sage [1] (sāj)
n.
One venerated for experience, judgment, and wisdom.
adj. **sag·er, sag·est**
1. Having or exhibiting wisdom and calm judgment.
2. Proceeding from or marked by wisdom and calm judgment: *sage advice.*

Percipient StorAGe for Exascale Data Centric Computing

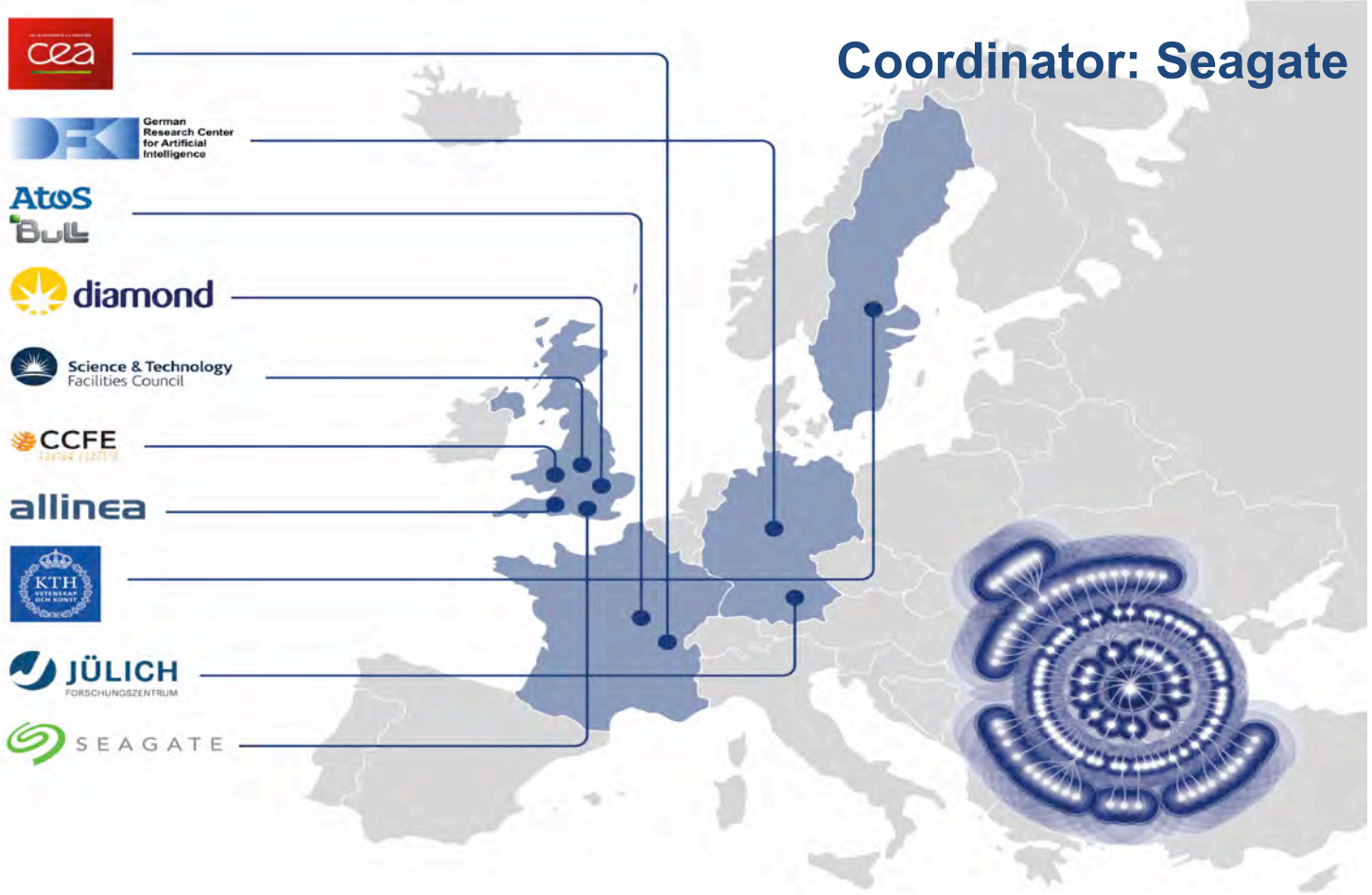
Project Summary and Status

Sai Narasimhamurthy
Seagate Systems UK

ExaIO Workshop SC'16 - Nov 2016

Per-cip-i-ent (pr-sp-nt)
Adj.
Having the power of perceiving, especially perceiving keenly and readily.
n.
One that perceives.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 671500



Coordinator: Seagate

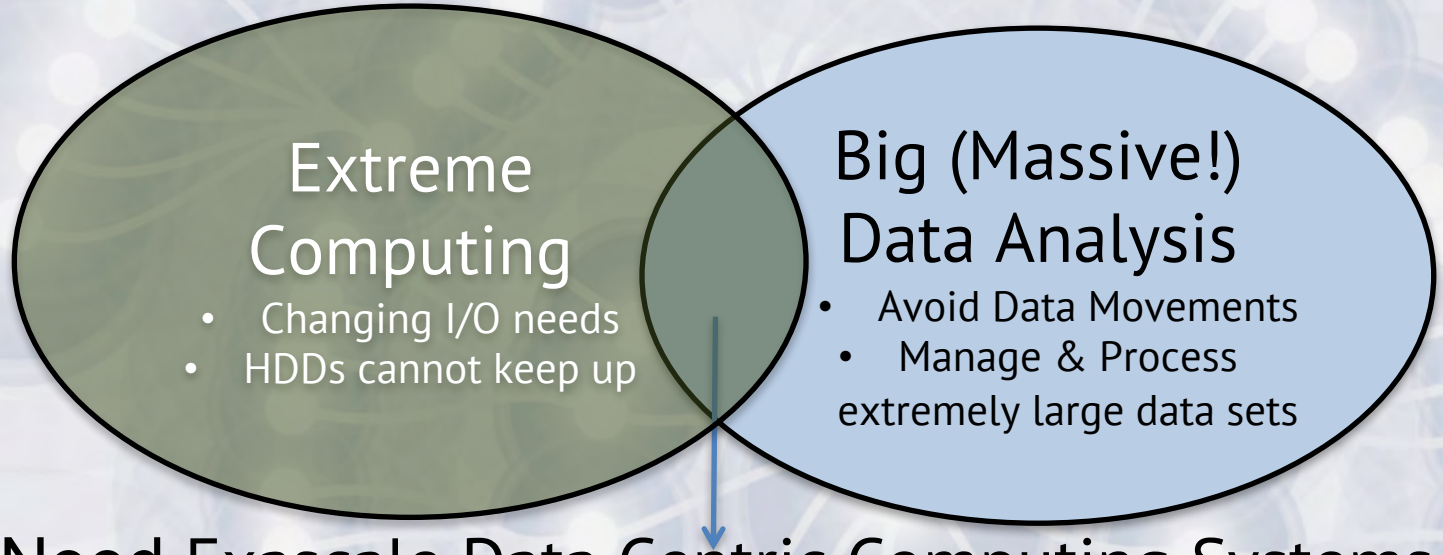
SAGE

Consortium

- *Storage cannot keep up w/ Compute!*
- *Way too much data*
- *Way too much energy to move data*
- *New Storage devices use unclear*
- *Opportunity: Big Data Analytics and Extreme Computing Overlaps*

SAGE

Storage Problems at Extreme Scale



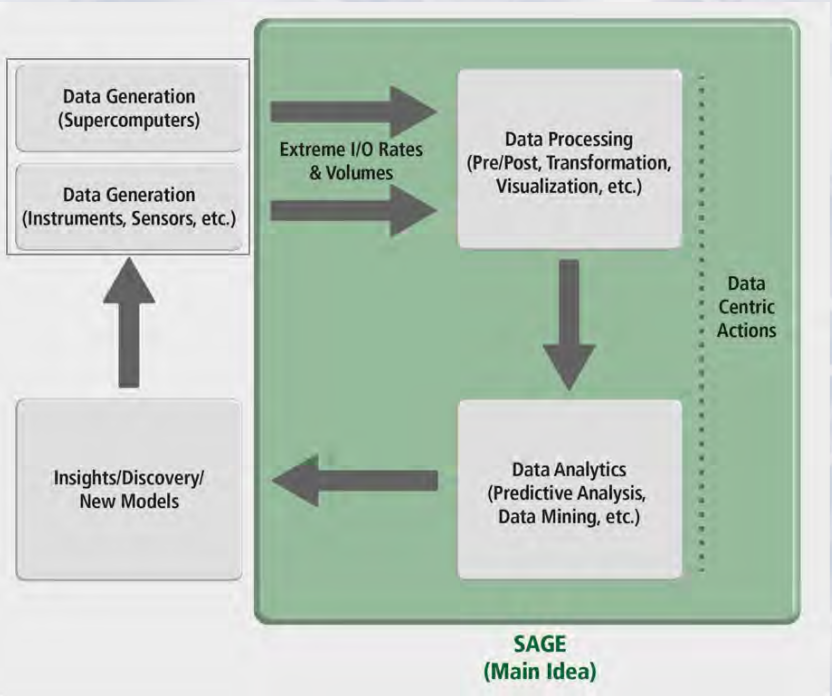
Need Exascale Data Centric Computing Systems

Big **D**ata **E**xtreme **C**omputing (*BDEC Systems*)

SAGE Validates a BDEC System which can Ingest, Store, Process and Manage extreme amounts of data

SAGE

SAGE Project Goal



- Deal with high I/O rates
  - Instruments
  - Running Simulations
- Performing data Processing
  - In parallel with ongoing simulation and Data collection
- Insights could be fed back into
  - Running Simulations
  - Data collection decisions

- **Provide Novel Storage Architecture**
  - Object Storage w/ a very flexible API , driving
    - Multi-tiered Hierarchical Storage System
    - Providing integrated compute capability

- **Purpose: Increase overall scientific throughput!**
  - Co-designed with Use cases
  - Integrated with Ecosystem tools

- **Provide roadmap of component technologies for achieving Extreme Scale**

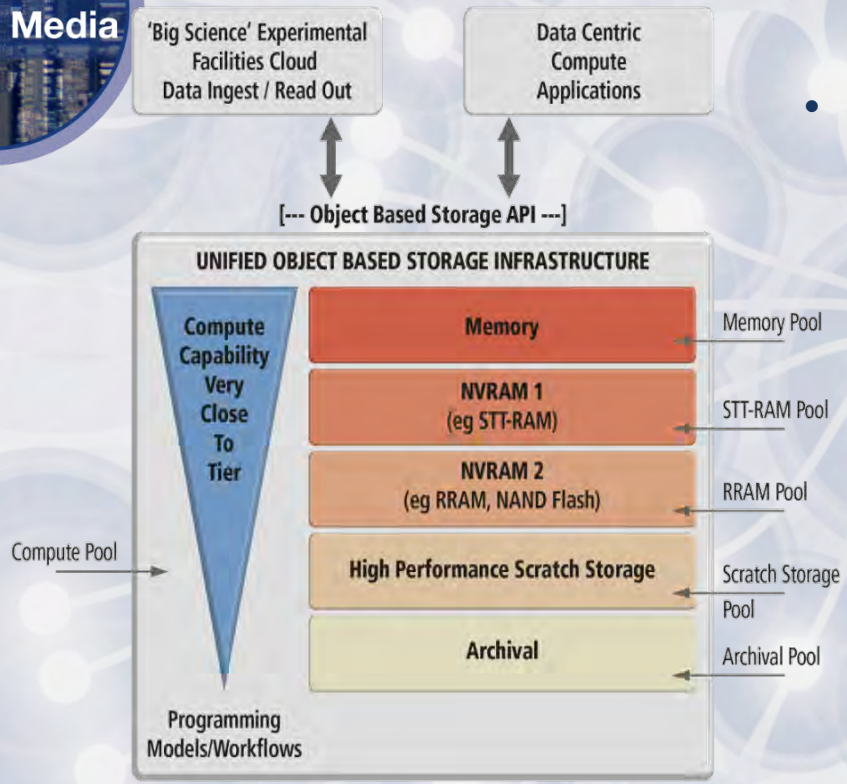
- **Key Targets for SAGE**
  - HPC & Big Data technology Influencers
  - Scientific Communities & Infrastructure
  - Wider Markets

SAGE

SAGE Overall Objectives

SAGE

Research Areas

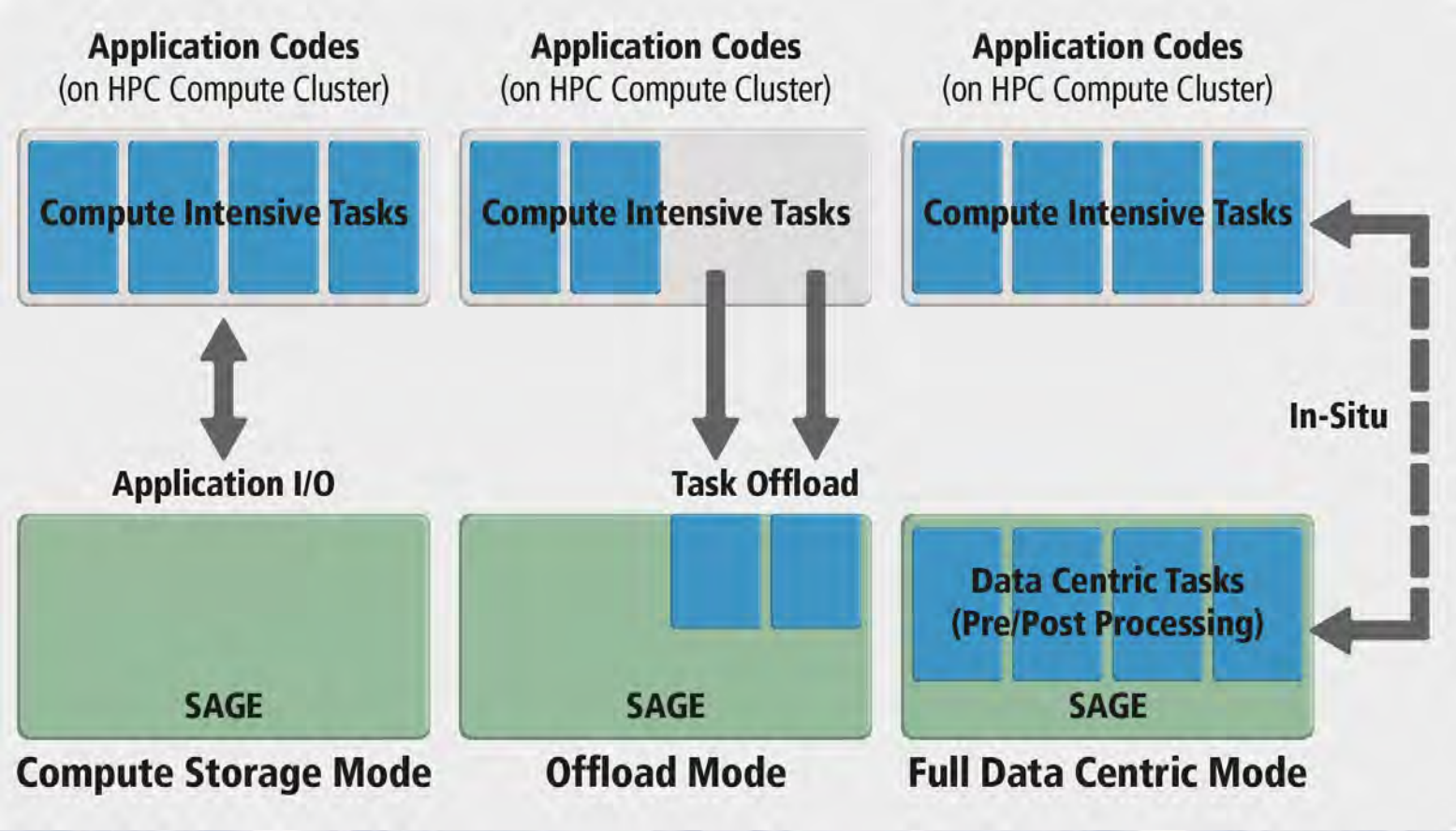


- **Goal**
  - Build the data centric computing platform
- **Methodology**
  - Advanced Object Storage
  - New NVRAM Technologies in I/O stack
  - Ability for I/O to Accept computation
    - Incl. Memory as part of storage tiers
  - API for massive data ingest and extreme I/O
  - Commodity Server & Computing Components in I/O stack

SAGE

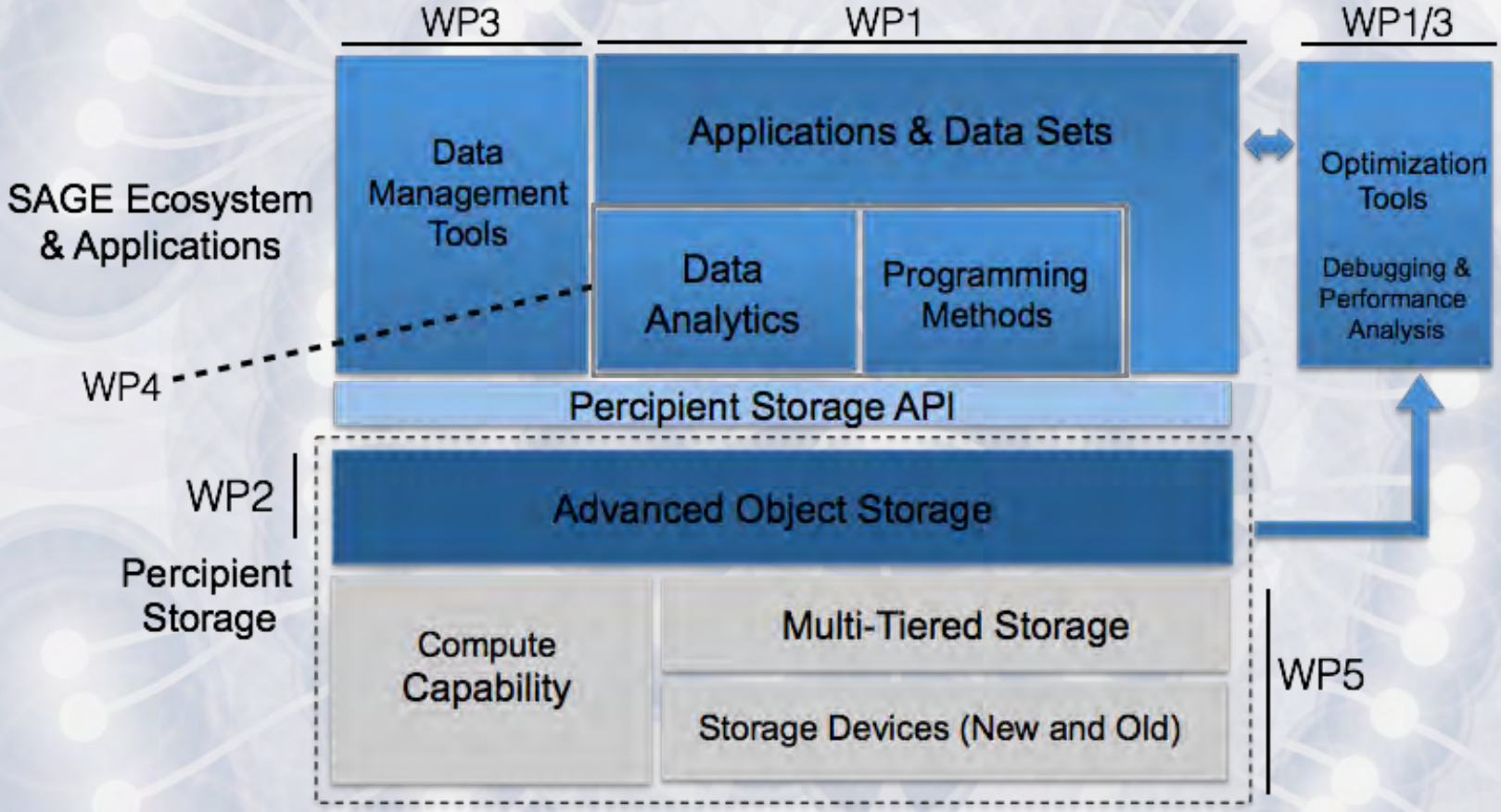
Percipient Storage Overview

Percipient Storage Methods



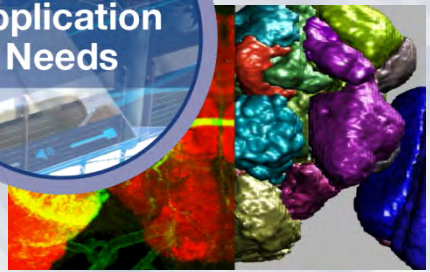
SAGE

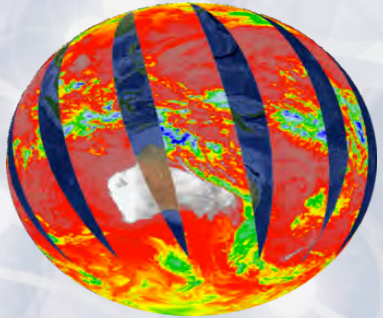
Percipient Storage Modes of Operation

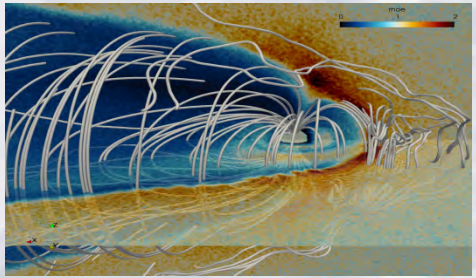


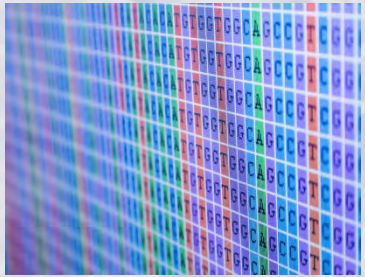
SAGE

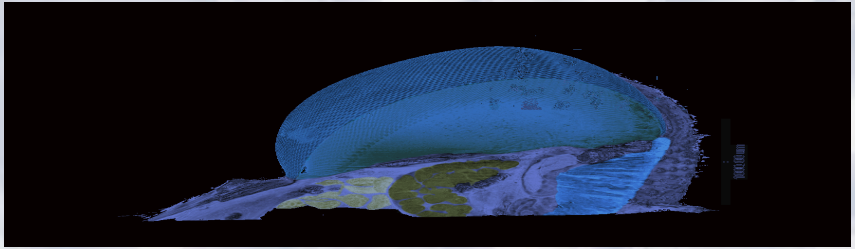
Percipient Storage Stack







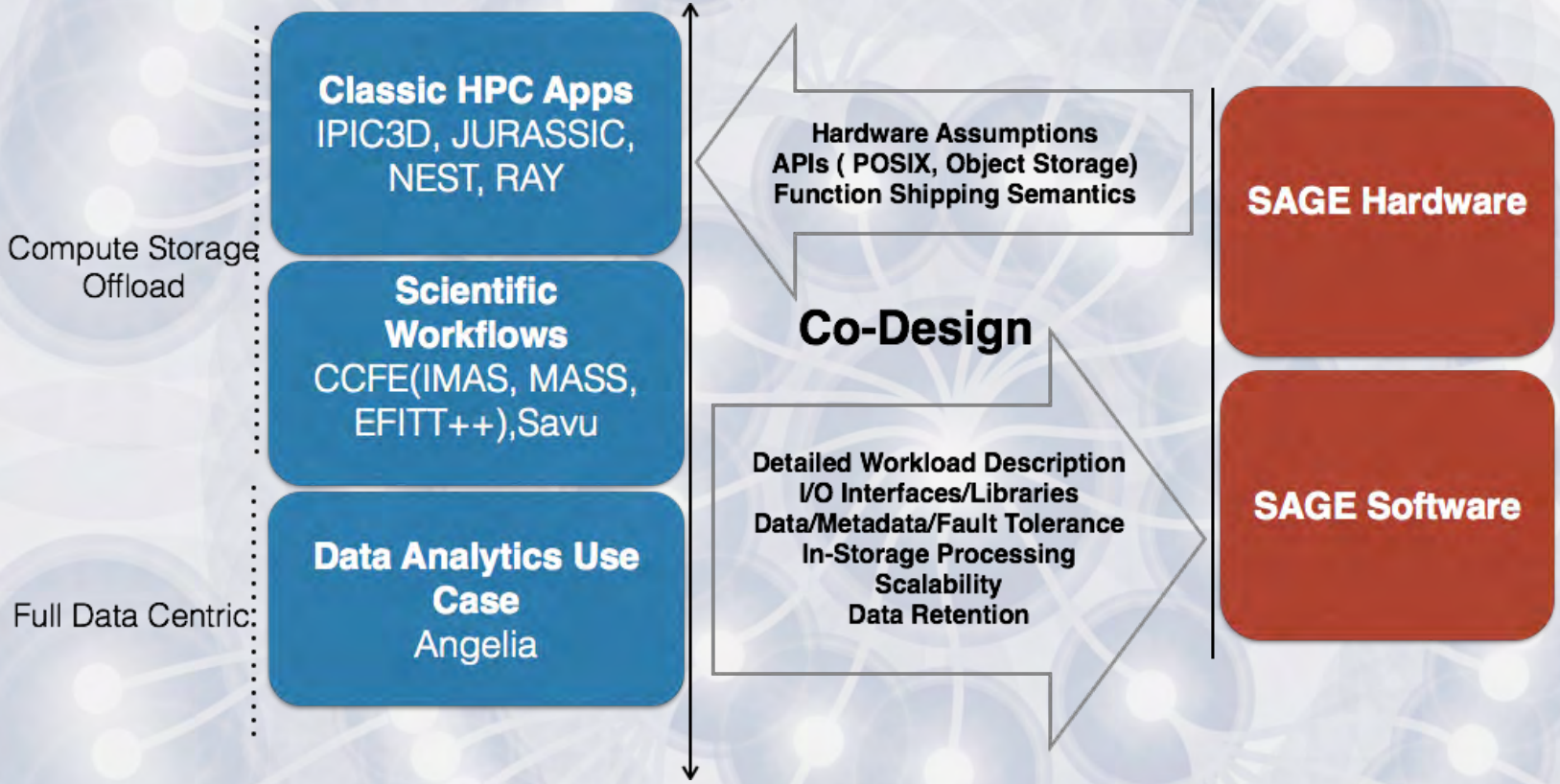




- **Primary Goal**
  - Demonstrate Use Cases & Co-Design the system
- **Methodology**
  - Obtain Requirements from:
    - Satellite Data Processing
    - Bio-Informatics
    - Space Weather
    - Nuclear Fusion (ITER)
    - Synchrotron Experiments
    - Data Analytics Application
  - Detailed profiling supported by Tools
  - Feedback requirements from platform

SAGE

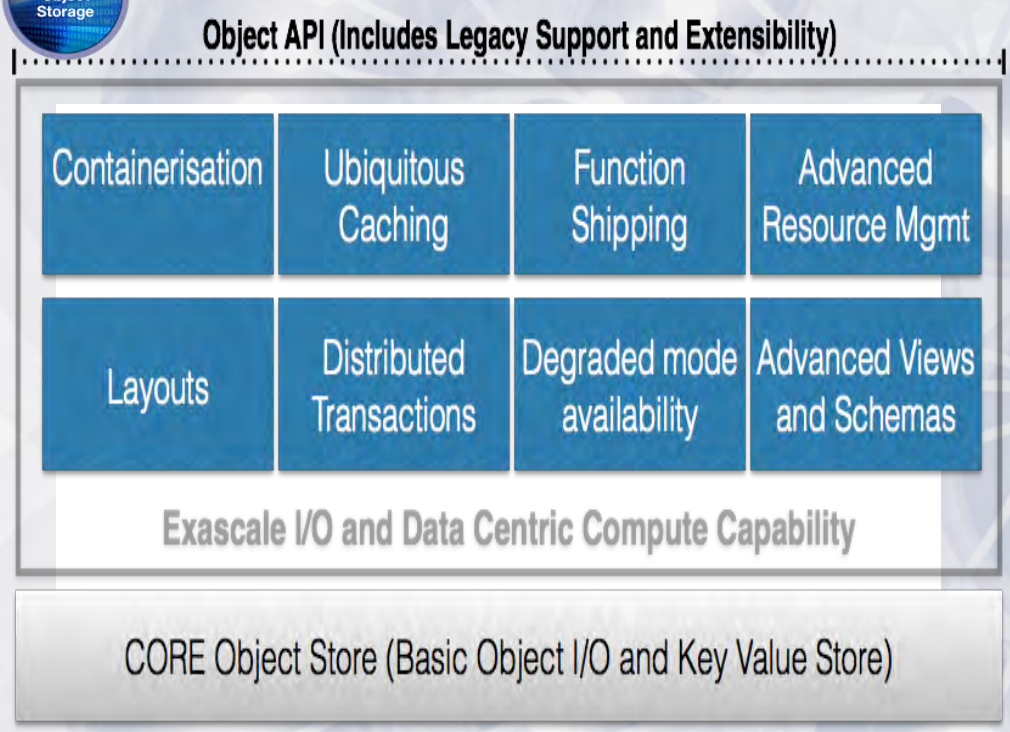
Applications & Co-Design



✓ Completed Co-Design Process – First Phase

SAGE

Applications & Co-Design..



- **Goal**
  - Object Storage Platform Development(w/ **Clovis** API)
  - Evaluate NVRAM Options

Advanced Object Store

Methodology

- Extreme I/O handling
- Infrastructure fault handling
- Enable Application Resiliency
- Handle Computations from apps
- Different views to same data
- Complex data layouts & caches in I/O hierarchy

- **Concept & Architecture of Key Mero components Available**
  -Good progress in key areas: Eg Function Shipping

SAGE

Object Store Extreme scale Features & NVRAM

- **Goal**
  - Explore tools and services on top of Mero
- **Methodology**
  - "HSM" Methods to automatically move data across tiers
  - "pNFS" parallel file system access on Mero
  - Scale out Object Storage Integrity checking service provision
  - Allinea Performance Analysis Tools provision

- ✓ **Completed Design of HSM**
- ✓ **Completed Design of pNFS**
- ✓ **Completed scoping/Arch of Data Integrity Checking**
- ✓ **Completed scoping/Arch of Performance Analysis Tools**

- **Goal**
  - Explore usage of SAGE by programming models, runtimes and data analytics solutions
- **Methodology**
  - Usage of SAGE through MPI and PGAS
    - Adapt MPI-IO for SAGE
    - Adapt PGAS for SAGE
  - Runtimes for SAGE
    - Pre/Post Processing
  - Volume Rendering
    - Exploit Caching hierarchy
  - Data Analytics methods on top of Clovis
    - Apache Flink over Clovis, looking beyond Hadoop
    - Exploit NVRAM as extension of Memory

- ✓ **MPI-IO for SAGE Design available**
- ✓ **Runtimes design available**
- ✓ **Framework for data analytics available**
- ✓ **PGAS for SAGE being studied**

SAGE

Programming Models and Analytics

- **Goal**
  - Hardware definition, integration and demonstration
- **Methodology**
  - Design and Bring-up of SAGE hardware
    - Seagate Hardware
    - Atos Hardware
  - Integration of all the software components
    - From WP2/3/4
  - Integration in Juelich Supercomputer Center
  - Demonstrate use cases
    - Extrapolate performance to Exascale
    - Study other Object stores vis-à-vis Mero

✓ **Design & Definition of base SAGE hardware complete**
Prototype being tested in Seagate

SAGE

Integration & Demonstration

- M1 (Sept 2015) – Project Start
- M3 – Draft System architecture
- M9 – Co-Design Inputs from Apps Complete
- M12 – Design of Key Software Components (1st Draft)
- M18 – Prototype system available (Seagate)
- M27 – Design of All Software Components (2nd Draft)
- M18 – M28 – Prototype system testing/system integration in Juelich
- M28 – M36 – Application use cases and performance tests of the SAGE Platform in Juelich

SAGE

Overall Project Timeline

Sage is extremely well aligned to the broader goals for Europe in the area of Storage, I/O and Energy Efficiency

- M-BIO-1: Tightly coupled **Storage class memory io systems** demo
- M-BIO-3: **Multi-tiered** heterogeneous storage system demo
- M-BIO-5: **Big data analytics tools** developed for hpc use
- M-BIO-6: '**Active Storage**' capability demonstrated
- M-BIO-8: Extreme scale **multi-tier data management** tools available
- M-ARCH-3: New compute nodes and **storage architecture use nvram**
- M-ENER –X: Addresses **Energy goals** by avoiding data movements
  - 100x more energy to move data compared to compute!!
- M-ENER-FT-10: Application survival on unreliable hardware

Alignment with European Goals [ETP4HPC SRA]

Primary European Storage Platform for Extreme Scale (Applicability: Big Science and BDEC)

Key inputs into European Road mapping

Key inputs into “Data Intensive” Research programs

Commercial and Market Impacts (Storage, Systems & Tools)

SAGE

Expected Impacts & Innovation

Questions ?

sai.narasimhamurthy@seagate.com