EMERGING TECHNOLOGIES: SELECTED RESULTS FROM EUROPEAN PROJECTS

Dirk Pleiter | ADAC6 Workshop | Zürich | 20.06.2018

JÜLICH
Forschungszentrum

Future and Emerging Technology Program of the EC

Political rational for H2020 program

- HPC is a strategic resource for Europe's future as it allows researchers to study and understand complex phenomena while allowing policy makers to make better decisions and enabling industry to innovate in products and services.

H2020 funding streams

- Future and Emerging Technology (FET)
  - HPC architecture and technology development projects
- Information and Communication Technologies (ICT)
  - Focus on lower-level technologies
- E-Infrastructure program (INFRA)
  - Pan-European High Performance Computing infrastructure and services (PRACE)
  - Centres of Excellence on HPC

Research agenda formulated by the ETP4HPC

Overview Running H2020 Projects

Excluding the Centres of Excellence

Compute

- **DEEP-EST**
- ECOSCALE
- EuroExa
- **ExaNoDe**
- **MB2020**
- Montblanc3

Interconnect

- ExaNeSt

Memory and storage

- NEXTGenIO
- **SAGE**

Mathematics

- ExaHYPE
- NLAFET

Algorithms

- Compat
- ExaFLOW
- ExCAPE

Data-intensive real-time

- MAGNO
- GreenFLASH

Programming tools

- AllScale
- ANTAREX
- ESCAPE
- INTERTWINE
- READEX

Upcoming projects

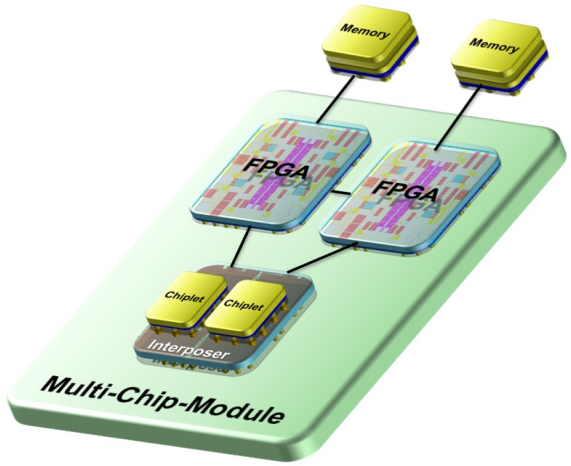
- **EPI**
- EXA2PRO
- **MAESTRO**
- SAGE2
- ….

PROCESSOR LEVEL

ExaNoDe: Overview

http://www.exanode.eu

ARM, BSC, Bull, **CEA**, ETHZ, FORTH, Fraunhofer, FZJ, Kalray, SCAPOS, U Manchester, VOSYS



Goal: Design of a new compute node architecture

- Design highly integrated heterogeneous computing device for HPC applications
- Exploit 3D silicon-on-silicon level fabrication technology
- Provide specification for integration into the larger HPC system

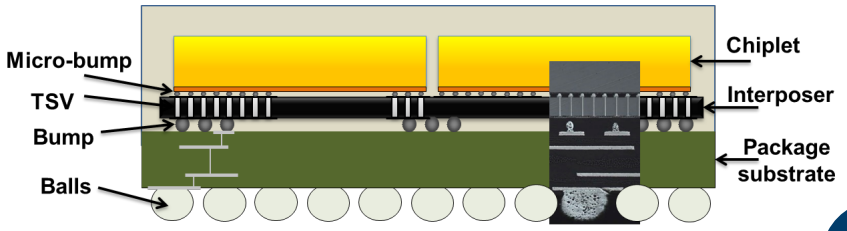
Key concepts

- System architecture
  - ARMv8
  - Coherent island
  - Global Address Space

} UNIMEM

- Silicon integration
  - 3D Integration (Chiplet, Active Interposer)
  - Multi-Chip-Module
  - FPGA, Memory

- System software
  - Firmware, operating system support
  - Virtualization
  - Programming models



Benefits of New Approaches to Integration

Challenge: The HPC market is a niche

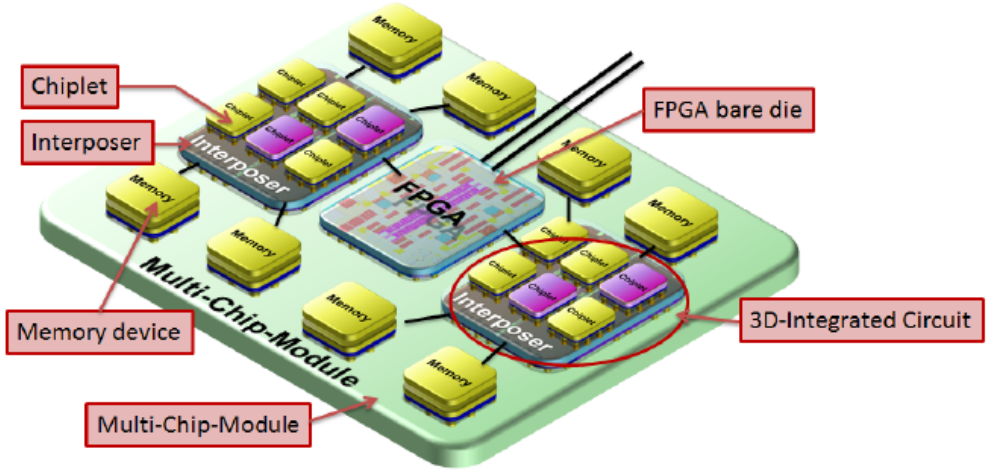
- Producing computing devices for just this market typically not affordable

ExaNoDe strategy

- Allow for flexible integration of different components on active/passive interposer
- Components
  - Compute chiplet and accelerators (e.g. FPGA)
  - Memory devices based on different memory technologies
  - Interconnect chiplet

Benefits

- Modular and customizable approach
- Increase of density



MB2020

http://montblanc-project.eu

ARM, BSC, **Bull**, CEA, FZJ, Kalray, Semidynamics

Project goals

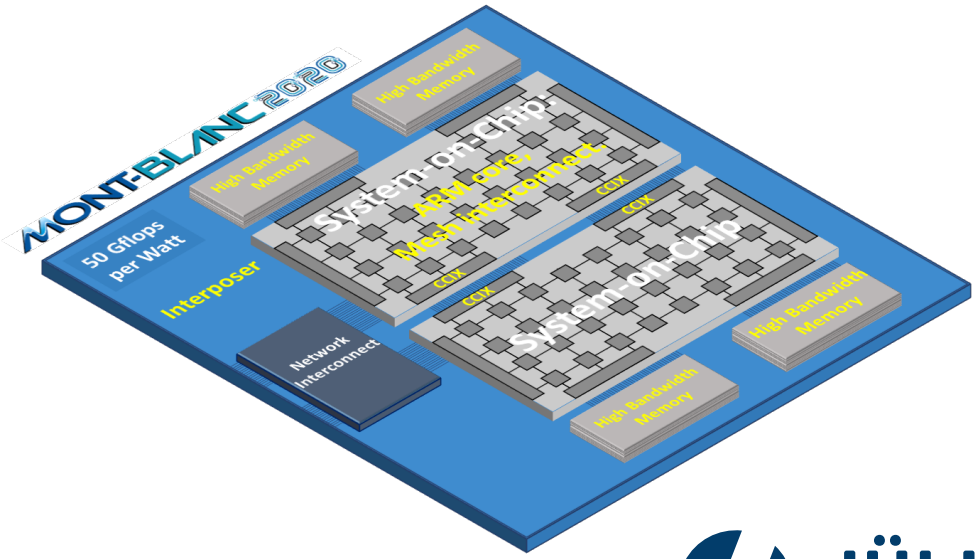
- Defining a low-power, ARM-based System-on-Chip architecture targeting Exascale
- Implementing new critical building blocks
- Delivering initial proof-of-concept demonstration of its critical components on real life applications

Focus on understanding trade-offs between

- Vector length and core count
- Intra-NOC interconnect performance
- Memory bandwidths/capacity
- Integration of accelerators

Key output: MB2020 demonstrator

- Emulator platform based on
  - Trace-based SVE traffic generators
  - RTL model of the NOC
  - Full memory hierarchy



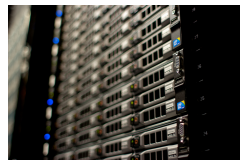
European Processor Initiative (EPI)

Expected impact

- Strengthening the competitiveness and leadership of European industry & science
- Covering important segments of the broader and/or emerging HPC and Big-Data markets

Targeted core technologies

- ARM-based CPU
- RISC-V-based (and other) accelerators



Target markets

- HPC
- Data centres and servers
- Autonomous vehicles

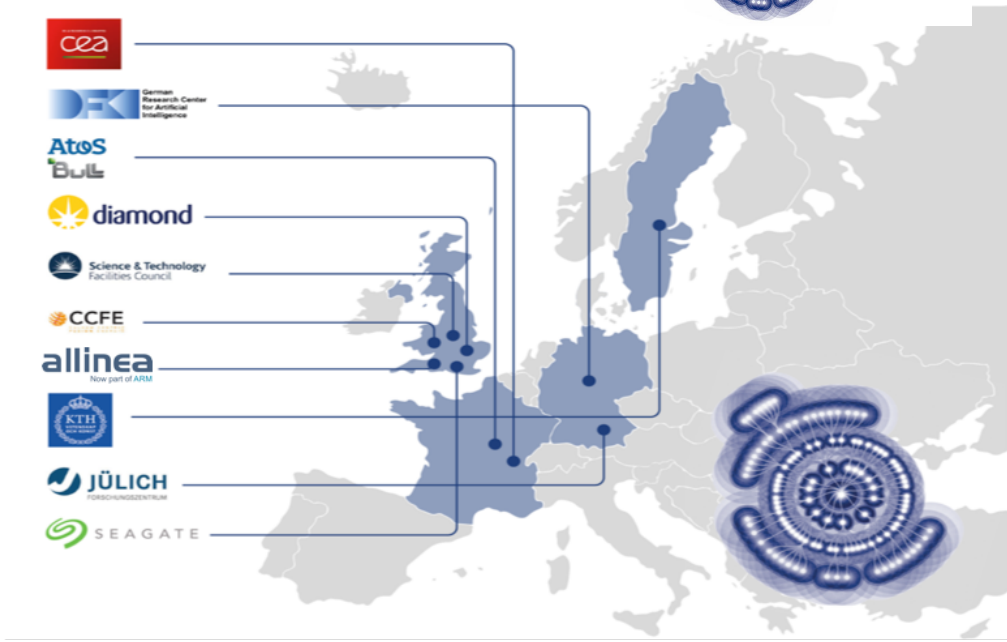
Status

- Framework Partnership Agreement awarded, project about to start

SYSTEM LEVEL

SAGE: Overview

http://www.sagestorage.eu/



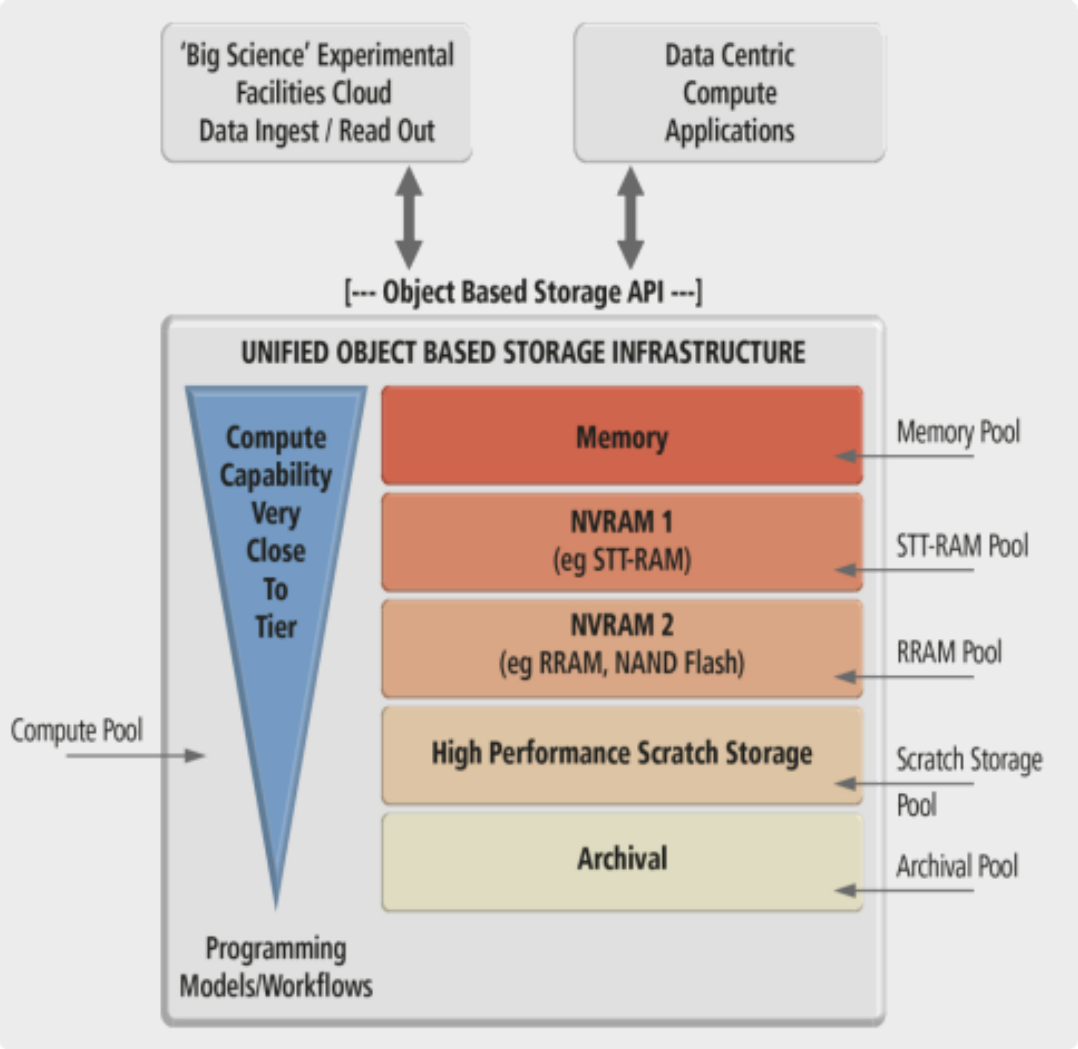
Goal

- Build a data centric computing platform
  = storage platform with integrated compute

Approach

- Create hierarchical storage architecture based on
  - Advance object storage technology = MERO
  - Multiple tiers with storage devices featuring different characteristics
  - Integrated compute capabilities
- Make new architecture usable

SAGE: Hardware Architecture



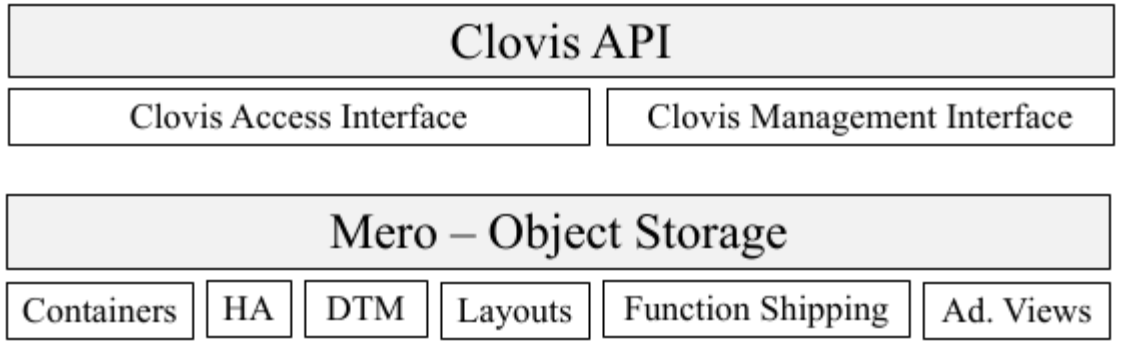
External user access program

SAGE prototype @ JSC

SAGE: Mero Architecture

Native object storage platform

- Scalable re-writable fault-tolerant data objects
- Index store with key-value indices
- Resource management capabilities



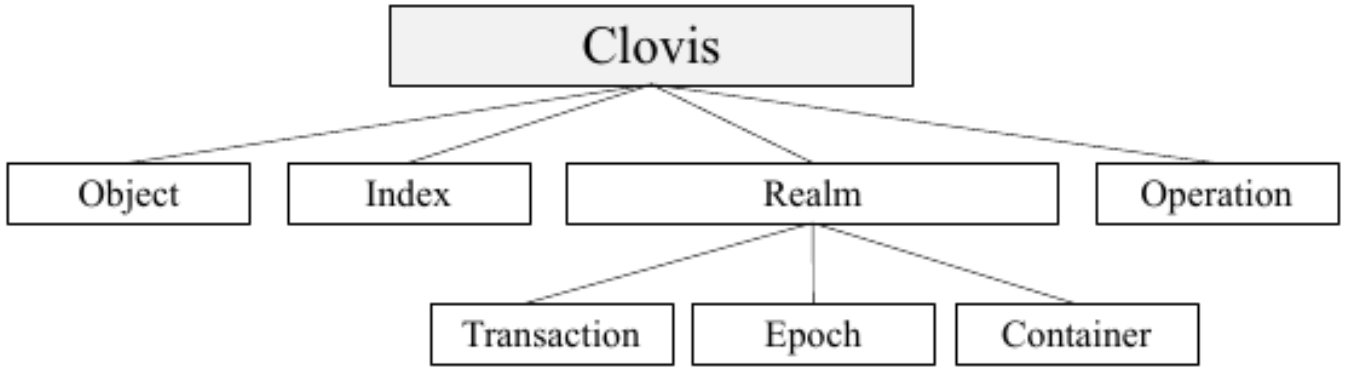
Key features

- Container abstraction
  - Allows for grouping of objects
- High-availability features
- Distributed transaction management
  - All I/O and metadata operations are organised as transactions
  - Transactions are atomic with respect to failures
- Layouts
  - Allow for mapping of different parts or regions of an object to storage tiers

SAGE: Clovis Interface

Clovis = API for Mero

- Access interface
- Management interface



Provided abstractions

- Object = array of fixed-size blocks of data
- Index = KVS
- Operation = process of querying and/or updating the system state
- Realm = spatial and temporal part of the system with a prescribed access discipline
  - Objects, indices and operations live in realms

SAGE: Selected Features

Different APIs

- Native object store
- POSIX via pNFS
- Native HDF5 support

Hierarchical storage manager

- Automatic movement of data across tiers

Performance tools

- Telemetry records available via Clovis interface

Data analytics frameworks integration

- Apache Flink

MPI-IO support

- Function shipping and in-storage compute

PGAS support

- Addressable access to storage devices

Function shipping

- Processing near the data

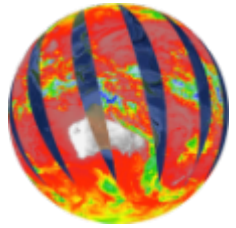
Run-time system

- Support for in-storage compute and steering

Publications

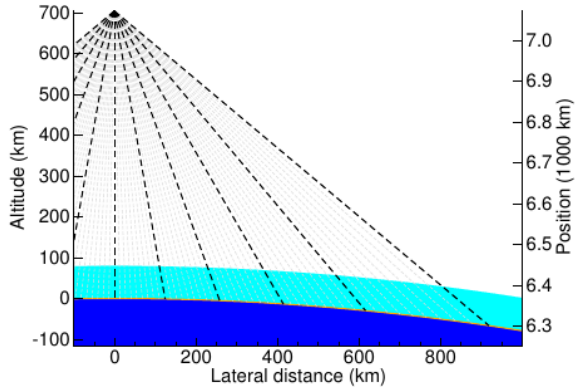
- S. Narasimhamurthy et al., 10.1016/j.parco.2018.03.002
- SAGE Early Experiences White Paper

SAGE: Run-time System Example



Science case: processing of large-scale satellite data

- Workflow steps
  - Read data object
  - Retrieve geophysical data (pressure, temperature, trace gas concentration) from inverse modelling



Semi-persistent cache approach

- Manage list of to be and already processed data objects in storage
- Cache manager pre-fetches data objects to cache
- Application queries list of objects available in cache

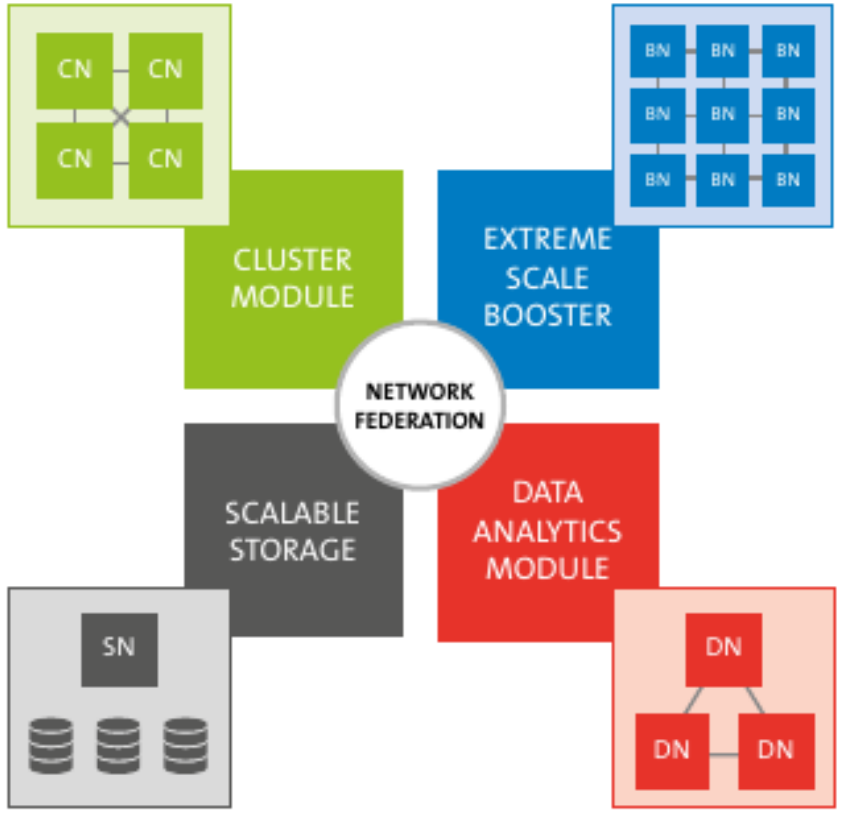
Targeted benefit of data-centric approach: Data available in fast storage tier when needed

DEEP-EST: Overview

http://deep-projects.eu/

Goals

- Develop an energy efficient system architecture that fits
  - HPC workloads and
  - HPDA workloads
- Build a fully working Modular Supercomputing Architecture system prototype made-up of three modules:
  - Cluster module
  - Extreme Scale Booster module
  - Data Analytics module



Development challenges

- Resource management and scheduling system
- Enhance and optimise programming models

Upcoming project: MAESTRO

MAESTRO
DATA ORCHESTRATION

Appentra, CEA, Cray, ECMWF, ETHZ/CSCS, FZJ, Seagate

Motivation

- Hierarchical memory and storage architectures will become generally available
- More data-intensive and complex workflows are in need of scalable compute resources

Typical shortcomings of today's middleware frameworks

- Lacking data awareness
- Lacking memory awareness

Project goals

- Develop a middleware providing consistent data semantics to multiple layers of the stack
- Demonstrate progress for applications through **memory-and-data-aware (MADA)** orchestration
- Enable and demonstrate next-generation systems software MADA features
- Improve the ease-of-use of complex memory and storage hierarchy

MAESTRO: Concepts

MAESTRO
DATA ORCHESTRATION

Core Data Model

- Logical Data Objects
  - Abstract description comprising references to Physical Data Objects + associated metadata
- Physical Data Objects
  - Description of concrete physical locations of copies of data
- Allow for different views on data referenced by a Logical Data Object
  - Example: row-major vs. column-major view

Memory System Model

- Abstract representation of physical memory and storage media
  - Captures few, key performance characteristics
- Physical Data Objects are assigned to an abstract memory level
  - Enable middleware to manage data transport

SUMMARY AND CONCLUSIONS

Summary and Conclusions

Broad landscape of HPC development projects

- Ranging from processor design to system-level hardware and software architecture

Key challenge: productisation and exploitation of results for supercomputing

- Many projects mainly driven by academic partners
- Co-design is a promise, but difficult to realise with desired results under given constraints