



FP7 Support Action - European Exascale Software Initiative

DG Information Society and the unit e-Infrastructures



Addressing the Challenge of Exascale

European Exascale Software Initiative EESI

Towards Exascale roadmap implementation

EESI2 – Recommendations

DATA CENTRIC APPROACHES

Philippe Ricoux (TOTAL, EESI2 Coordinator)



Extreme computing and Extreme Data should be tackled simultaneously.

Extract of EESI2 Vision:

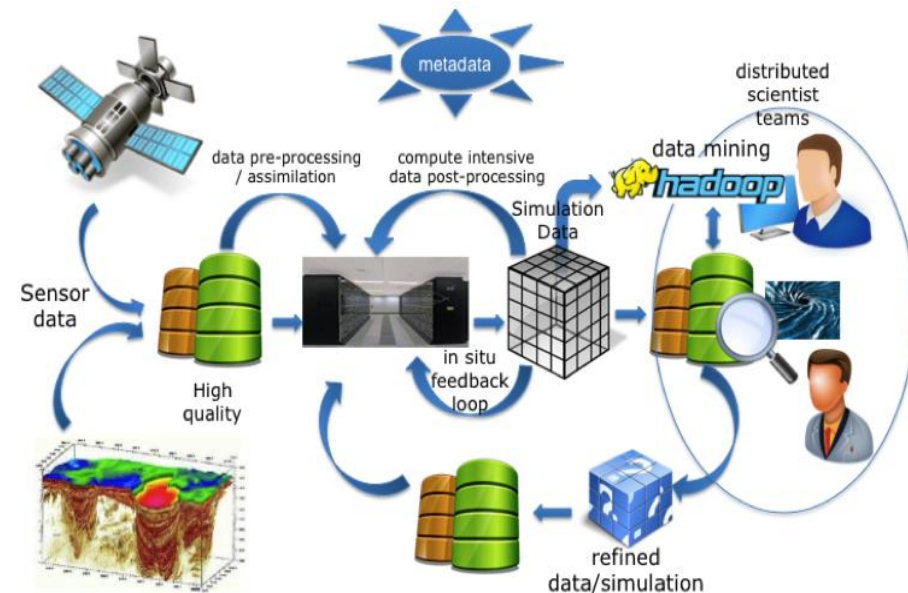
- At Exascale, Extreme Computing and Extreme Data (or Big Data) are intrinsically linked since supercomputers become mandatory to analyze efficiently huge flows of data generated by large scale instruments or by massive complex simulations.
- Exascale applications will be efficient only through developments by multidisciplinary teams, optimizing the interactions between
 - **architecture** (nodes, cores, **memories, interconnect, power, resilience ...**),
 - **algorithms** (programming, ultra scalable numerical methods, **asynchrony, fault tolerance ...**),
 - and **applications** (discretization of problems, engineering tools, **data processing ...**).

- At Exascale, the key factor will be power consumption and the **high power cost of the data movement**. Energy cost leads system operators to move from "charge by core-hours" to "charge by kilowatt-hours"
- Efficiency at Exascale level requires **breaking with the traditional scientific workflow** where simulation data are stored on disk for later analysis. This **disruption** comes in sync with **new memory technologies, new photonic networks** as well as the increasing cost of transistors.
- Holistic approaches considering **all data cycles** from sensors capture to **visualization, encompassing simulation, code coupling, in-situ, pre and post analysis** can guarantee that no bottlenecks are introduced in the scientific discovery process.
- In particular, it is strongly wished that **new systems simplify human-in-the-loop workflows**.

Motivations : foster synergies between data-intensive science and extreme-scale computing, **understanding the full cycle of data and integrating a coherent approach of uncertainties**

Proposal :

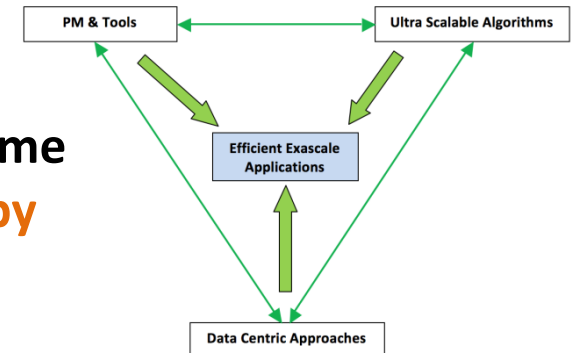
- Bridge the gap between scientists trained in both extreme-scale and data-intensive computing
- Data/compute research topics:
 - ✓ data transformational algorithms (extreme concurrency, asynchronous parallel data move, ...)
 - ✓ disruptive I/O and scalable multi-tiered data storage and parallel data management system
 - ✓ data analytics algorithms and techniques
- End-to-end simulation cycle:
 - ✓ simplify human-in-the-loop workflows,
 - ✓ libraries of scalable data analytics and data mining algorithms,
 - ✓ full-service data lifecycle management systems, in-situ vs.ex-situ data processing
 - ✓ **Need multidisciplinary teams**



Data Centric Approaches pillar

A vision “Software for Data Centric Approaches to Extreme Computing”, **In Situ Uncertainty Quantification Guided by Analytics**

- Uncertainty Quantification in massively parallel codes
- Towards flexible and efficient Exascale software couplers (direct or not, exchange of big data)
- In situ data processing: Toward real time Extreme Data Processing and better science through I/O avoidance in High-Performance Computing systems
- Declarative processing frameworks for big data analytics, extreme data fusion
- Identification of turbulent flow features from massively parallel Exaflops and Exabytes simulations



THANK YOU FOR ATTENTION