Advancing Object Storage for Extreme Scale

July 2021

Sai Narasimhamurthy
Eng. Director, EU R&D,
Seagate Systems
Project Co-ordinator, Sage2
www.sagestorage.eu

Object Stores in the realm of Extreme scale Computing

★ Object Stores – Designed originally to store unstructured data

★ Object stores - stores data in a flat hierarchy with each "objects" uniquely accessible, as opposed to traversing a file system tree

★ Extremely relevant for cloud storage (Think Amazon S3 buckets) as a storage backend to many web applications, etc

★ Are Object stores relevant for HPC and Extreme Scale computing?

★ A series of requirements gathering workshops with HPC community (Circa ~2012 onwards)

  ☆ Reported in: http://www.pdsw.org/pdsw-discs16/wips/danilov-wip-pdsw-discs16.pdf

  ☆ Led to the architecture of Cortx/Motr

Object store: Provision of the functionalities for extreme scale
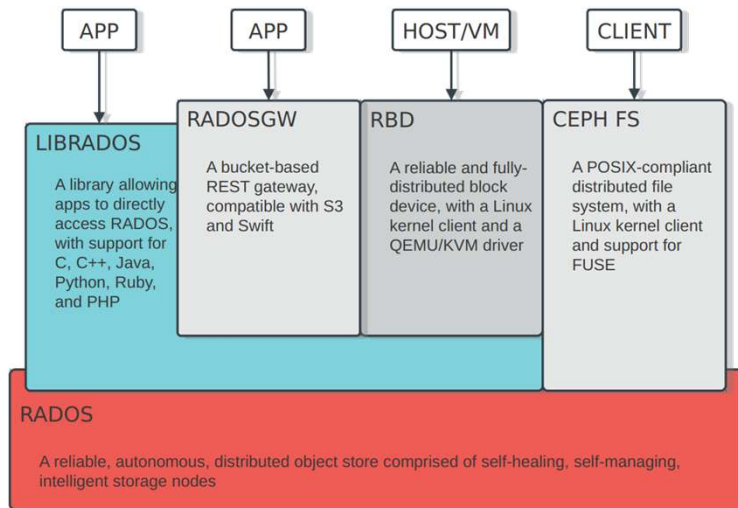

Motr is the open source version of formerly "Mero"
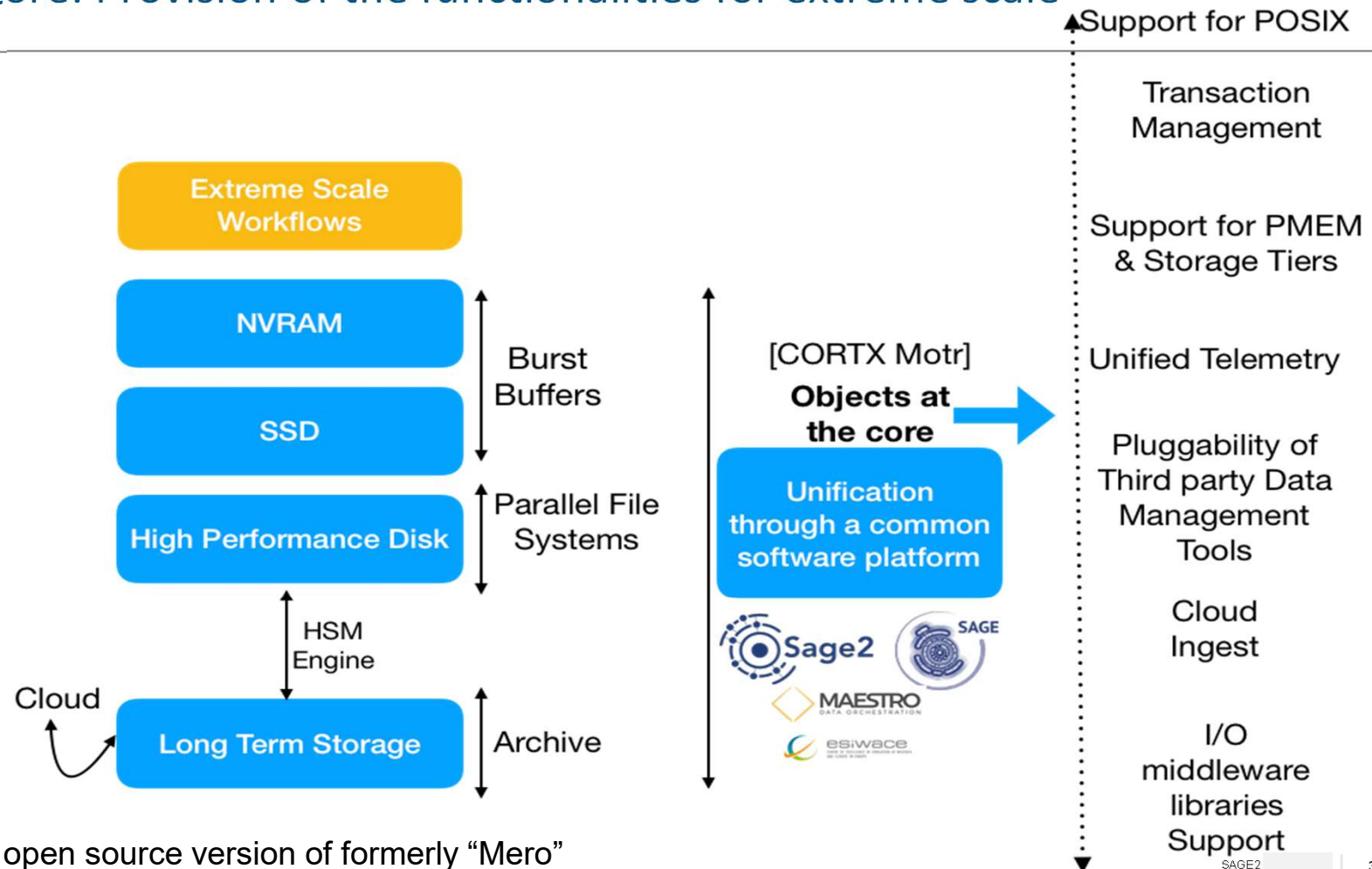
CORTX Motr Architecture
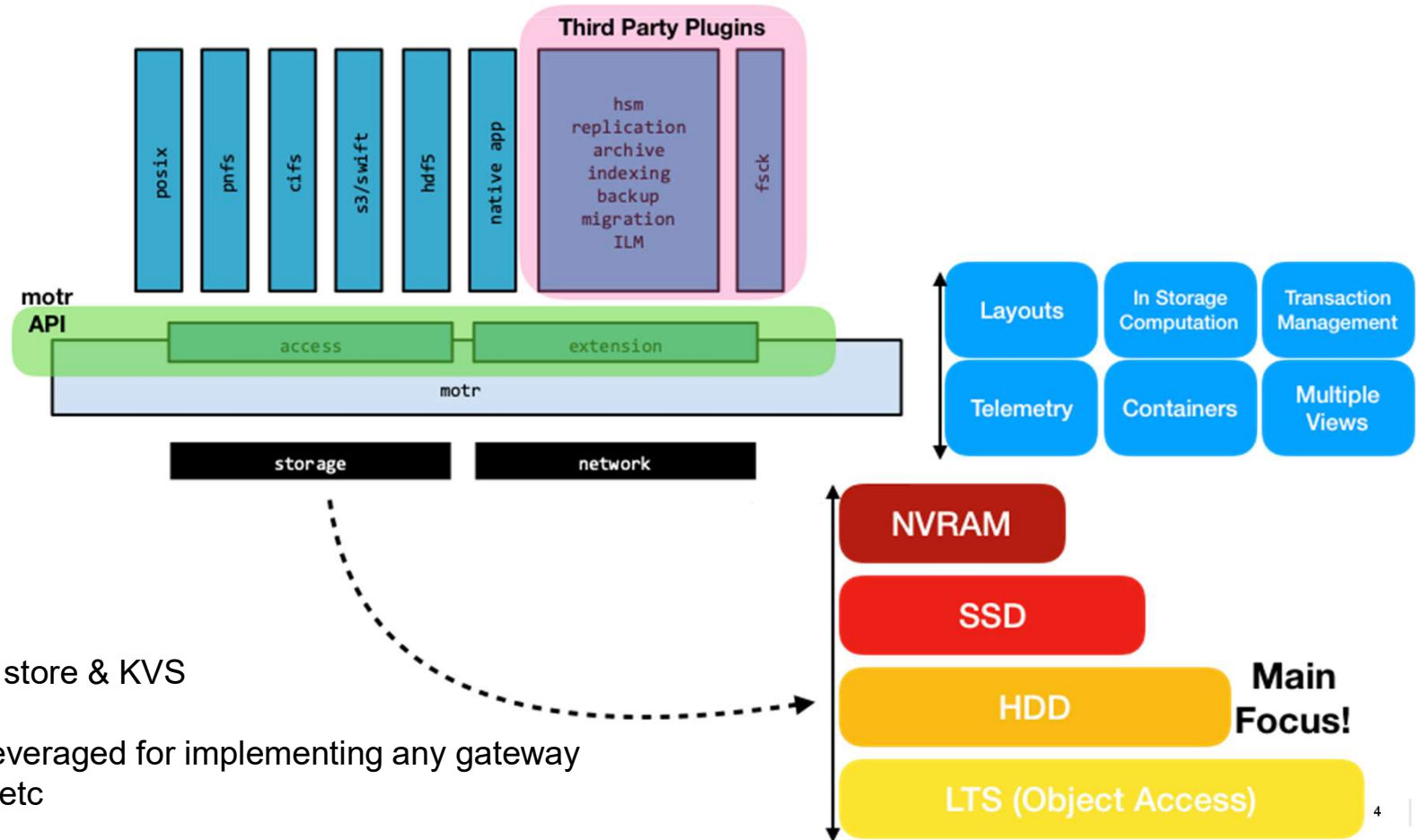
Motr – Object store & KVS

KVS can be leveraged for implementing any gateway
Eg: S3, NFS, etc

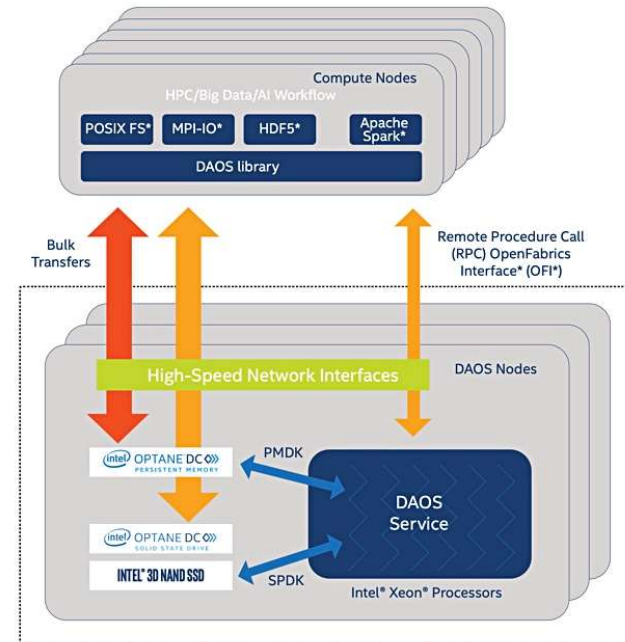# Other Object Stores approach (eg: Ceph and DAOS)

Ref: Redhat

Ref: Blocks and Files

DAOS Service = Motr Services = RADOS

LIBRADOS = DAOS Library = Motr API

Note: Interesting complementarities!
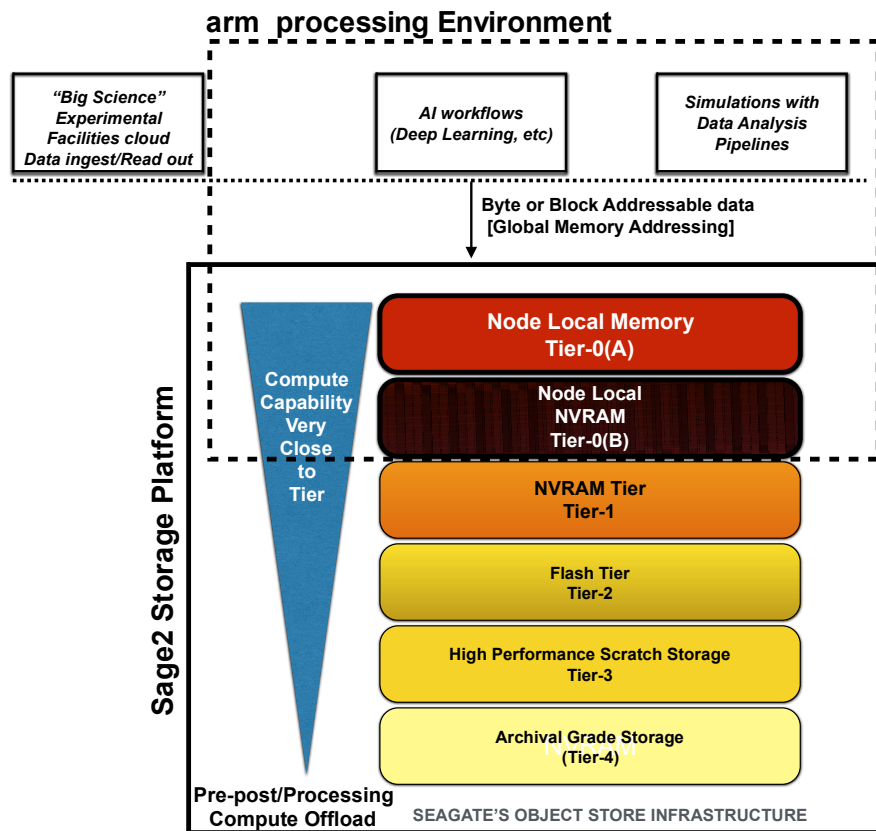**They all have a role to play**

Sage2 EU R&D project (2018 – 2021)

Cortx demonstrated in real world Exascale HPC/AI use cases

Sage2 Summary: Exploring the next generation of Object Storage Features

arm processing Environment

"Big Science" Experimental Facilities cloud Data ingest/Read out

AI workflows (Deep Learning, etc)

Simulations with Data Analysis Pipelines

Byte or Block Addressable data [Global Memory Addressing]

Sage2 Storage Platform

Compute Capability Very Close to Tier

Node Local Memory Tier-0(A)

Node Local NVRAM Tier-0(B)

NVRAM Tier Tier-1

Flash Tier Tier-2

High Performance Scratch Storage Tier-3

Archival Grade Storage (Tier-4)

Pre-post/Processing Compute Offload

SEAGATE'S OBJECT STORE INFRASTRUCTURE

Vision:

Extending storage systems into Compute nodes & blurring the lines between memory & storage

Four primary Innovations

1. **Compute node local Memories** part of storage stack
2. **Byte Addressable extensions** into Persistent storage (Global Memory Abstraction)
3. **Co-design** with new workflows: Mainly Data analytics pipelines w/ **AI/Deep learning**
4. **Co-design** with **ARM based environments** – moving towards European HPC Ecosystem Goals.

AI/DL use cases expected to be memory intensive & will exploit node local memory which will need to be extended

Sage2 - Key Stack Components
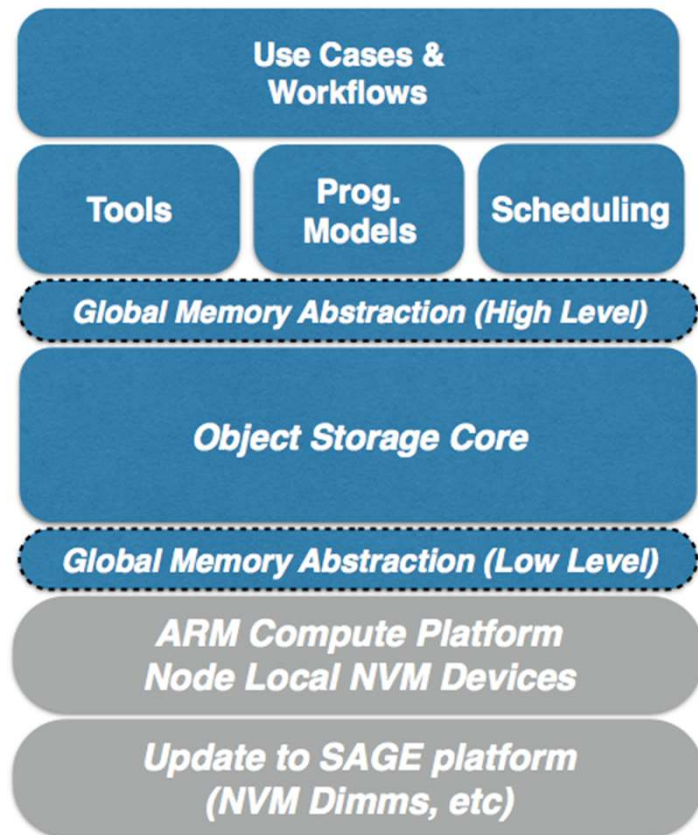
Tools/ Prog. Models/Schedulers

- dCache, High Speed Object Transfer, I/O Containers, TensorFlow, Slurm for Mero, Object access Prog. Mod, Simple Access Interface

GMA

- High Level – API for mapping Objects in Memory
- Low Level – Incorporating NVDIMMs

Object Storage Core

- Motr for GMA
- Motr extreme scale comps. - QoS, DTM, Function Shipping
- Motr for Sage2 (Incl. ARM port)

ARM

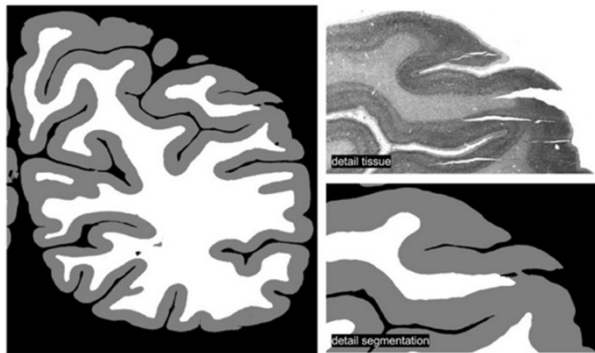- ARM support for NVDIMMs

Sage2 Use Cases

AI Based Data Analysis
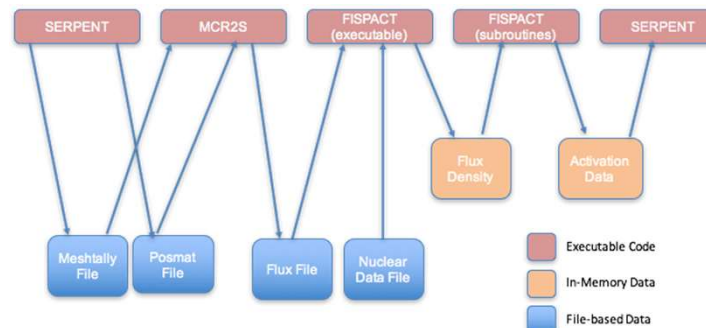[1]Cervical Cancer Diagnosis

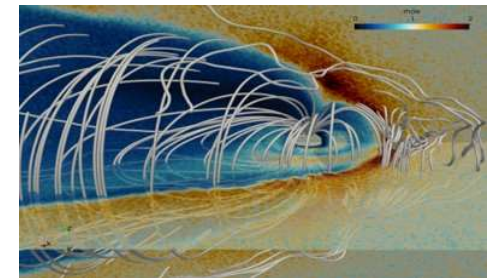**AI Based Data Analysis**
[2] Multi-label Classification of Large Videos

[4] Radio Astronomy Data Analysis

[3] Brain Image Data Analysis

Machine Learning
[6]Tensorflow for machine learning monitoring data

SERPENT   MCR2S   FISPACT (executable)   FISPACT (subroutines)   SERPENT

Flux Density   Activation Data

Meshtally File   Posmat File   Flux File   Nuclear Data File

Executable Code
In-Memory Data
File-based Data

[5] Multi-Physics Multi-stage workflows (Nuclear Fusion)

[7] Classic HPC Applications

SAGE Prototype at Juelich

Rack2

Tier 2-a
Tier 3-a
ARM Nodes
Tier 3-1
Tier 3-2
Tier 4-1
Tier 4-2

Rack1

Network
Clients
ThunderX2
Tier 1-1
Tier 1-2
CMU Login
Tier 2-1
Tier 2-2
Tier 2-3
Tier 2-4
OceanStor

Rack3

Datawarp
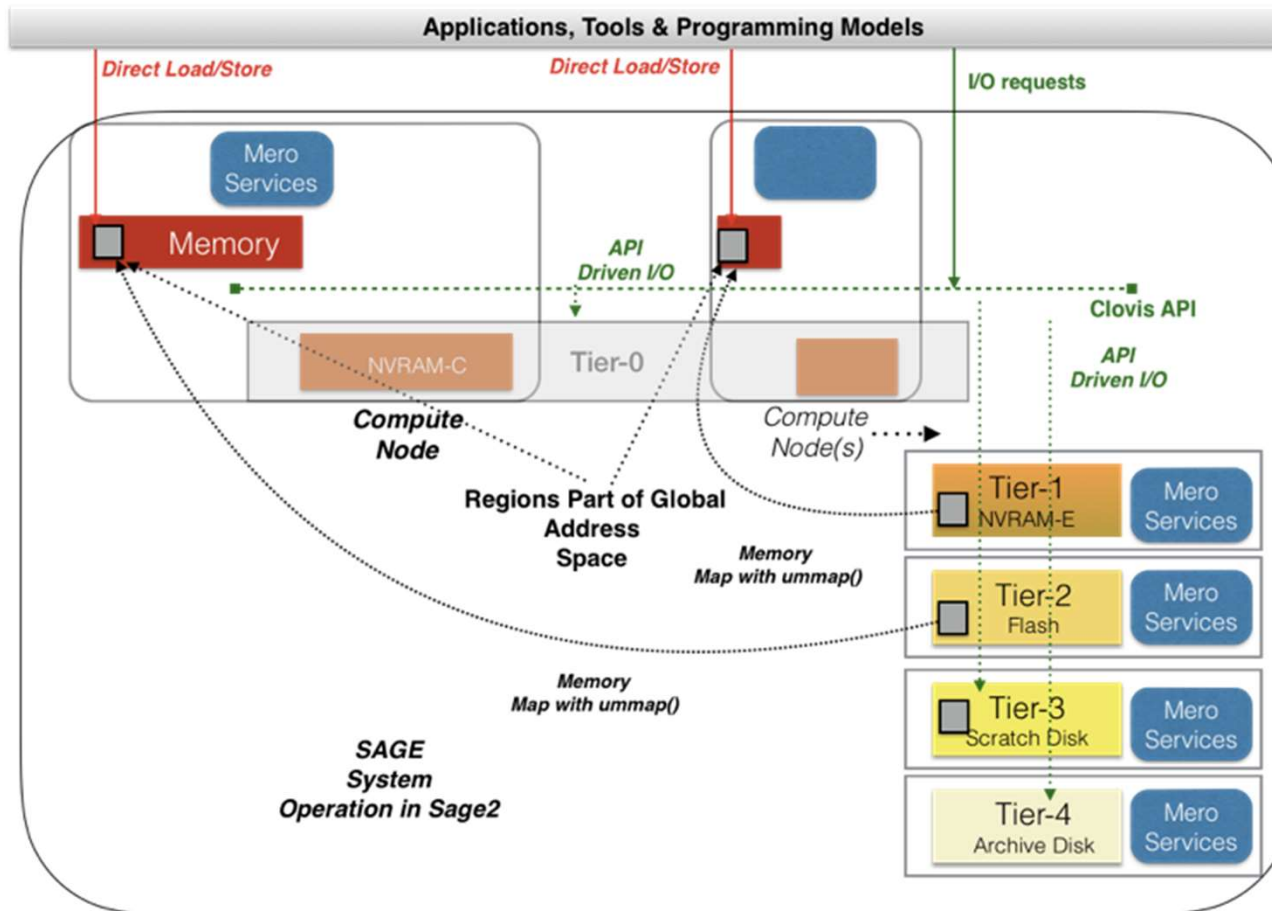Vis Nodes
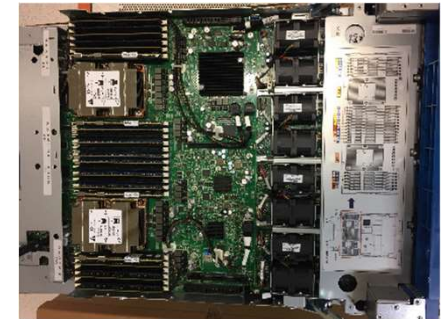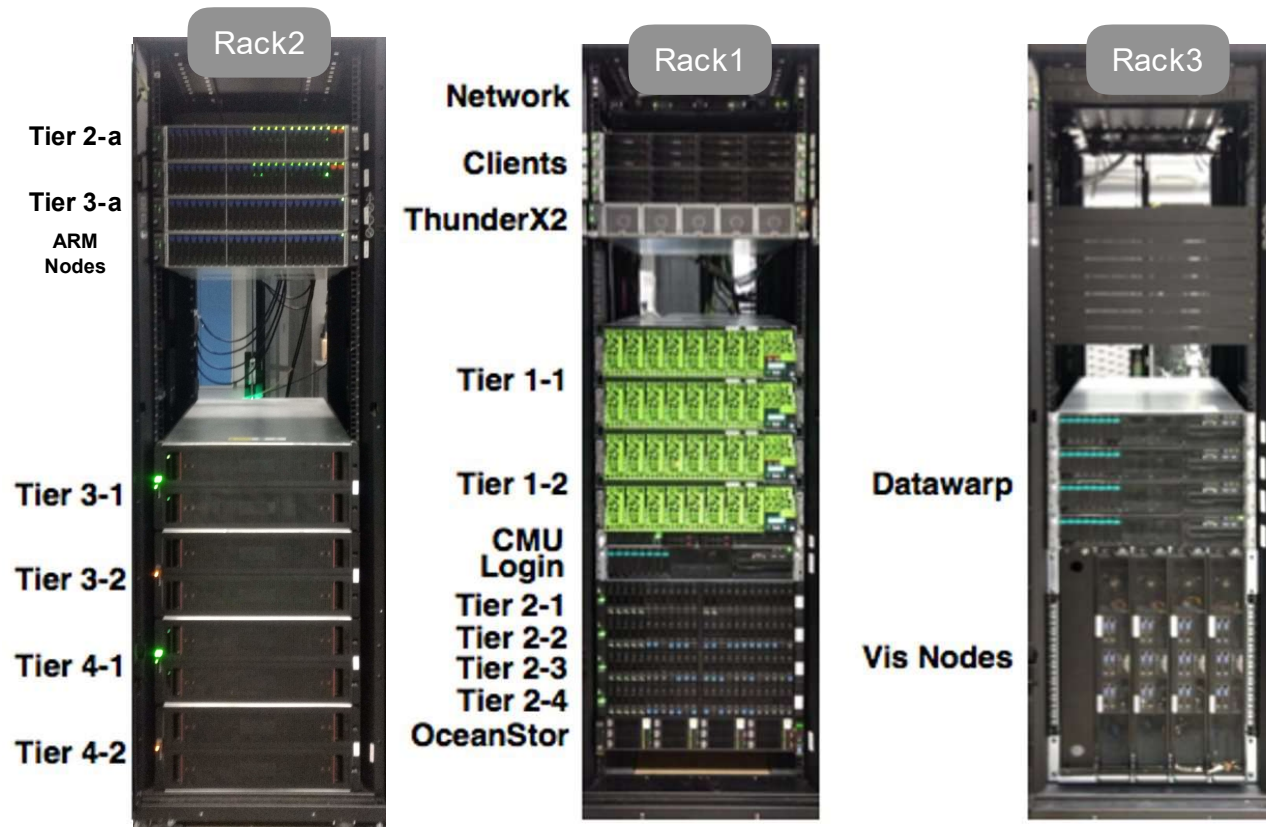
Note:
2 NVDIMM
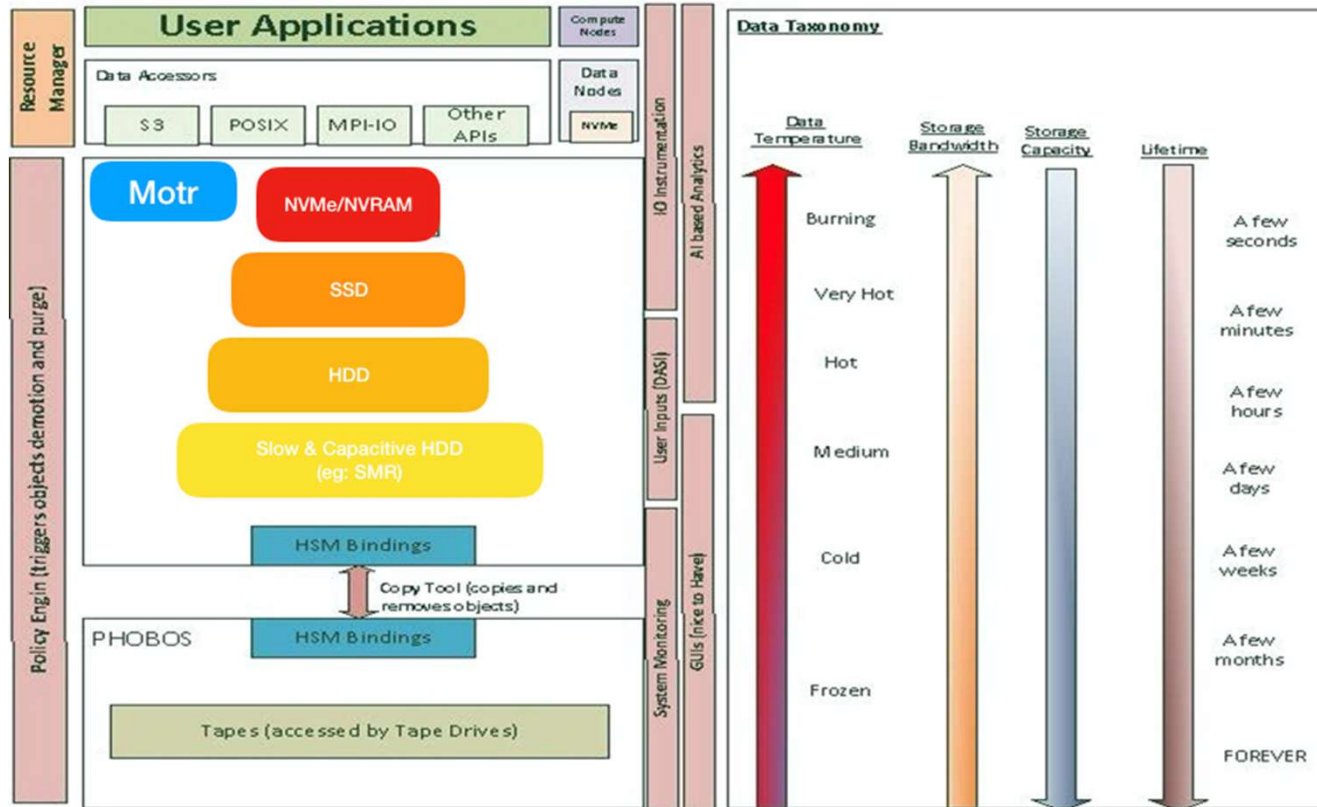Equipped
Nodes also
implemented
at ATOS

Sage2 System Operation with Motr

Pls note NVRAM-C & NVRAM-E are Sage2 specific terms

Going forward – Motr in the IO-SEA Project

Sage2 Status – July 2021

- **Application Porting on Motr on the SAGE platform in progress**

- **Implementation of the Motr components & GMA & ARM porting**

- **Ecosystem tools – Implementation & Porting**

- **Performance Analysis**

Interested to test drive Motr/CORTX?

https://github.com/Seagate/cortx

Interested to test drive SAGE?

info@sagestorage.eu

Hackathons actively in progress, where we invite external community

Thank You

www.sagestorage.eu
Twitter: @sagestorage
sai.narasimhamurthy@seagate.com

Sage2