IO-SEA

Project Summary

Contact: info@io-sea.eu
Project Duration: April 2021 – March 2024

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955811. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, the Czech Republic, Germany, Ireland, Sweden, and the United Kingdom.

Project Partners

(In Alphabatical order below)

- Atos-Bull (France)
- CEA (France) – **Project Co-ordinator**
- CEITEC (Czech Republic)
- ECMWF (International)
- Forschungszentrum Jülich (Germany)
- ICHEC (Ireland)
- IT4I (Czech Republic)
- JGU Mainz (Germany)
- KTH (Sweden)
- ParTec (Germany)
- Seagate (UK)

Goals

- Input/Output, Data Management aspects to tackle increased pressure on storage systems at Exascale.

- Design and development of a novel management and storage platform
  - Usage of Object Stores
  - Hierarchical storage management (HSM)
  - On-demand/Ephemeral provisioning of storage services & Scheduling
  - IO Instrumentation & AI Based telemetry analytics

- Co-design with next generation I/O intensive HPC oriented applications
  - Includes development of new flexible application Interface (“DASI”)

IO-SEA

IO-SEA is part of the "SEA" Projects

All addressing Modular Supercomputing Architectures

- Funded by the EuroHPC 2019-1 call focused on Software and Applications
  - The EuroHPC Joint Undertaking targets Exascale computers in Europe in 2023-24
- Coordinated with other on-going European projects, particularly the European Processor Initiative

DEEP-SEA: DEEP Software for Exascale Architectures

- Better manage and program compute and memory heterogeneity
- Targets easier programming for Modular Supercomputers
- Continuation of the DEEP projects series

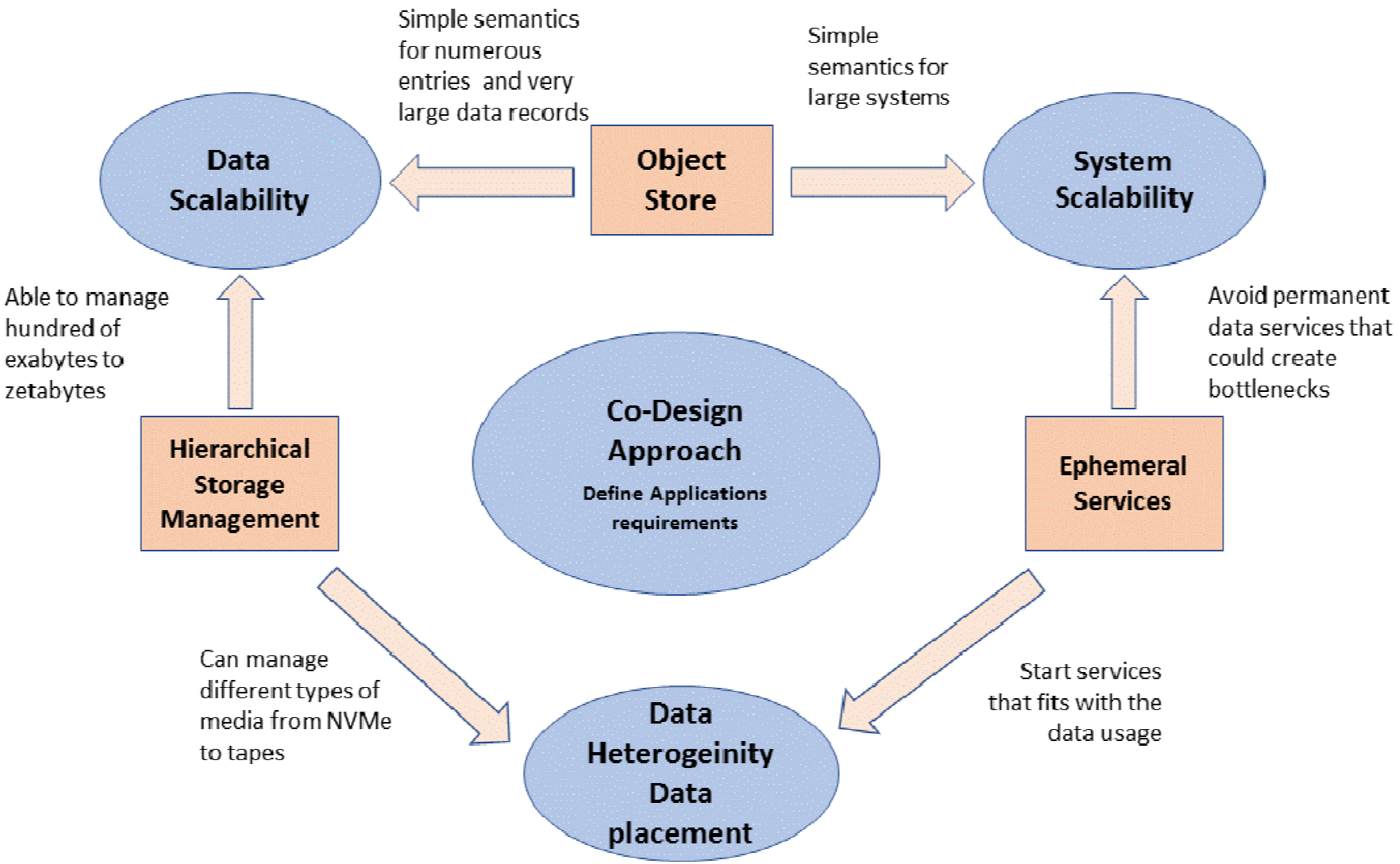
IO-SEA: Input/Output Software for Exascale Architectures

- Improve I/O and data management in large scale systems
- Builds upon results of SAGE1-2 projects and MAESTRO

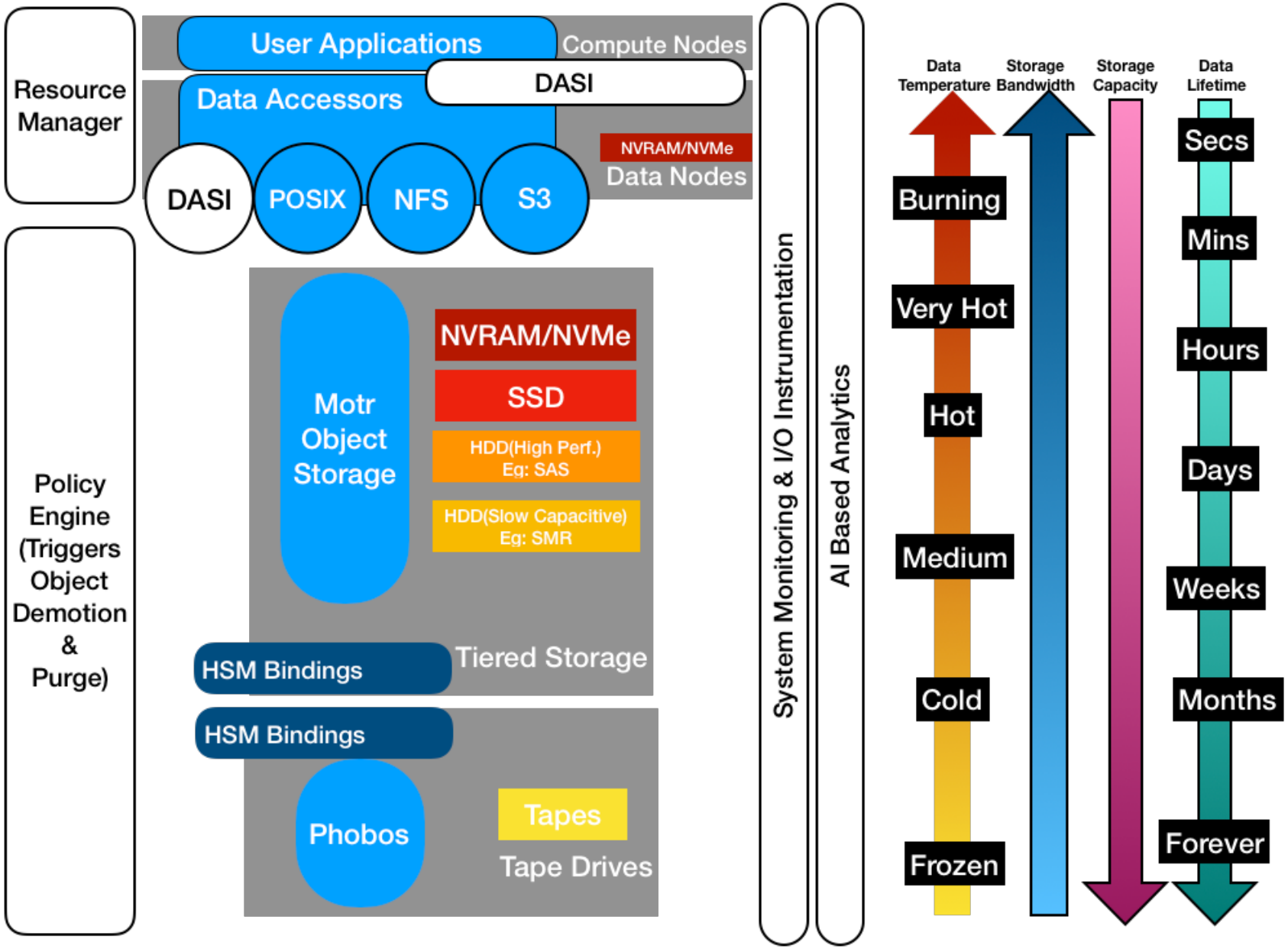
RED-SEA: Network Solution for Exascale Architectures

- Develop European network solution
- Focus on BXI (Bull eXascale Interconnect)

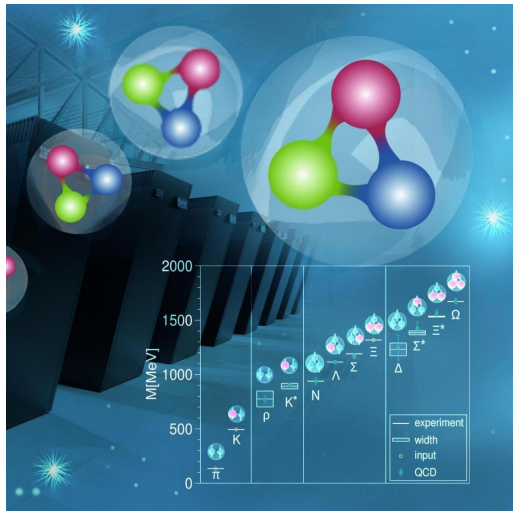
The IO-SEA Philosophy



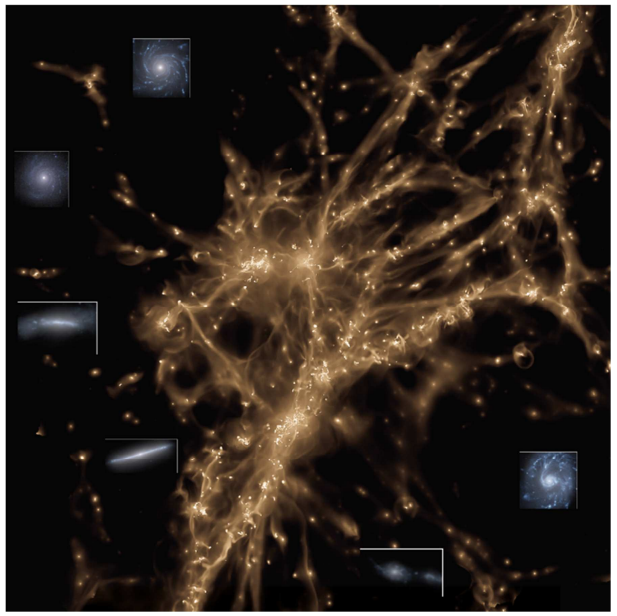
The Big Picture – IO-SEA



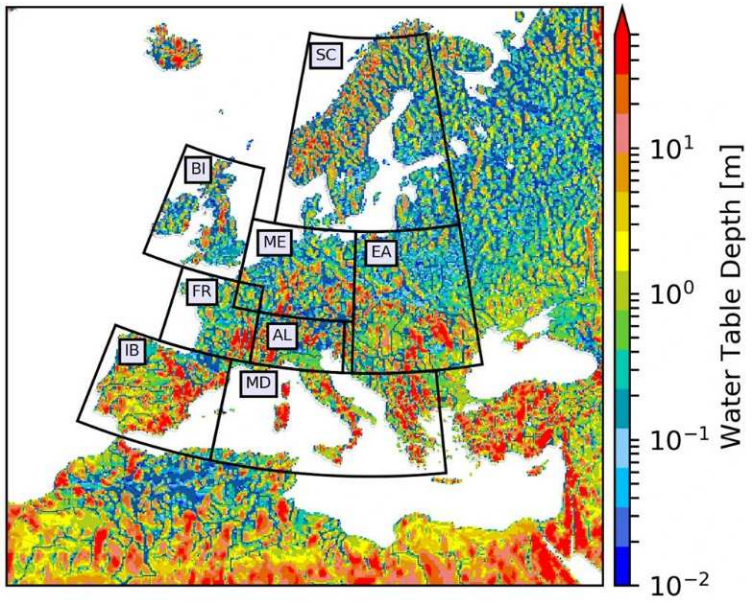
The Applications (& Co-Design)



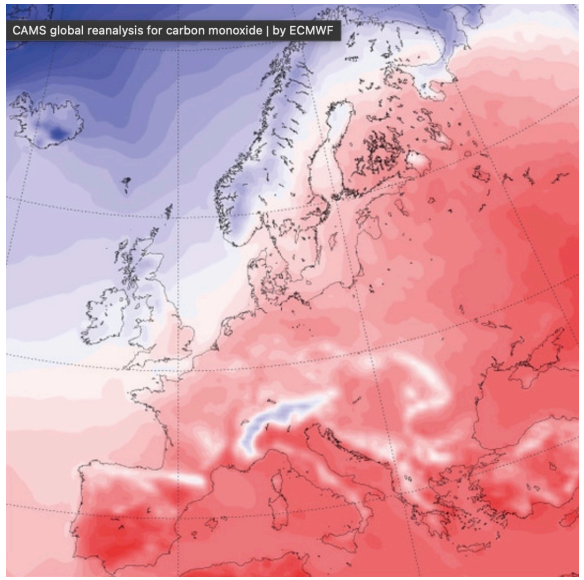
Lattice QCD (Particle Physics)



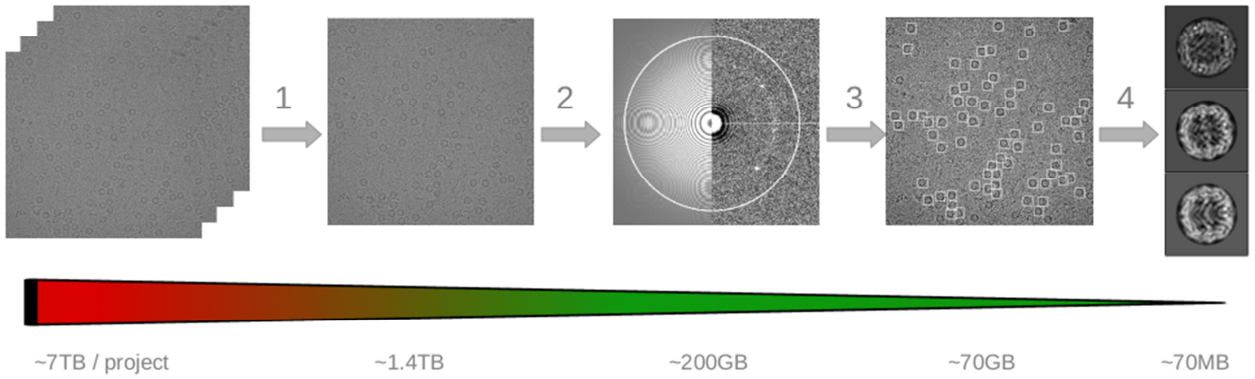
RAMSES (Astrophysical Systems)



Terrestrial Systems Modeling (TSMP)



Weather Forecasting Workflows

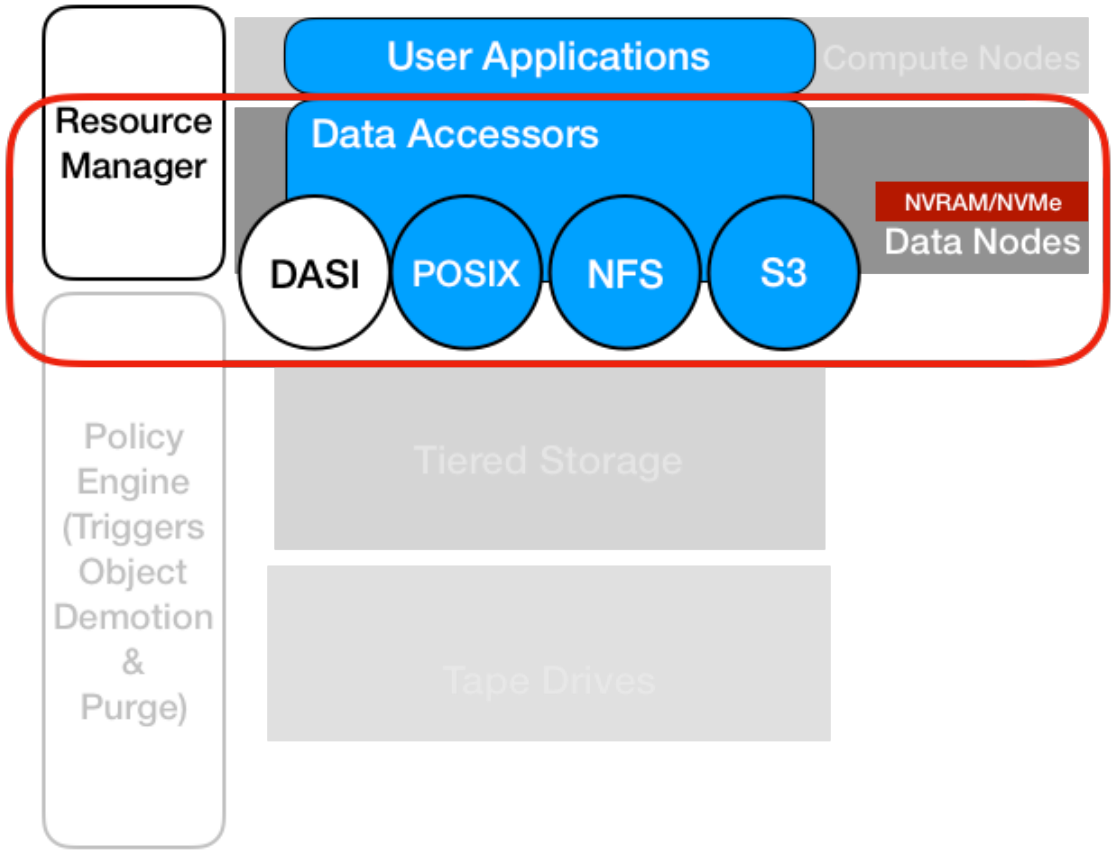


Electron Microscopy Imaging

IO-SEA

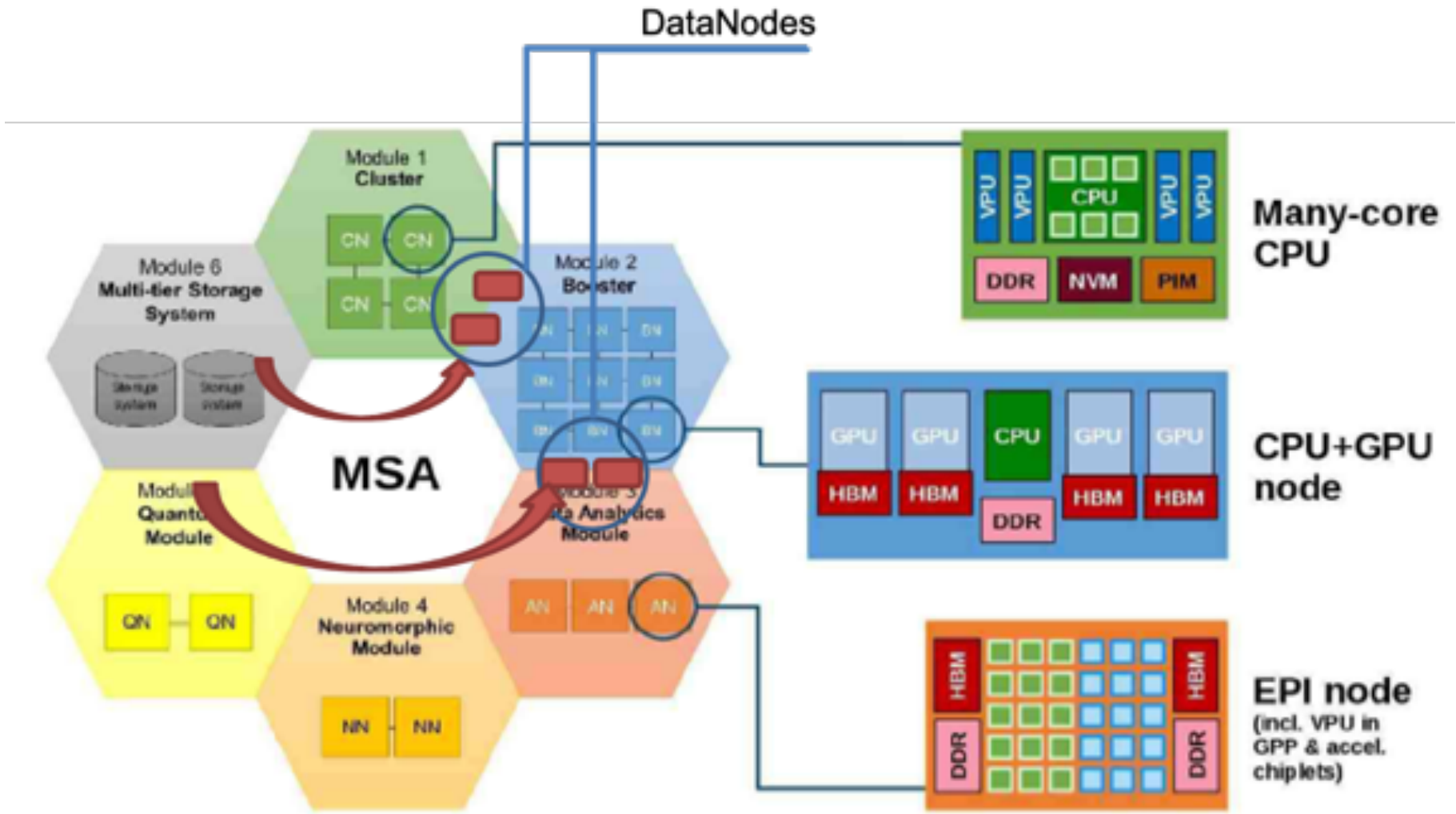
Ephemeral Data Access Environment

- Specialised data access environment suitable for applications and workflows
- Concept of Data Nodes
  - Specialised nodes that are used for I/O sits between Compute Nodes and Persistent Storage
- Ephemeral Services on Compute nodes
  - Ephemeral services are run on demand and in run time as needed by applications
    - POSIX
    - S3
    - NFS
    - etc
- Will Leverage NVRAM/NVMe resources available on data nodes
- Will develop mechanisms to schedule data accesses on demand through the Ephemeral Services

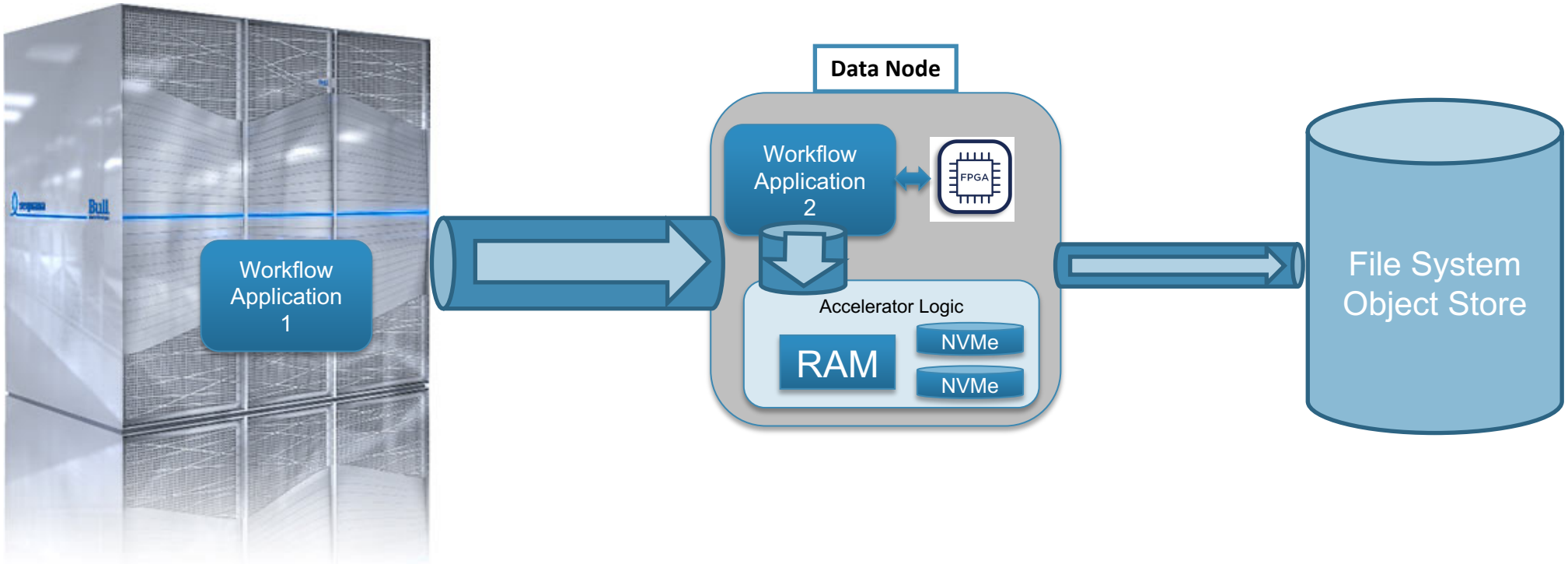


Ephemeral Data Access Environment

Envisioned Location of Data nodes in the MSA



Ephemeral Data Access Environment Example

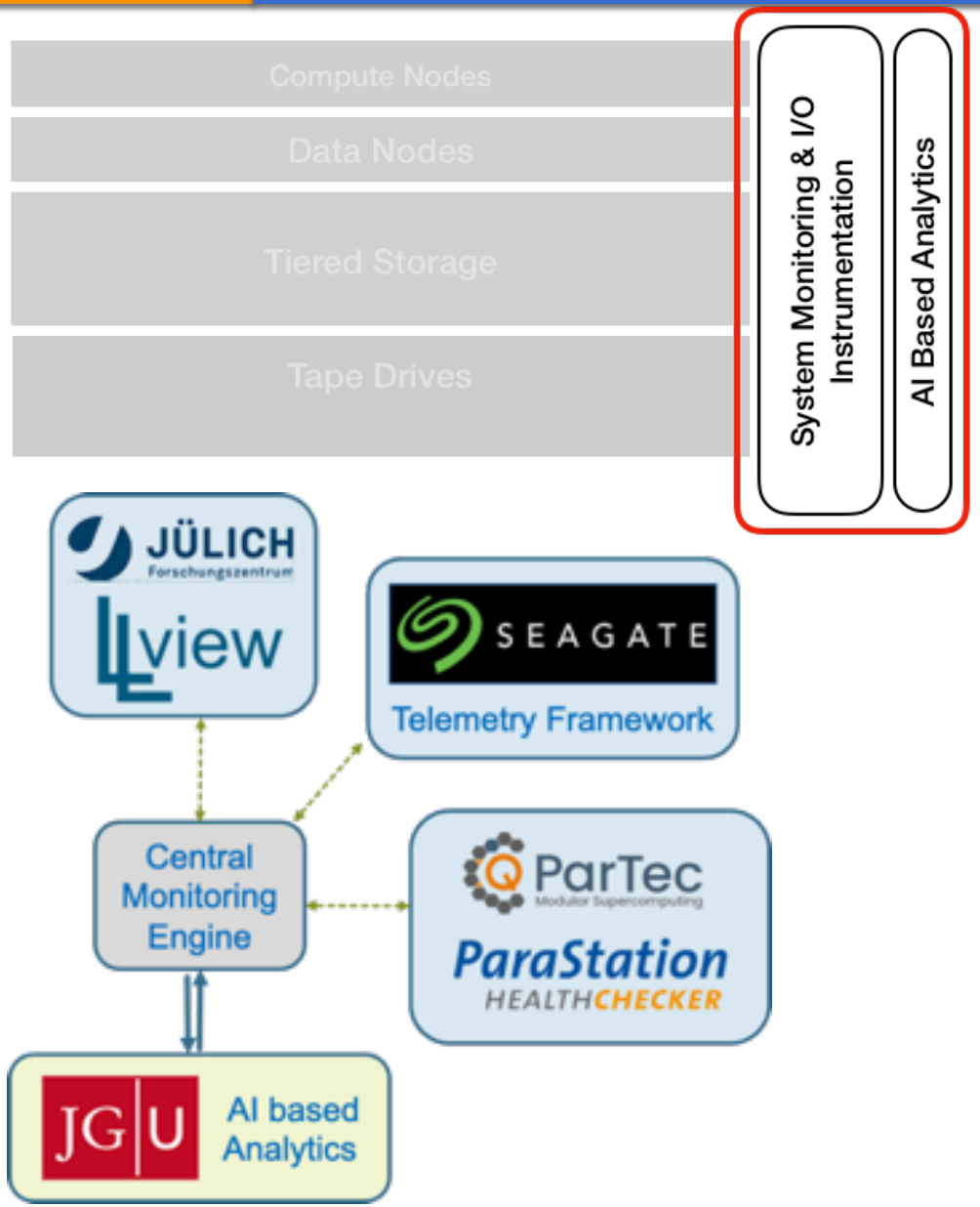


- Pre-processing: Prefetching in fast storage, decompressing, formatting, …
- Post-processing: Formatting results, generating images, compressing, …
- On the Fly processing: Processing intermediate data.

IO-SEA

Instrumentation & Monitoring

- Gathering knowledge on I/O behaviour of applications & workflows
  - Analyse collected data using AI based techniques
- Knowledge will feed algorithms that will allocate I/O services & data nodes resources
- Gathering knowledge about infrastructure resources to make efficient scheduling decisions
  - AI algorithms will complement scheduling decisions made by users
- I/O & instrumentation tools will be adapted to each protocol (S3, NFS, POSIX, etc)
- Live view of the infrastructure will be provided through the tools techniques and methods within IO-SEA

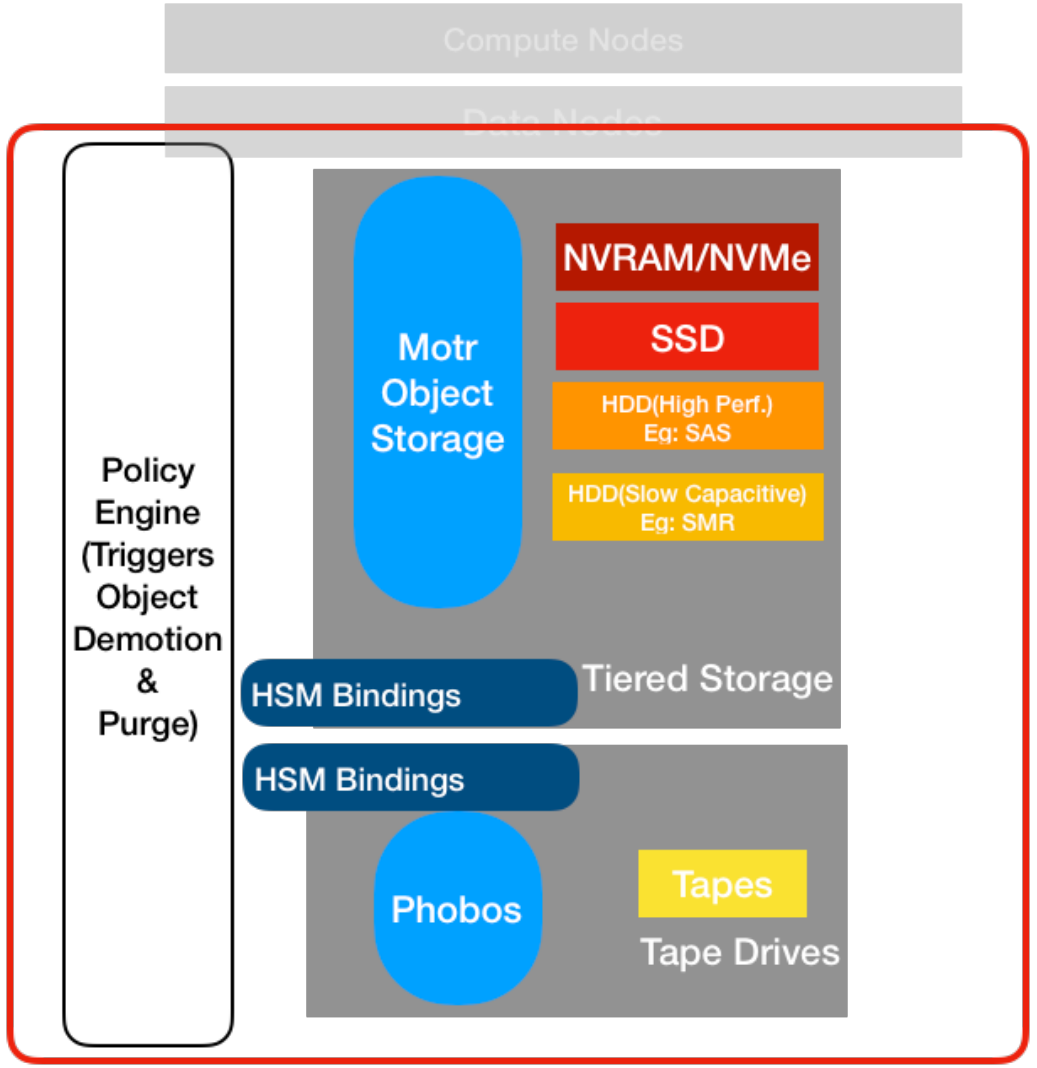


Components of Infra. Monitoring

IO-SEA

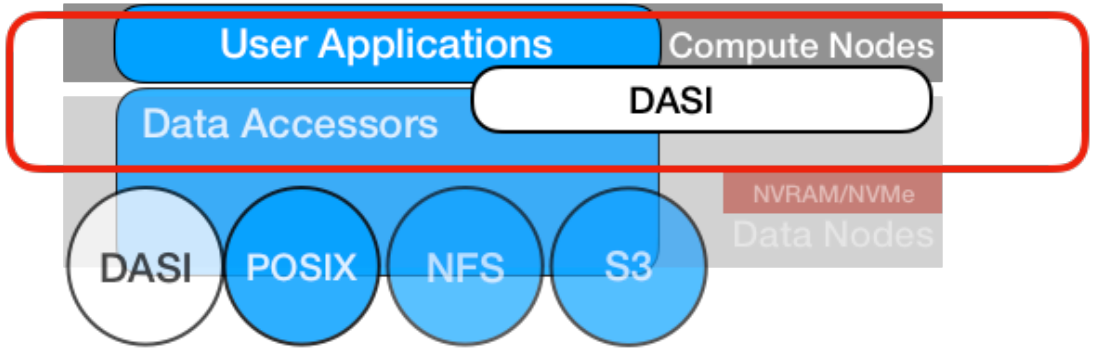
Hierarchical Storage Management Features

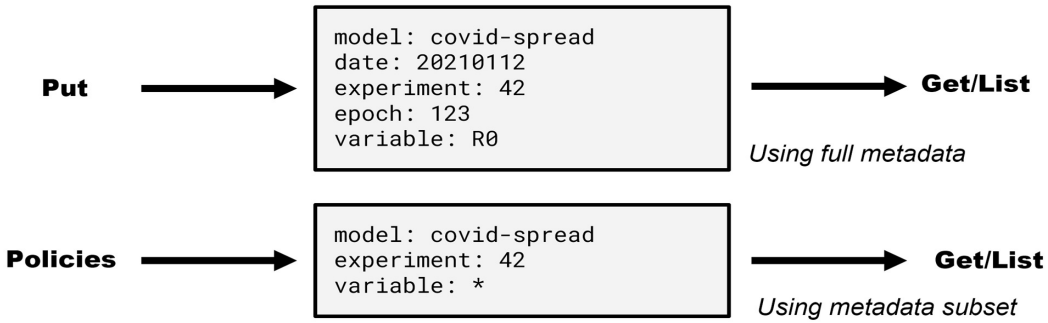
- HSM Mechanism for managing data movements b/w multiple tiers of Persistent Storage tiers
  - NVMe/NVRAM
  - SSD
  - Disk
  - Tape
- Copy Tool mechanism of coupling of two object stores
  - Motr (NVRAM, SSD, Disk)
  - Phobos (Tape)
- Capability to move data between IO-SEA and POSIX namespaces
  - POSIX Support within Copy Tool



Application Interfaces

- Data Access and Storage Interface (DASI) Layer (Language) will be built to abstract the complex storage layer
- Will use semantic description of data – speaking the language of the scientific domain
  - Eg: COVID model data addressed by scientifically meaningful keys – rather than files & objects
- Data can be queried using subsets of these keys
- Possibility to set access policies by data lifetime, data hotness, etc
- DASI language could be used to describe data management for workflow scheduling and in-situ processing
  - May also be exposed via a POSIX interface





Project Innovation - Summary

Ushering the next generation in Exascale HPC with better data management in MSAs

- New Methods of Deploying Storage Services for Exascale HPC applications
  - **Ephemeral Services**
- New methods to access extreme data using semantic descriptions – Codesigned with applications
  - **DASI**
- Novel Methods to access and stage data across deep storage hierarchies, NVMe/NVRAM through Tape
  - **Advanced HSM and Policy Engines**
- New methods to understand I/O and intelligent use of System Telemetry
  - **AI based Analytics**
- Continued Utilization of Object stores in the realm of Exascale HPC
  - **Motr**
  - **Phobos**

IO-SEA