SAGE

Percipient StorAGE for Exascale Data Centric Computing

www.sagestorage.eu

Research

The SAGE project tasks itself with building a data centric infrastructure for handling **extreme data** in the **Exascale/ Exabyte era** centred on a storage oriented solution, 'Percipient Storage'. Sage will work with I/O intensive codes and data centric applications that enable the derivation of **scientific and business insights** from very large data sets. SAGE incorporates cutting edge innovations in an ecosystem of critical supporting research into highly distributed storage platforms, data analytics frameworks and programming tools.

SAGE Ecosystem

Extreme Data Management

- **Hierarchical** Data Management tools for Extreme Scale
- Extreme scale **data integrity** checking techniques

Extreme Data Analytics

- Tools for **Big Data** Analytics and adaptations for extreme scale HPC
- Exploiting the computation capability associated with **tiered data** and non volatile memory
- **Apache Flink** for batch and stream data processing

Programming Methods

- **MPI:** Offloading computational kernels to the different tiers of SAGE I/O
- MPI-IO: Adapting **MPI-IO** for **extreme object stores**
- PGAS: Integrating the **NVRAM tier** to the virtual memory of the compute nodes

Optimisation Tools

- Exposing the **true cost** of I/O at **scale**
- **Debugging** capabilities for computational **offload** to storage

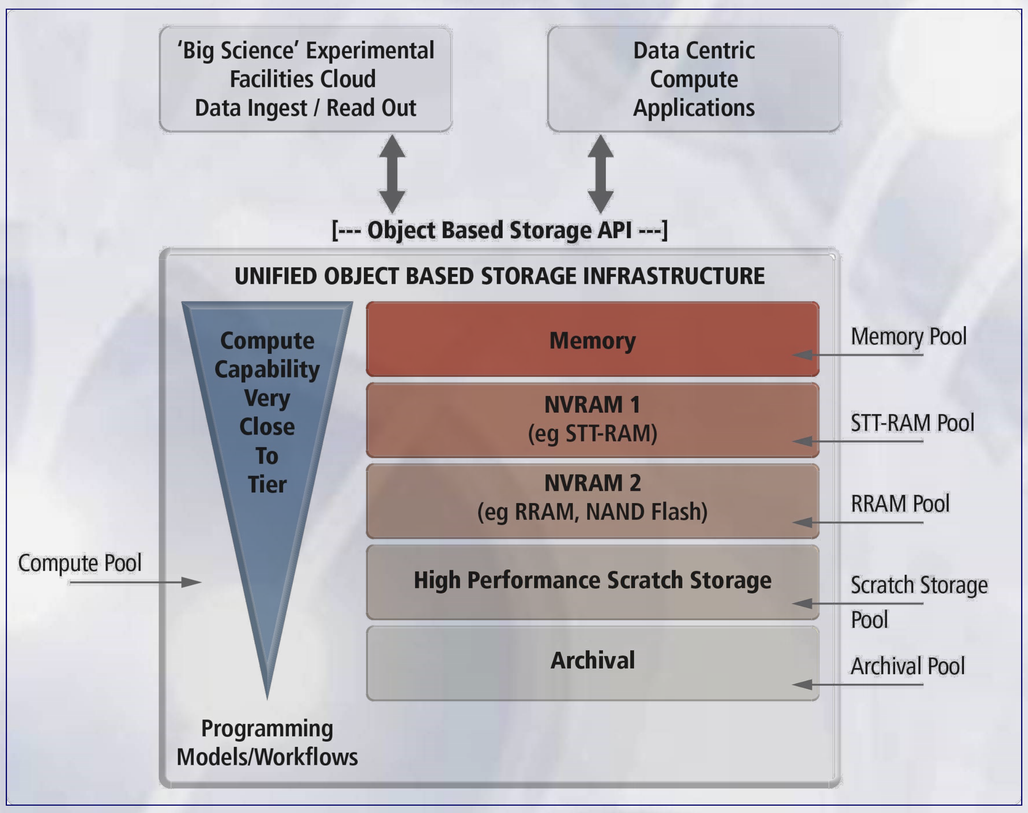
Percipient Storage

We propose an advanced object based storage solution, termed *Percipient Storage*.

- A new and **flexible API**
- Exploit **deep I/O** hierarchies
- **Computation** capabilities in **all tiers**
- Homogeneous view of data throughout the stack

The SAGE architecture (depicted below) aims to meet the needs of next generation storage solutions through:

- Improved **energy efficiency** by reducing data movement
- Incorporate the technology trend towards **non-volatile memory** technologies



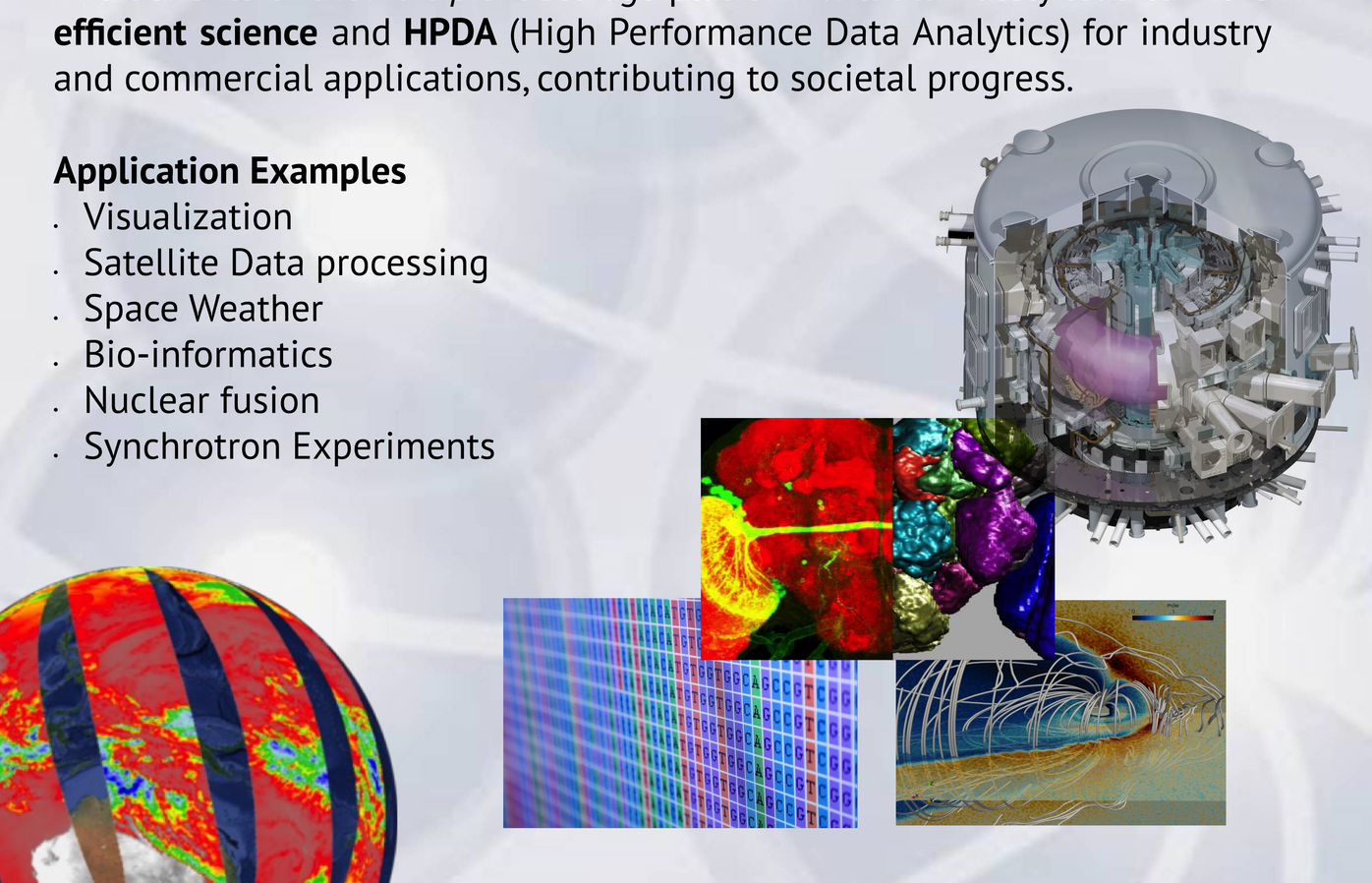
Co-Design

The SAGE Project will deal with use cases that are typical examples of **cutting edge** problems in the domains of **Extreme Computing I/O** and **Data Intensive Science.**

The benefits of the *Percipient Storage* platform will ultimately lead to more **efficient science** and **HPDA** (High Performance Data Analytics) for industry and commercial applications, contributing to societal progress.

Application Examples

- Visualization
- Satellite Data processing
- Space Weather
- Bio-informatics
- Nuclear fusion
- Synchrotron Experiments



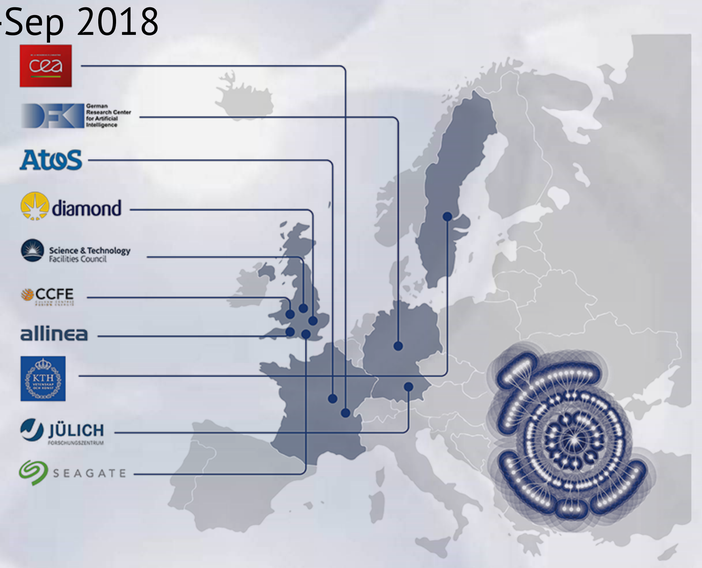
SAGE at a Glance

BUDGET: € 7.9 million

TIME PERIOD: Sep 2015-Sep 2018

CONSORTIUM:

- Co-ordinator: Seagate
- 10 Partners
- 4 European Countries



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