OBJECTIVES & BENEFITS

OBJECTIVES DEEP-EST PROJECT:

- Build a fully working, **energy efficient prototype** of the Modular Supercomputing Architecture (MSA).
- Extend a proven **resource management** and **scheduling system** to fully support the MSA.
- Enhance and optimise the **programming environment** based on **MPI** and **OpenMP,** and add support for **data analytics** and **machine learning** frameworks.
- **Validate** the full **hardware (HW) / software (SW) stack** with relevant HPC and HPDA applications clearly demonstrating the MSA benefits.
- **Accelerate** and **support take-up** and further development of **key European technologies** in e.g. network fabrics, system integration and system software.

BENEFITS OF MSA AND THE DEEP-EST SYSTEM:

- **Mix & Match:** heterogeneous applications/workflows run on exactly matching compute and data resources.
- **Highest flexibility:** each user selects the combination of resources best suited for its code.
- **Significant reductions** in time and energy to solution.
- Improvements in **system throughput** for heterogeneous workload mixes.

IN SHORT: **The MSA is an ideal fit** for advanced supercomputing centres.

IN A NUTSHELL

BUDGET:

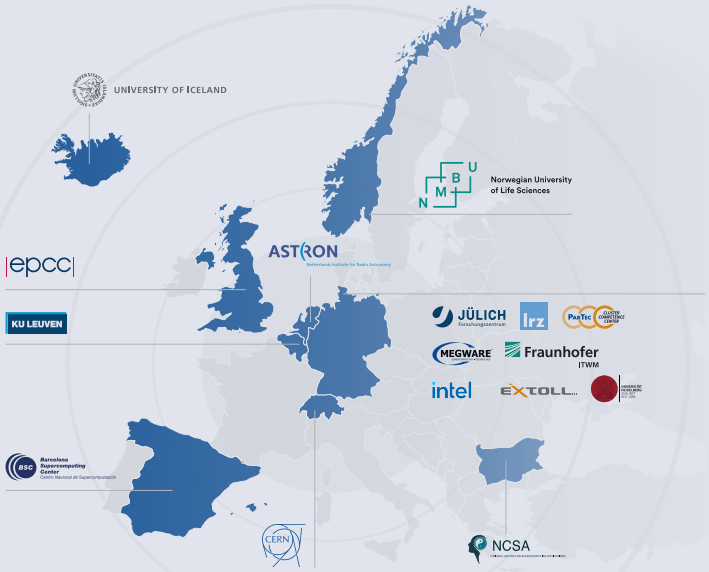
- € 15 Mio. European Union funding

PROJECT TERM:

- July 2017 – March 2021

CONSORTIUM:

- Coordinator: Jülich Supercomputing Centre
- 16 Partners
- 8 European countries



GET CONNECTED

Email: pmt@deep-projects.eu
Web: www.deep-projects.eu
Twitter: @DEEPprojects
Facebook: fb.com/deepprojects
LinkedIn: www.linkedin.com/groups/DEEPprojects-6534965

This project has received funding from the European Commission's Horizon 2020 Programme for research, technological development and demonstration under Grant Agreement n° 754304.

DEEP

Extreme Scale Technologies

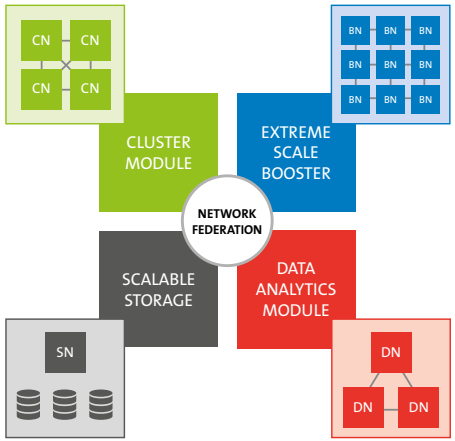
TOWARDS A MODULAR SUPERCOMPUTING ARCHITECTURE

THE MODULAR SUPER-COMPUTER ARCHITECTURE

The Modular Supercomputer Architecture (MSA) is a blueprint for heterogeneous HPC systems supporting the divergent computation and data processing requirements of high performance compute and data analytics with highest efficiency and scalability. The DEEP-EST prototype integrates three compute modules with different performance characteristics:

- **Cluster Module (CM):** reliable performance for all codes, including complex and irregular data structures and control structures.
- **Extreme Scale Booster (ESD):** highest delivered performance for regular and vectorizable codes, plus high energy efficiency.
- **Data Analytics Module (DAM):** highest performance for data analytics and machine learning codes.

The modules are connected to each other via a **Network Federation (NF)** solution. **Network Attached Memory (NAM)** and **Global Collective Engine (GCE)** nodes offer fast globally accessible memory and acceleration of MPI collectives. Finally, storage is provided by the **Scalable Storage Service Module (SSSM).**



SOFTWARE ENVIRONMENT

DEEP-EST provides an **integrated programming environment,** using standards such as MPI, OpenMP and Tensorflow or similar frameworks for machine learning and data analytics. In the DEEP-EST project, R&D activities focus on:

- **SLURM resource manager/scheduler:** Extensions will help to determine the optimal resource allocation for each workload.
- **Adaptive scheduling and dynamic resource reservation:** Aggregation will be used to achieve efficient use of the overall system.
- **I/O and resiliency techniques:** Proven developments of the DEEP-ER project will be adapted to the MSA.
- **Scalability projections** to pre-Exascale performance levels will be provided.

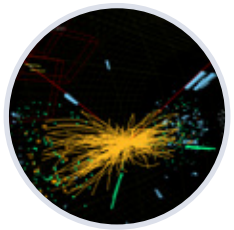
ENERGY EFFICIENCY

The project addresses **energy efficiency** on multiple layers of the HPC system:

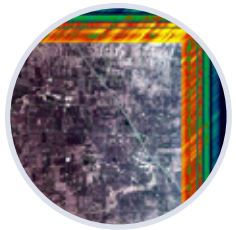
- **Highly efficient code execution:** application requirements are matched with available system resources.
- **Optimal system throughput** leads to reduced time and energy to solution.
- **Efficient system operation** achieved by leading-edge hardware and use of direct warm-water cooling is combined with a **sophisticated monitoring system.**

CO-DESIGN APPLICATIONS

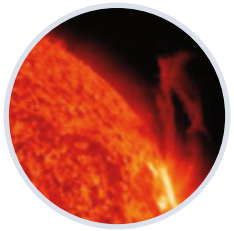
Six ambitious HPC and HPDA applications drive the co-design process and will be used to evaluate the HW and SW technologies developed in DEEP-EST. The application variety will exploit the system flexibility, each one utilizing different module combinations and show that the Modular Supercomputer Architecture is beneficial for a wide range of users. The dynamic scheduling and resource management extensions will ensure highest throughput of applications.



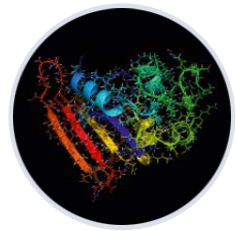
High Energy Physics



Earth Science

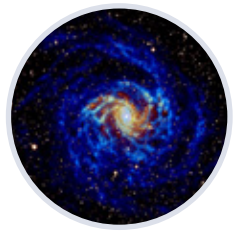


Space Weather



Molecular Dynamics

Neuroscience



Radio Astronomy

evolve

Leading the Big Data Revolution

Big data, a big opportunity for growth and innovation

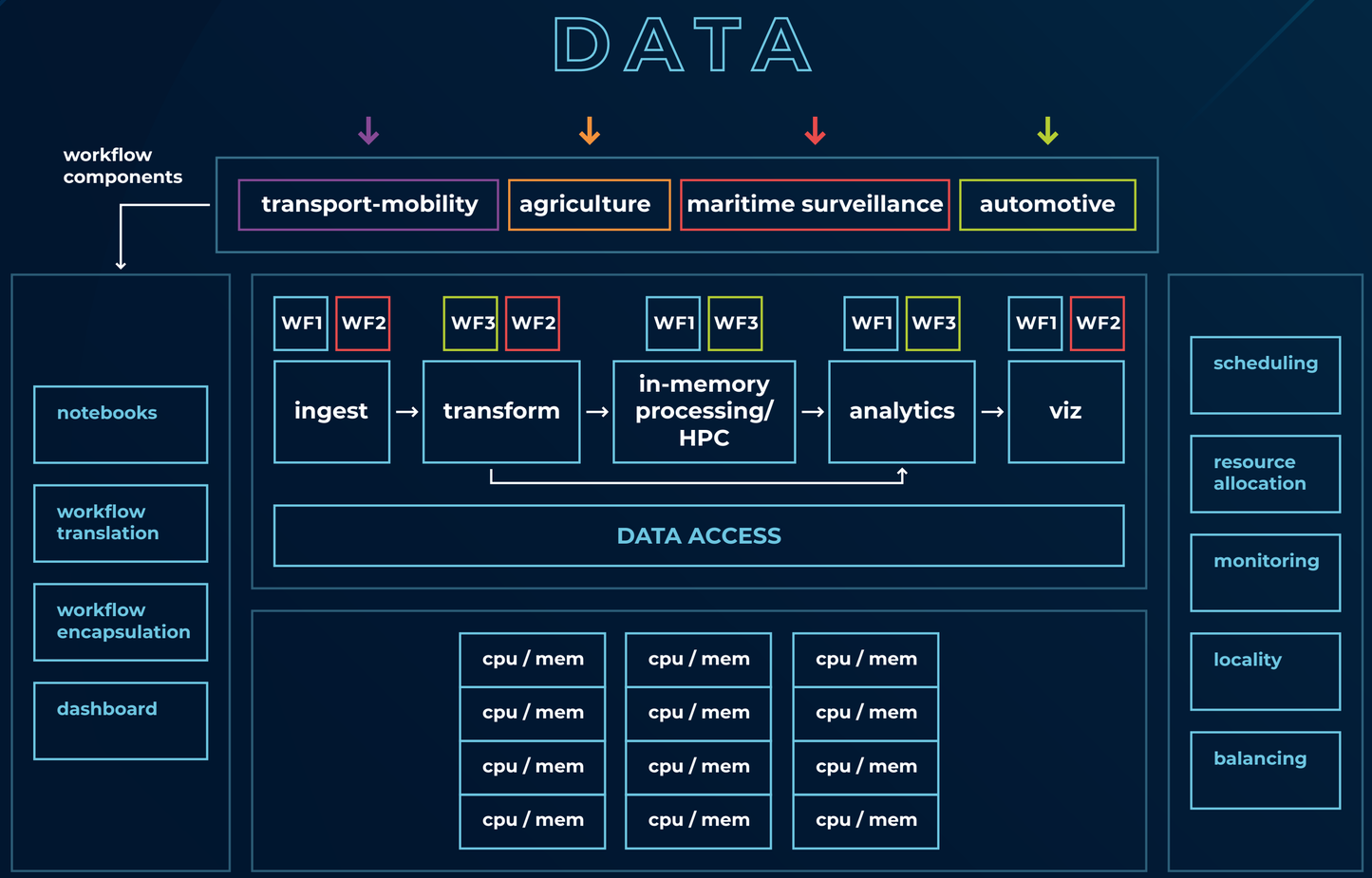
As data becomes the centre of innovation in modern economy and society, we start to face new challenges and limitations. Although tremendous progress has happened over the past several years on increasing productivity for data processing over commodity systems and providing new services with Big Data and Cloud technologies, the projected data deluge brings business, consumers, and the society in general at a new frontier: how can we process massive data that require demanding computation?

Creating new data-intensive services in terms of dataset size and data processing is an onerous and costly process that requires deep expertise. It requires high performance beyond what commodity systems can achieve, describing business logic typically by writing applications code, complex software stacks that are hard to deploy and maintain, and the need to use dedicated, per application, testbeds for achieving the desired performance levels. However, most organisations today lack these resources and the associated expertise.

EVOLVE is addressing these issues as it offers new (High Performance Computing) HPC-enabled capabilities in data analytics for processing massive and demanding datasets without requiring extensive IT expertise.

Leading the Big Data Revolution

At the centre of EVOLVE lies an advanced HPC-enabled testbed that is able to process unprecedented dataset sizes and to deal with heavy computation, while allowing shared, secure, and easy deployment, access, and use.



EVOLVE's testbed is a concrete step in bringing together features from the Big Data, High Performance Computing (HPC) and Cloud worlds, directly contributing to their convergence.

HPC

An advanced computing platform with HPC features and systems software.

Big Data

A versatile big data processing stack for end-to-end workflows.

Cloud

Ease of deployment, access, and use in a shared manner, while addressing data protection.

Evolve's **testbed** is based on:

Advanced Computing Platform

The main aspects of EVOLVES's hardware platform are its large scale, fast interconnect and memory. EVOLVE core architectural contribution is harnessing accelerators. The testbed will support accelerated nodes by GPU, FGPA and specialised processors.

Storage Subsystem Architecture

In EVOLVE the storage is envisioned as a tiered architecture. The storage subsystem uses a shared Infinite Memory Engine (IME) and fast local "non-volatile memory express" storage devices. Storage will be extended with advanced data protection, compression an encryption features.

Safety & Ease of Deployment, Access & Use:

EVOLVE will provide shared access to the testbed for improving productivity and Total Cost of Ownership (TCO). Cloud native technology will be used for the deployment of containerised high performance applications. End-to-end encryption will ensure safety and privacy.

End-to-end Workflows

EVOLVE is using end-to-end workflows that express full data-processing pipelines, including data ingest from external sources with time constraints. The Extract-Transfom-Load process will be fully supported for all pilot applications.

Versatile Software Stack

To realize workflows, EVOLVE is providing a versatile software stack that employs existing data processing engines that have proven flexibility and breadth of applicability.

Performance

Reduced turn-around time for domain-experts, industry (large and SMEs) and end-users.

Experts

Increased productivity when designing new products and services, by processing large datasets.

Businesses

Reduced capital and operational costs for acquiring and maintaining computing infrastructure.

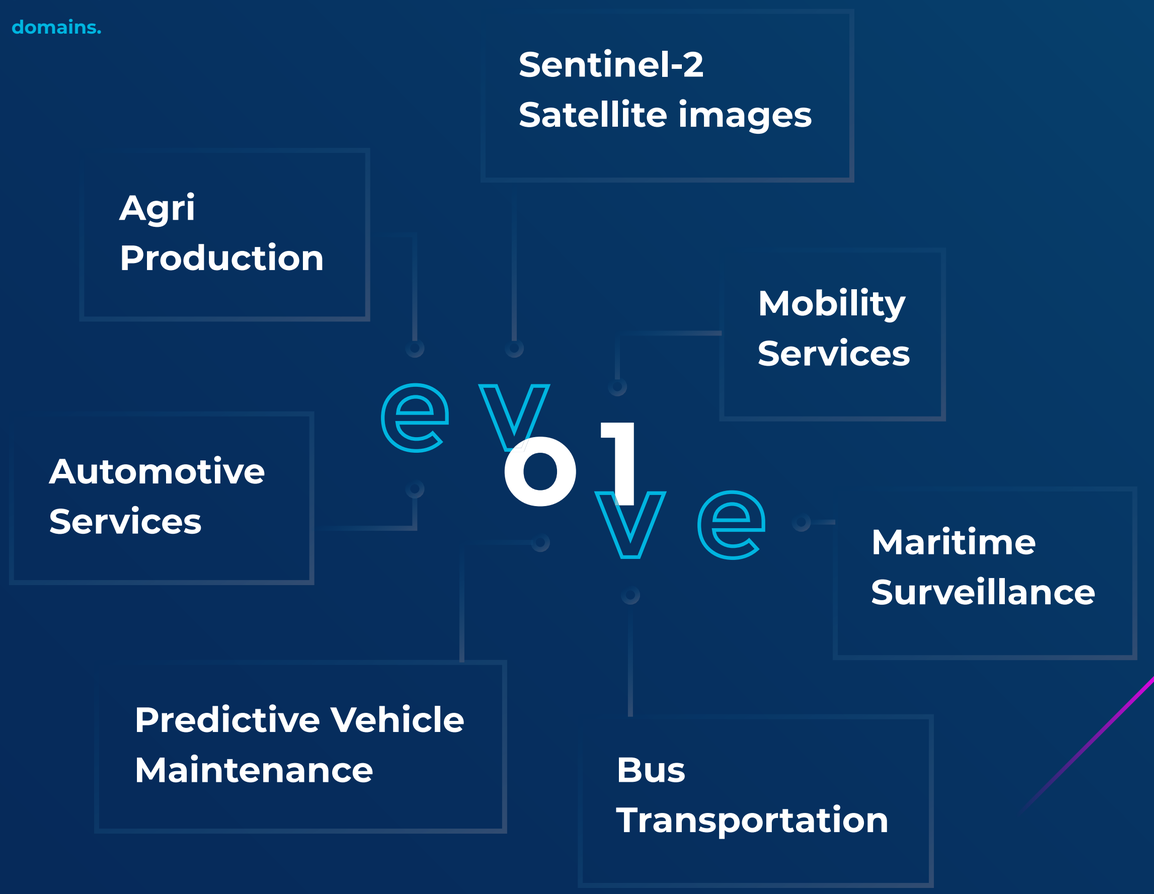
Society

Accelerated innovation via faster design and deployment of innovative services that unleash creativity.

7 Pilots & Domains

putting EVOLVE's testbed into practice

The benefits of EVOLVE's testbed will be demonstrated through pilots implemented in seven domains.



In all cases, domain experts are working on models that provide accurate predictions, data processing and validation techniques over massive datasets and have the potential to improve substantially the efficiency of existing or introduce new services in the respective domains.

When technology meets social needs

EVOLVE will use technologies in markets where data capability is already the source of disruption, or is the turn point of being disrupted, these markets, used as case studies in EVOLVE, are socially critical for European citizens, like mobility (autonomous vehicle, ground mass transportation, maritime transport), agriculture and urban planning.

EVOLVE is not a pure technology project but frames itself in the more global perspective of data ownership in an open society

. Jean-Thomas Acquaviva, DDN Storage .
coordinator of the EVOLVE Project

Consortium

DDN STORAGE
www.ddn.com

BULL
www.atos.net

IBM
www.ibm.com

FORTH
www.ics.forth.gr

OnApp
ww.onapp.com

Institute of communications and computer systems
www.microlab.ntua.gr

MemoScale
www.memoscale.com

webLyzard technology
www.weblyzard.com

LOBA
www.loba.pt

Thales Alenia Space
www.thalesgroup.com

Space Hellas
www.space.gr

CybeleTech
www.cybeletech.com

Neurocom Luxembourg
www.neurocom.eu

MemEX
www.memexitaly.it

Tiemme SPA
www.tiemmespa.it

Virtual Vehicle
www.v2c2.at

AVL List GmBH
www.avl.com

BMW AG
www.bmw.com

KOOLA
www.koola.io

Follow us on:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825061

Contact Us
info@evolve-h2020.eu

Abstract

The LEXIS project is building an advanced engineering platform at the confluence of HPC, Cloud and Big Data, which leverages large-scale geographically-distributed resources from existing HPC infrastructure, employs Big Data analytics solutions, and augments them with Cloud services.

Driven by the requirements of several pilot testcases, the LEXIS platform relies on best-in-class data management solutions and advanced, distributed orchestration solutions, augmenting them with new, efficient hardware and platform capabilities in the form of Data Nodes and federation, usage monitoring and accounting/billing supports to realize an innovative solution.

Use cases

AERONAUTICS

This pilot aims to assess the industrial applicability of LEXIS through sophisticated CFD analyses referring to aircraft engines turbomachinery and rotating parts.

WEATHER AND CLIMATE

The Weather & Climate Use Case focuses on a complex system, to provide a diverse set of forecasts: weather, flood, fire, energy, air pollution.

EARTHQUAKE AND TSUNAMI

The Earthquake and Tsunami large scale pilot is about building with LEXIS a time-constrained HPC emergency response simulation and decision support flow.

Contacts

COORDINATOR
Jan Martinovic
IT4Innovations, VSB – Technical University of Ostrava

CO-DESIGN MANAGER
Olivier Terzo
LINKS Foundation

GENERAL CONTACT TO THE LEXIS PROJECT
info@lexis-project.eu

@LexisProject
@LEXIS2020
LEXIS Project

Consortium

COORDINATOR
Supercomputing Centre
VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

Atos — Industry
LINKS — Research Organization
EIFFAGE TESEO — Industry
ECMWF — Supercomputing Centre
Irz — Supercomputing Centre
cea — Research Organization
IPAC — Research Organization
Avio Aero — Industry
cima — Research Organization
GFZ Helmholtz-Zentrum Potsdam — Research Organization
AWI — Research Organization
Outpost24 — SME
NUMTECH — SME
BAYNCORE LABS — SME
CYCLOPS — SME

Project Ref. H2020-825532
Starting Date: 01/01/2019
Ending Date: 31/12/2021

LEXIS PROJECT

Large-scale EXecution for Industry & Society

INDUSTRIAL INNOVATION Targets

lexis-project.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825532

Application domains

HETEROGENEOUS DATA STORAGE | HPC-AS-A-SERVICE | SPEED UP | OPTIMIZE DATA MANAGEMENT OPERATIONS & ANALYTICS ALGORITHMS | SIMPLE AND SECURE HPDA SERVICE | INTEROPERABILITY WITH EXTERNAL DATA SOURCES

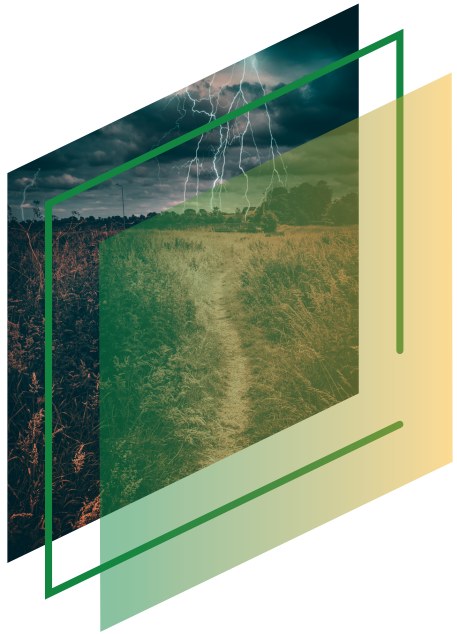
Foundation | Innovation | Extension

AERONAUTICS

Benefits by LEXIS

- By leveraging Lexis Technology, the goal of Aeronautics Large-scale pilot is to obtain a marked step-change in the context of turbomachinery and rotating parts engineering: less time-consuming simulations for the former one while a completely newly developed CFD methodology for the latter one, but always based on GPU-accelerated HPC platforms.
- Furthermore, the big data produced as a result will require proper solutions to be put in place for quick data access, management and post-processing.

WEATHER AND CLIMATE



Benefits by LEXIS

- This system will exploit Copernicus and GEOSS data services, supplemented and complemented by global and local in-situ unstructured observations.
- Data will be filtered and pre-processed before being assimilated to ensure quality. Moreover, each layer in the use case will produce its own large data set of forecast data, varying from hundreds of Terabytes for global weather models, to Megabytes at the decision maker level.

EARTHQUAKE AND TSUNAMI

Benefits by LEXIS

- The Tsunami and Earthquake large scale pilot builds an event-driven, time-constrained, always up to date flow aimed at providing timely warnings and well informed decision support for tsunami and earthquake emergencies.

MAESTRO

DATA ORCHESTRATION

Middleware for Memory and Data-Awareness in Workflows

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

cea

appentra make code parallel

ETH zürich

CSCS

ECMWF

SEAGATE

CRAY a Hewlett Packard Enterprise company

Maestro consortium is building a middleware library that characterises application data, reasons about how to load and store that data, assesses the cost of moving it and automates data movement across diverse memory systems

Motivation

- HPC and HPDA workloads are more and more I/O-intensive
- Performance bottlenecks are usually in the memory and storage systems
- Reducing and minimising data movement is very hard in general
- The HPC software stack was designed in a different era, to solve a different problem
- Few abstractions exist that capture data semantics of applications, so reasoning about data movement and memory in software is impossible
- Few useful models of memory systems and data movement exist, so estimation of cost of data movement is hard
- The memory-storage hierarchy is becoming more heterogeneous and complex, so a unified API and automatic promotion are needed

Approach and Methodology

- Co-design: ascertain data movement and access requirements of target applications
- Develop new data-aware abstractions:
  - Used in any level of software (compiler, runtime, application)
  - Relevant for any type of data (array, file, unspecified)
- Design a middleware and library that enables:
  - Modelling of memory hierarchy
  - Reasoning based on cost of moving data objects
  - Automatic movement and promotion of data in memories
  - Powerful data transformations and optimisation
- Explore data-based performance portability of Maestro applications

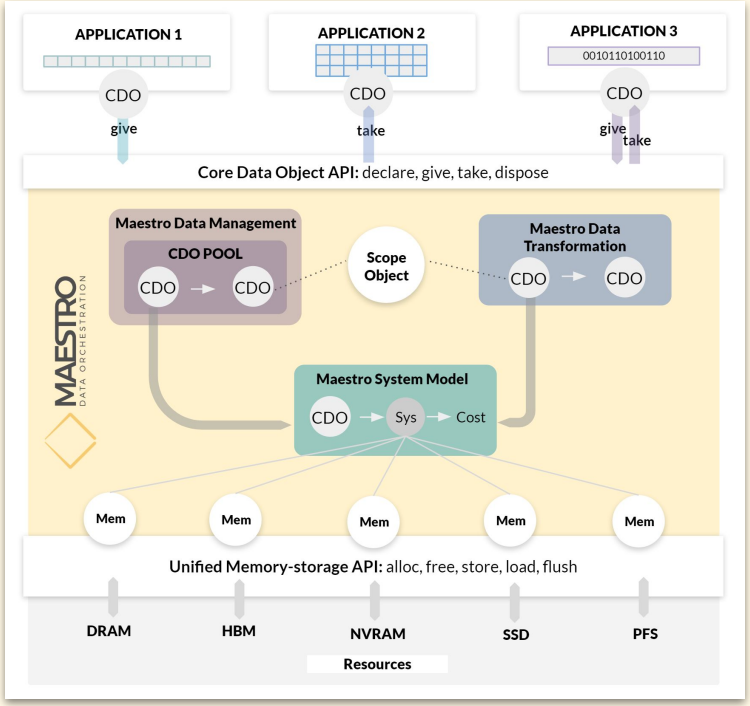


Figure: Design of the Maestro middleware. The CDO (Core Data Object) is at the heart of Maestro's design. It is used to encapsulate data and meta-data.

This project has received funding from the European Union's Horizon 2020 research and innovation program through grant agreement 801101.

The landscape for extreme computing and big data analysis is changing with AI/ Deep learning workflows augmenting the proliferation of enormous volumes of data created by scientific instruments and sensors, in addition to data from simulations. Sage2, follow on to the SAGE H2020 project, co-ordinated by Seagate, intends to validate a next generation storage system building on top of the already existing SAGE platform to address new use case requirements in the areas of extreme scale computing scientific workflows and AI/deep learning.

Sage2 will also provide global memory addressing capability to persistent storage resources and include new arm based in-storage processing environments.

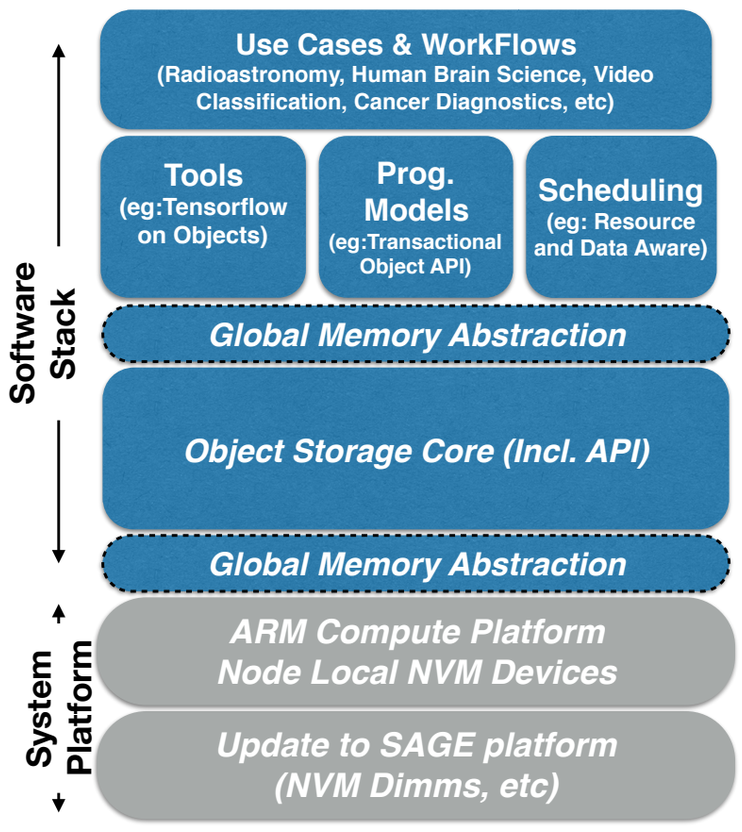
Contact: info@sagestorage.eu/Twitter: @sagestorage/ Web: www.sagestorage.eu

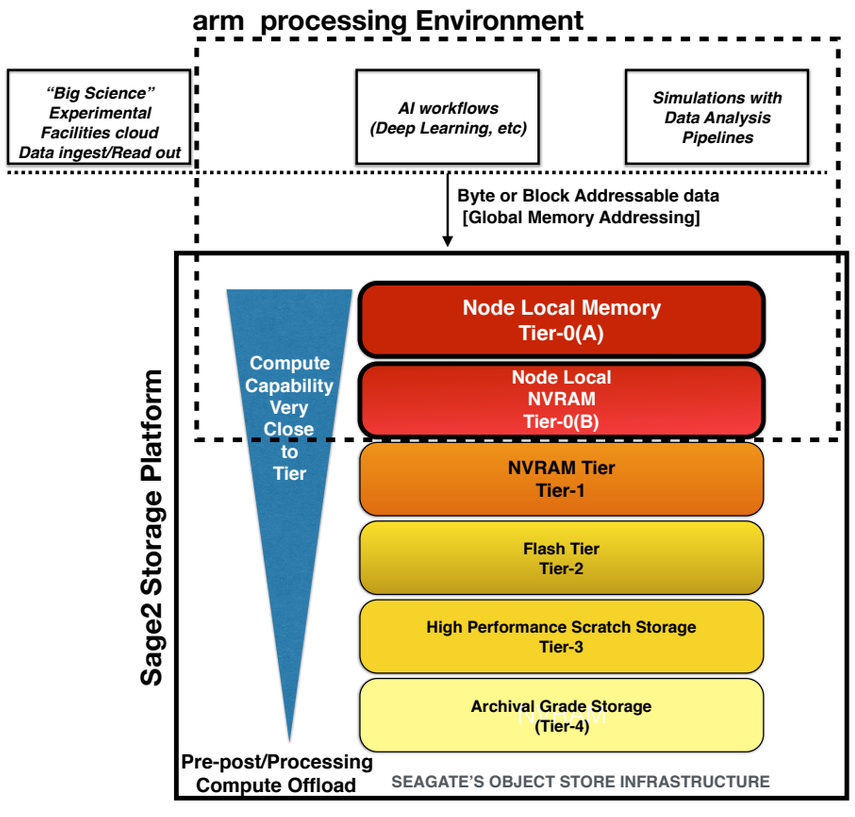
epcc | JÜLICH Forschungszentrum | Kitware

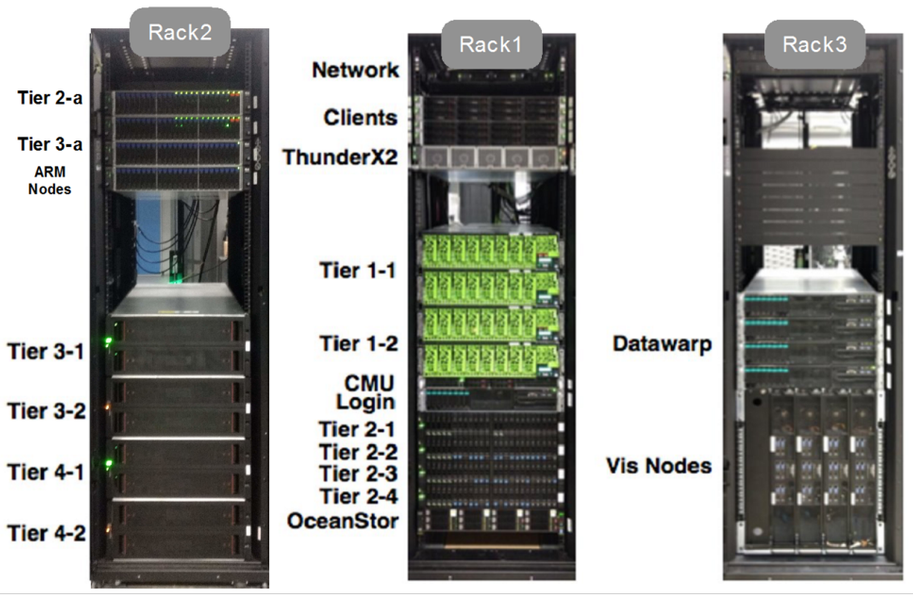
CCFE CULHAM CENTRE FOR FUSION ENERGY | Sage2 | KTH VETENSKAP OCH KONST

cea DE LA RECHERCHE À L'INDUSTRIE

SEAGATE | Bull atos technologies | arm







Development platform (Juelich Supercomputing Centre)

The project has received funding from the European Union's Horizon2020 Research & Innovation Programme under grant agreement 800999. The project is part of H2020-FETHPC2-2016-2017 (Transition to Exascale Computing) call.
Size: €3.997M, Duration: 36 Months starting September 2018

ExaQUte

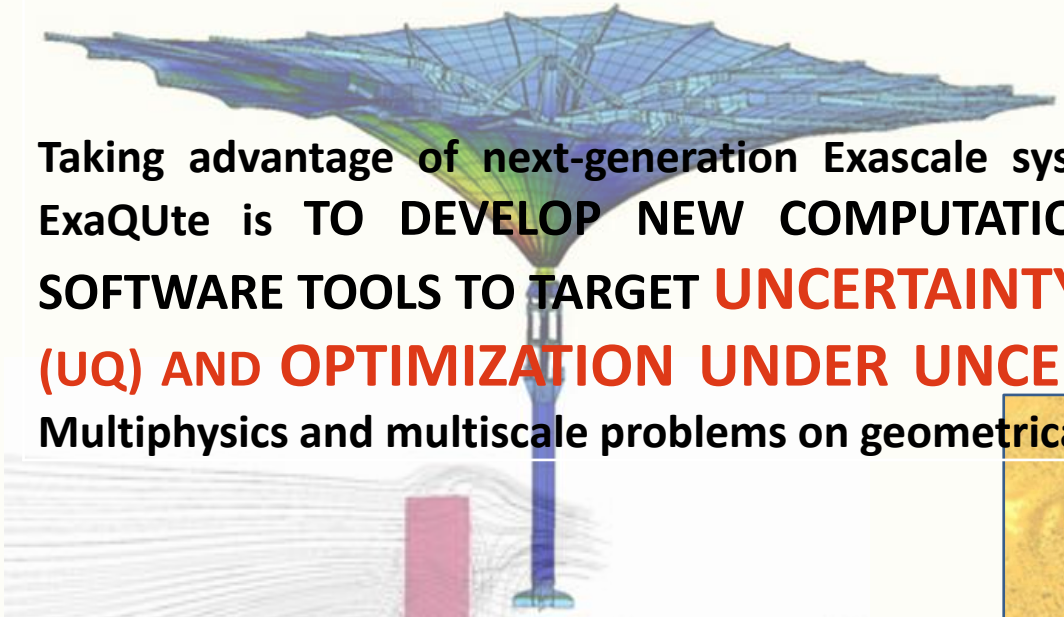
Exascale **Q**uantification of **U**ncertainties for **Te**chnology and Science Simulation

CIMNE
EXCELENCIA SEVERO OCHOA
BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación
Inria INVENTEURS DU MONDE NUMÉRIQUE
VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER
EPFL
UPC UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
str.ucture LIGHTWEIGHT DESIGN. MADE IN STUTTGART.

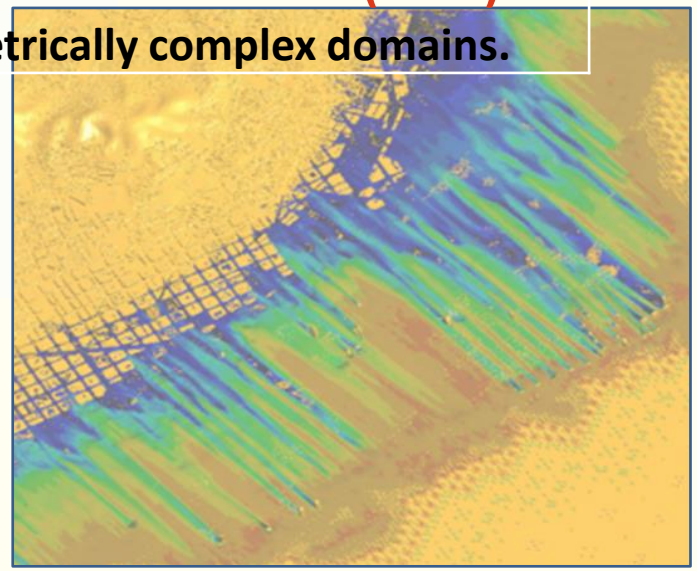
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No **800898**

THE GOAL

ExaQUte GENERAL OBJECTIVE



Taking advantage of next-generation Exascale systems, the main goal of ExaQUte is TO DEVELOP NEW COMPUTATIONAL METHODS AND SOFTWARE TOOLS TO TARGET UNCERTAINTY QUANTIFICATION (UQ) AND OPTIMIZATION UNDER UNCERTAINTIES (OUU) for Multiphysics and multiscale problems on geometrically complex domains.



ExaQUte

RATIONALE

HPC SYSTEMS HAVE AN INCREASING LEVEL OF PARALLELISM.

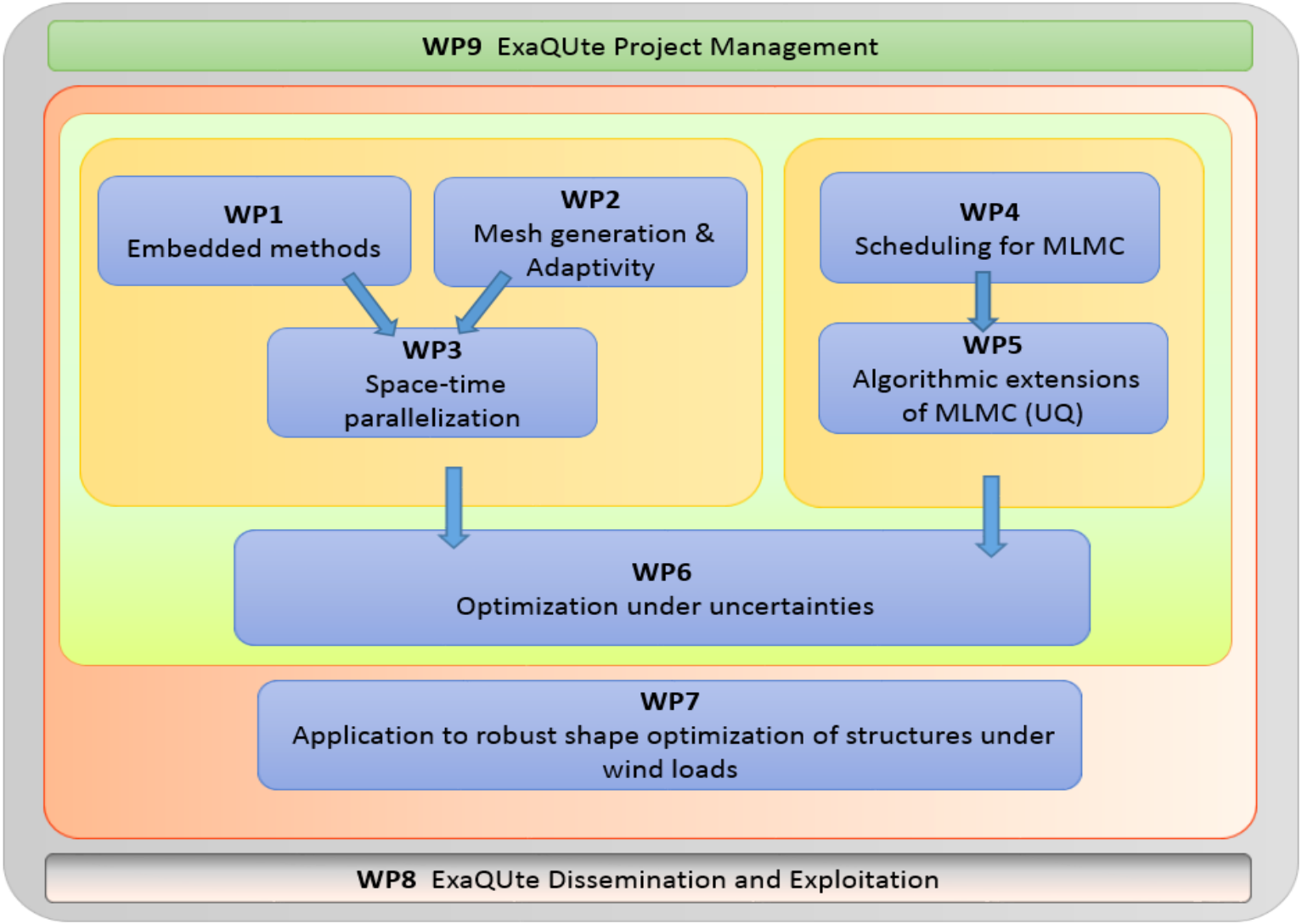
Instead of targeting extremely-large simulations

we target **many smaller problems** of engineering interest

by **considering also uncertainties.**

- possible to occupy any size of HPC systems −→ **Target full exascale systems as Proof Of Concept**
- but also opens a door for non-urgent computing (use when resources are available) cloud/grid computing in the future?

THE WORKFLOW



ExaQUte

INVOLVED TECHNOLOGIES

Through the combination of the following technologies

- Adjoint based methods for optimization (AO)
- Multilevel Montecarlo (MLMC) for UQ
- Embedded methods to avoid body-fited mesh generation
- Space-time adaptive mesh refinement (AMR) to capture missing multiscale features
- High performance computing (HPC) platforms
- Dynamic task scheduling (DTS) to efficiently exploit them

Application

ExaQUte

TARGET APPLICATIONS

Multiscale: The problem presents different length scales and different time scales which can be orders of magnitude different.

Multiphysics: The underliying physics may involve different materials (fluids, solids) and regimes.

Geometrically complex domains: “Beyond the cube”.

DRIVING APPLICATION: Robust optimization of structures subjected to wind action

ExaQUte

ASPIDE PROJECT

Extreme Data is an incarnation of Big Data concept distinguished by the massive amounts of data that must be queried, communicated and analyzed in (near) real-time by using a very large number of memory/storage elements and Exascale computing systems. Immediate examples are the scientific data produced at a rate of hundreds of gigabits-per-second that must be stored, filtered and analyzed, the millions of images per day that must be mined (analyzed) in parallel, the one billion of social data posts queried in real-time on an in-memory components database. Traditional disks or commercial storage cannot handle nowadays the extreme scale of such application data. Following the need of improvement of current concepts and technologies, ASPIDE's activities focus on data-intensive applications running on systems composed of up to millions of computing elements (Exascale systems). Practical results will include the methodology and software prototypes that will be designed and used to implement Exascale applications. The ASPIDE project is contributing with the definition of a new programming paradigms, APIs, runtime tools and methodologies for expressing data-intensive tasks on Exascale systems, which can pave the way for the exploitation of massive parallelism over a simplified model of the system architecture, promoting high performance and efficiency, and offering powerful operations and mechanisms for processing extreme data sources at high speed and real-time.

ASPIDE proposed a new programming model for data intensive applications based on a MPI+X model, adapted for distributed memory systems at large scale. In this project, we propose a new task-based back-end for distributed environments employing different communication channels (queues) for moving tasks and data among the different nodes comprising a given distributed architecture. The ASPIDE monitoring system exposes and associates a vast set of collected metrics with a main goal to expose potential application bottlenecks, which execute in Exascale systems. The key insight behind such an approach is that the source of a bottleneck in a data-intensive applications is often not place where it is detected (i.e. where the data is processed with a high communication or thrashing overhead), but where it is allocated. This project approaches the implementation and testing processes of a distributed ad-hoc in-memory storage system. The application will follow a client- server design model where the client itself will be responsible for the server entities deployment. There will be two manners available to accomplish the previous task: a lightweight deployment constrained to application's nodes and a heavier one considering offshore nodes. The application itself will be focusing in data locality exploitation alongside the implementation of multiple I/O patterns providing numerous data distribution policies. Besides, the storage system will follow a key-value structure mapped to a get-set interface trying to relax the POSIX standard interface constraints, such as unavailability for non-contiguous read and write operations.