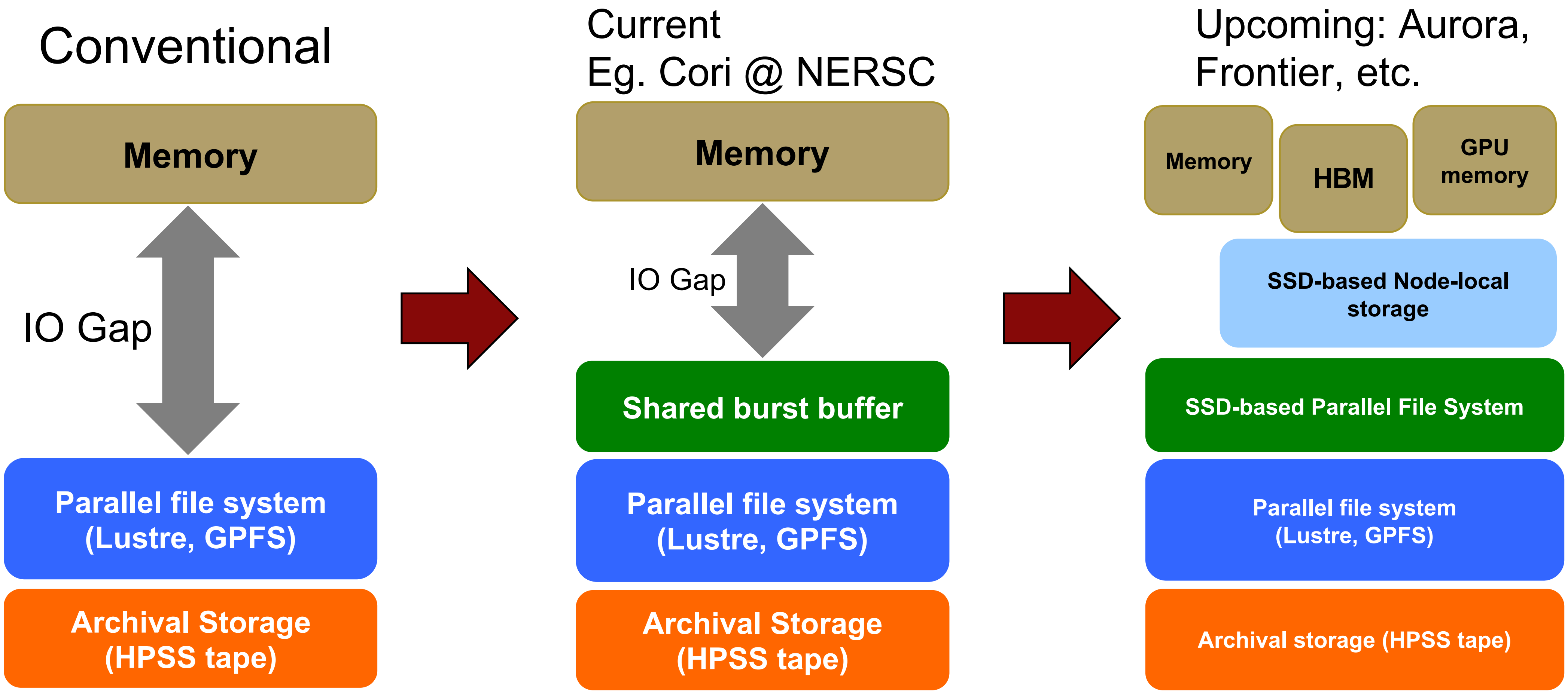


Figure Source: "Data Storage Trends and Technologies" talk by Mike Leonhardt (StorageTek) at THIC meeting

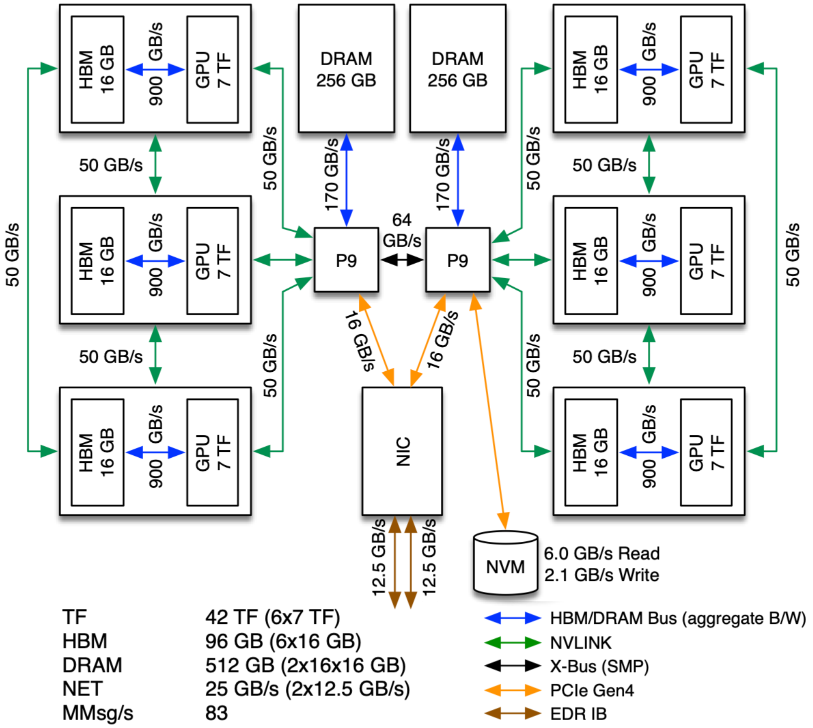
Alternative disk technologies

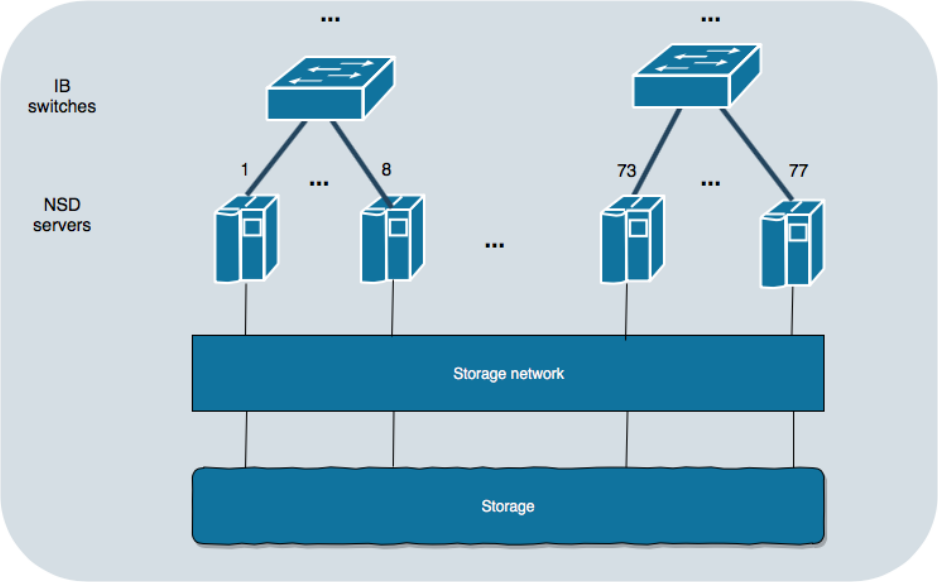
- Optical disks
  - CD, DVD, DVR, Blue-ray, HD DVD
- Holographic storage
  - Magnetic and optical data storage devices
    - Individual bits are stored as distinct magnetic or optical changes on the surface of the medium
  - Holographic data storage records information throughout the volume of the medium
  - Multiple images in the same area using light at different angles can be recorded.

Storage systems in high performance compute systems

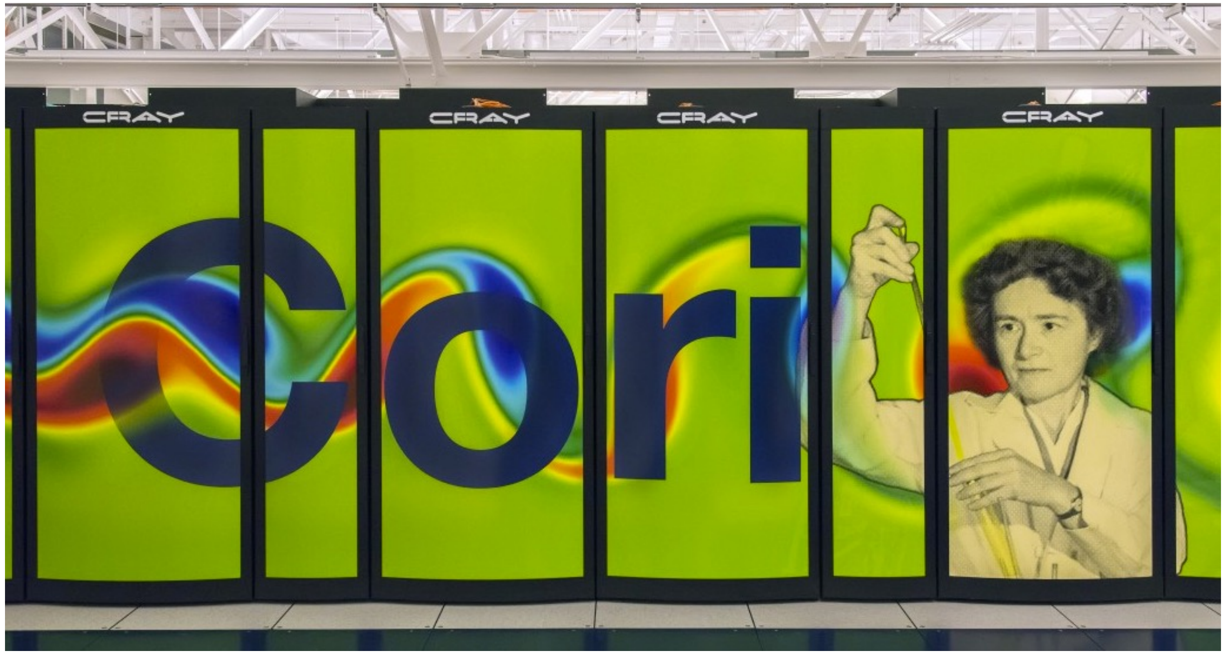


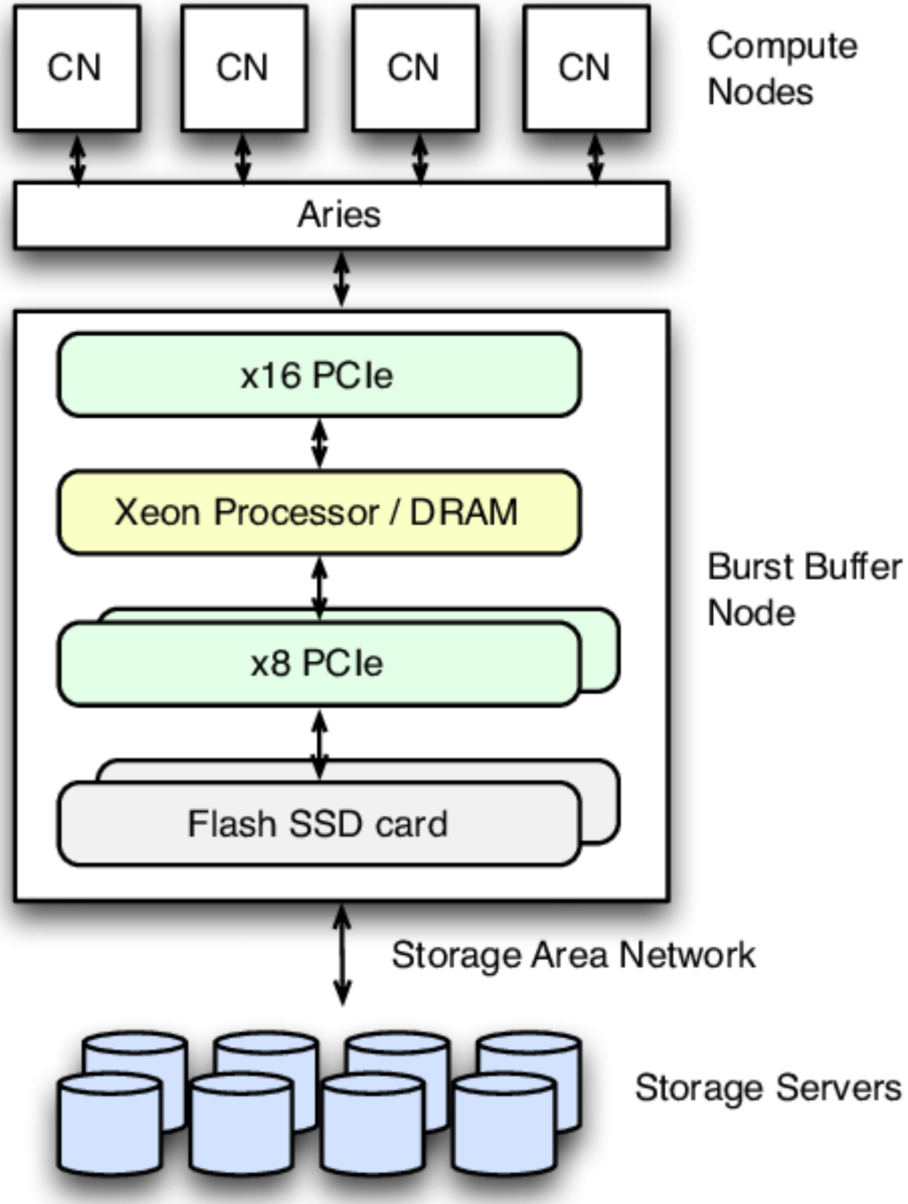
Summit system at Oak Ridge Leadership Computing Facility (OLCF)



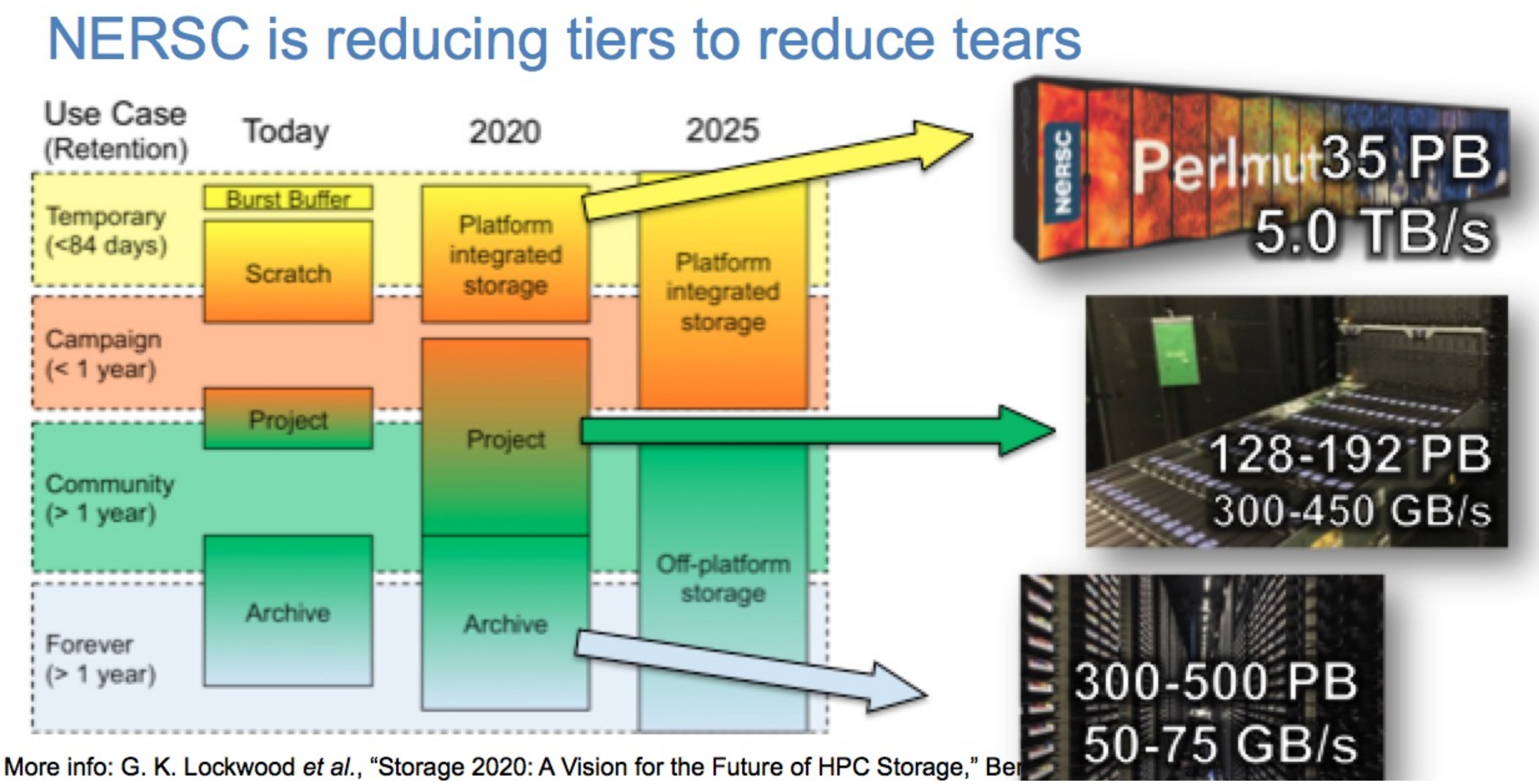


NERSC Cori system

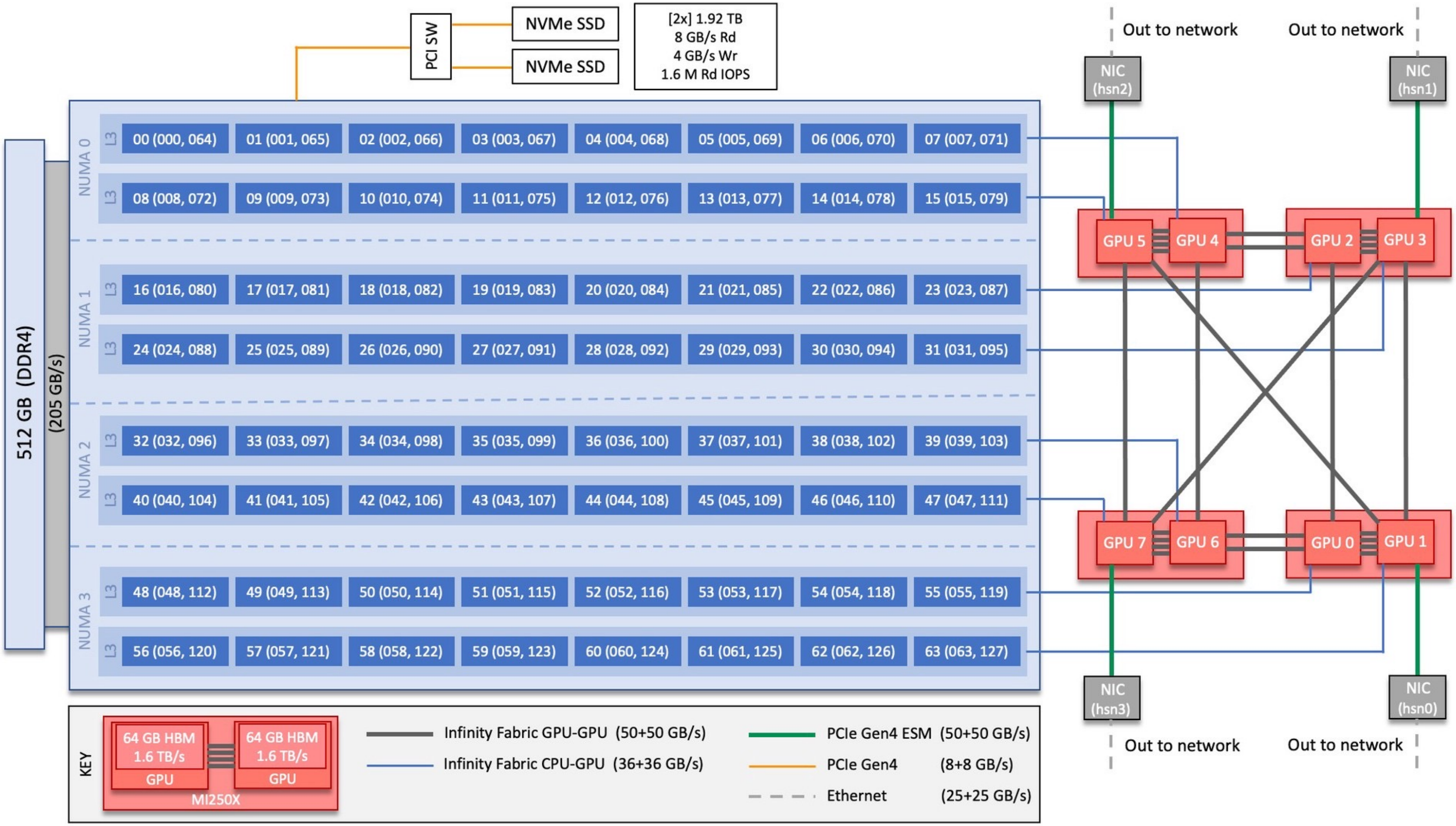




NERSC Perlmutter storage architecture



OLCF Frontier storage architecture



Class projects

1. File format comparison

- A comparison of various file formats in performing I/O operations on sequential and parallel storage systems
- Prior work
  - https://arxiv.org/pdf/2207.09503.pdf
- Deliverable: A short paper comparing performance using real scientific data

2. A retrospection of metadata standards in scientific data

- Numerous metadata standards are available
- Question: What's their readiness to be used for finding desired datasets and knowledge in massive amounts of data?
- Deliverable: A short paper with a survey of metadata standards and their usefulness / readiness for querying desired data.

Class projects

3. Performance tuning of High Energy Physics I/O benchmarks

- Question: What's the performance of a realistic use case from a high energy physics benchmark that's representative of the CMS and the ATLAS experiments (from the Large Hadron Collider data sets)
- Benchmark: https://github.com/Dr15Jones/root_serialization
- Deliverable: A short paper describing the current performance and improved performance by applying various tuning options

4. Study of parallel I/O problems and solutions/optimizations explored so far

- Questions
  - What was the parallel I/O problem?
  - How did the authors find a parallel I/O problem?
  - What was the solution?
  - How was the solution applied to fix the problem?
- Background: Various papers available in literature
- Deliverable: A short paper surveying I/O problems, solutions applied, and exposing research gaps (an advanced version of this is a cookbook for I/O performance)

Class projects

5. Performance comparison of sub-filing in HDF5 and PnetCDF

- Background: Sub-filing is an approach to split a very large file into smaller files. However, there are pros / cons with the approach on how the data is organized.
- Question
  - Which of the HDF5 and PnetCDF sub-filing approaches are best?
  - What better strategies for sub-filing are there?
- Deliverable: A short paper describing

- Before the next class, look at the project topics
  - Discuss with me in the next class
  - Think about why are you interested in any of the project topics

Summary of today's class

- Common data formats in science
  - **<u>Homework</u>** – Present a few data structures in the next class
- Brief intro to data storage hierarchy
  - Hardware
  - Software
- Class projects
  - **<u>Homework</u>** – Look at the project options and discuss in the next class

Send me an email if you have any questions regarding the homework or project topics

Next class

- Data format – Student presentations (2 min each)
- Class projects – questions
- I/O libraries
  - HDF5