2nd ASTERICS-OBELICS Workshop

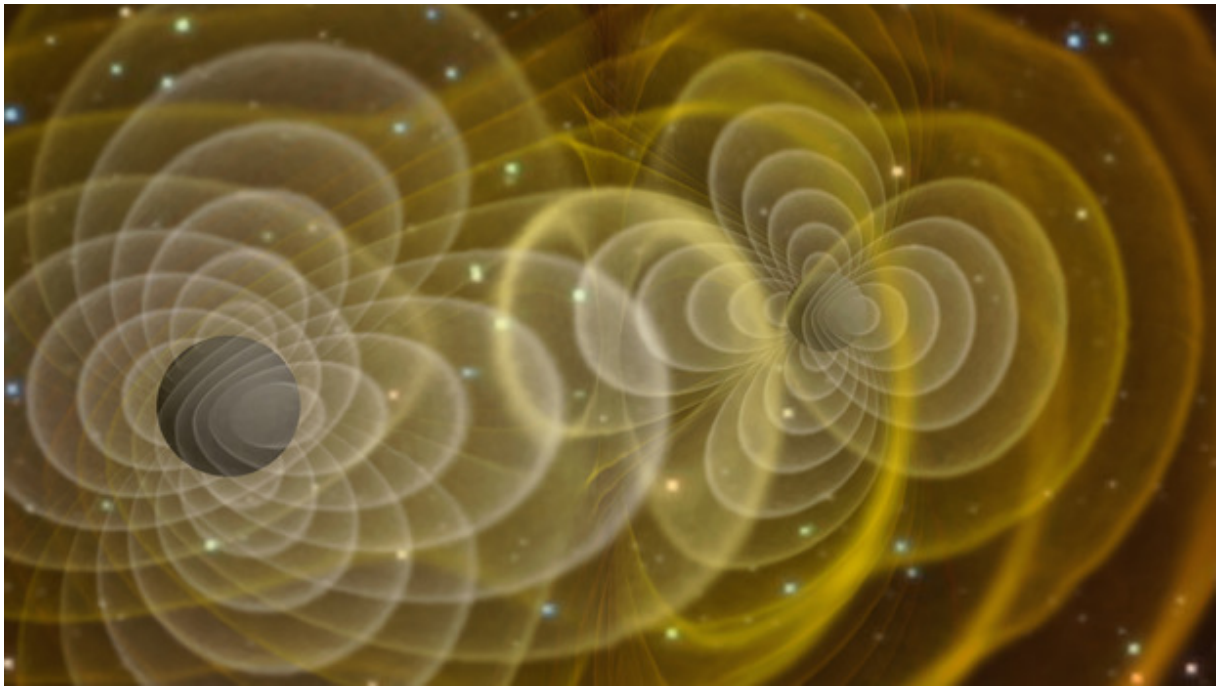
16-19 October 2017, Barcelona, Spain.

E4 Projects, Collaborations and Expertise

Daniele Gregori Ph.D.

H2020-Astronomy ESFRI and Research Infrastructure Cluster
(Grant Agreement number: 653477).

TABLE OF CONTENTS

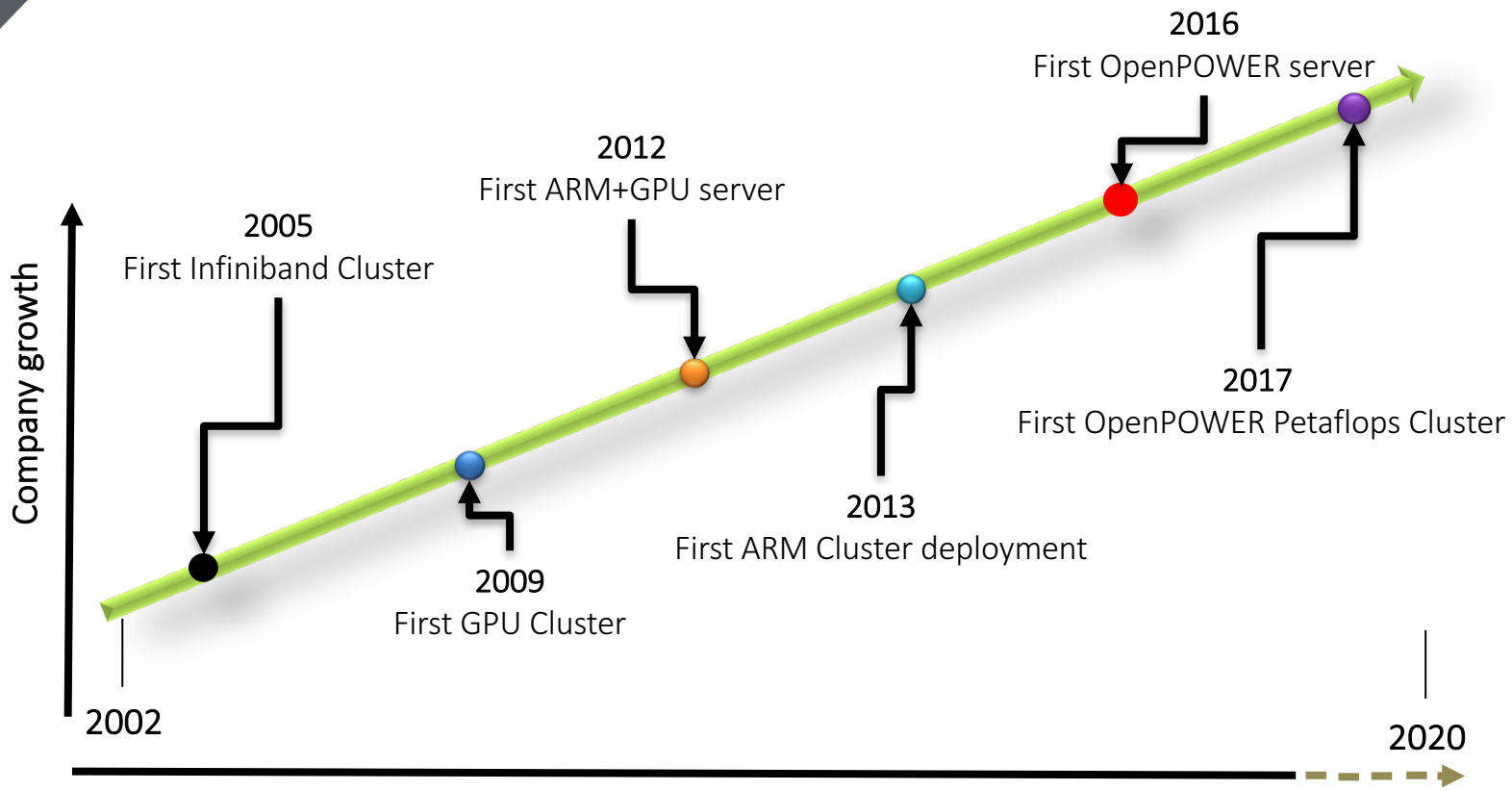


THE COMPANY

Since 2002, E4 Computer Engineering has been innovating and actively encouraging the adoption of new computing and storage technologies. Because new ideas are so important, we invest heavily in research and hence in our future. Thanks to our comprehensive range of hardware, software and services, we are able to offer our customers complete solutions for their most demanding workloads on: HPC, Big-Data, AI, Deep Learning, Data Analytics, Cognitive Computing and for any challenging Storage and Computing requirements.

E4. When Performance Matters.

COMPANY MILESTONES



WHAT WE DO

E4 COMPANY PILLARS

- Hardware Products
- Extreme Computing Solutions
- R&D Prototyping
- Services and Support

Collaboration with Universities and Research Center are the key element to attend to EU Projects

Cooperations:

MEMBERSHIPS

- **OpenPOWER Foundation**
- **OpenPower for Physical Science WG**
- **ETP4HPC European Technology Platform for HPC**
- **HPC Advisory Council**
- **Open Compute Project**
- **MAX Center of Excellence**
  http://www.max-centre.eu/

Artificial Intelligence for Astrophysics

AlfA Project

The final purpose of the collaboration is to assign a Co-funded PhD scholarship by H2020 ASTERICS/OBELICS, INAF-Osservatorio Astronomico di Roma and E4 Computing Technologies S.p.A.. This scholarship will have to be assigned within the XXXIII cycle of Italian Ph.D. to issued by the University of Rome “Tor Vergata” on June 2017 . *Such a project will be dedicated to the development of a new data analysis technique for IACTs, based on machine-learning and using Deep Neural Networks (DNNs), for analyzing images and classify within a software-hardware integrated system adopting new hardware architectures.*

E4 Alfa Responsability

- E4 Computer Engineering S.p.A. grants access to the internal data centre and to a proper hardware infrastructure to realize a neural network, big data and deep learning system to recognize CTA experiment images.
- E4 Computer Engineering S.p.A. commits to train PhD student to system engineering activities.

R&D LAB

- 30 m^2
- temperature 27/30°c
- 6 x Rack 19'
- 4 x Chiller 22 kw
- Active Power available ~100 kw

- Hardware Management via OpenDCIM open source
  Remote access available on demand

EU Project Involved

Proposal Submitted to:

- FETHPC-01-2016: Co-design of HPC Systems and applications https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/fethpc-01-2016.html
- FETHPC-02-2017: Transition to Exascale Computing https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/fethpc-02-2017.html
- FETHPC-03-2016: Exascale HPC ecosystem development https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/fethpc-03-2017.html
- ICT-42-2017: Framework Partnership Agreement in European low-power microprocessor technologies http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/ict-42-2017.html

EU Project Involved

- **EU-funded projects**
- **Awarded PRACE-3IP PCP Pre-Commercial Procurement concerning R&D services on "Whole System Design for Energy Efficient HPC"**
- **D.A.V.I.D.E. (Development of an Added Value Infrastructure Designed in Europe) #299 in TOP500, #14 in GREEN500**

D.A.V.I.D.E. SUPERCOMPUTER

COMPUTE

(Development of an Added Value Infrastructure Designed in Europe)

PRACE Awards Third and Final Phase of Pre-Commercial Procurement (PCP)

After successfully completing phase II, during phase III, E4 proposed an innovative design that makes avail of the most advanced technologies, to produce a leading edge HPC cluster showing higher performance, reduced power consumption and ease of use.

PCP PHASE III – D.A.V.I.D.E. SUPERCOMPUTER
(Development of an Added Value Infrastructure Designed in Europe)

COMPUTE NODE:

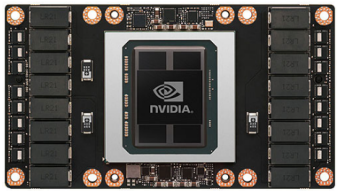
- Derived from the IBM® POWER8 System S822LC (codename Minsky).
- 2 IBM POWER8 NVlink and 4 NVIDIA Tesla P100 HSXM2 with the intra node communication layout optimized for best performance.
- While the original design of the Minsky server is air cooled, its implementation for DAVIDE uses direct liquid cooling for CPUs and GPUs.
- Each compute node has a peak performance of 22 TFLOPS and an power consumption of less than 2kW.

Total number of nodes	45 (compute) + 2 (login)
Form factor	2U
SoC	2xPOWER8 NVlink
GPU	4xNVIDIA Tesla P100 HSMX2
Network	2xIB EDR, 1x 1GbE
Cooling	SoC and GPU with direct hot water
Max performance (node)	22 TFlops
Storage	1xSSD SATA, 1x NVMe
Power	DC power distribution

PCP PHASE III – D.A.V.I.D.E. SUPERCOMPUTER
(Development of an Added Value Infrastructure Designed in Europe)

ACCELERATOR

- NVIDIA Tesla P100 (HSMX2)
- NVIDIA Tesla P100 was built to deliver performance for the most demanding compute applications, providing:
  - 5.3 TFLOPS of double precision floating point (FP64) performance
  - 10.6 TFLOPS of single precision (FP32) performance
  - 21.2 TFLOPS of half-precision (FP16) performance



NVLINK BUS

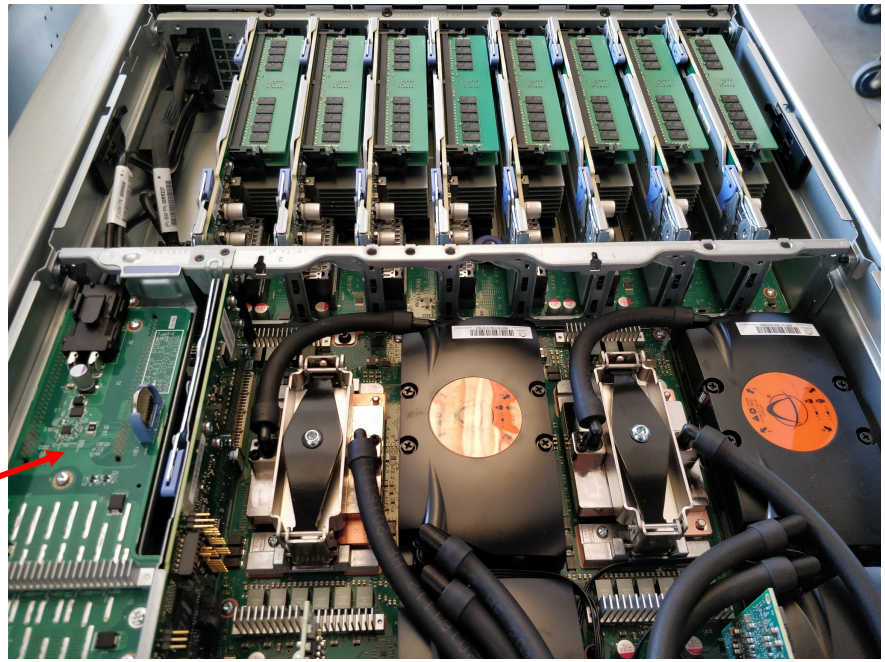
- NVIDIA's new High-Speed Signaling interconnect (NVHS).
- NVHS transmits data over a differential pair running at up to 20 Gb/sec.
- Eight of these differential connections form a Sub-Link that sends data in one direction, and two sub-links—one for each direction—form a Link that connects two processors (GPU-to-GPU or GPU-to-CPU).
- A single Link supports up to 40 GB/sec of bidirectional bandwidth between the endpoints.
- The NVLink implementation in NVIDIA Tesla P100 supports up to four links, enabling ganged configurations with aggregate maximum bidirectional bandwidth of 160 GB/sec.

PCP PHASE III – D.A.V.I.D.E. SUPERCOMPUTER
(Development of an Added Value Infrastructure Designed in Europe)

OPEN RACK LIQUID COOLED

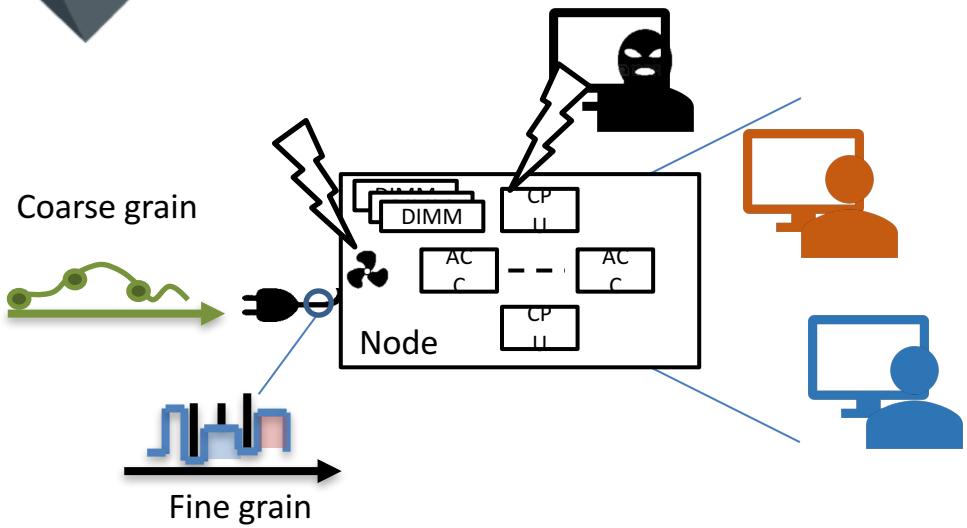
- Direct hot-water cooling (35-40 °C) for the CPUs and GPUs.
- Capable to extract about 80% of the heat produced by the compute nodes.
- Extremely flexible and requiring minor modifications of the infrastructure.
- Each rack has an independent liquid-liquid or liquid/air heat exchanger unit with redundant pumps.
- The compute nodes are connected to the heat exchanger through pipes and a side bar for water distribution.

Total number (racks)	3
Form factor	2U
Cooling Capacity	40 kW
Heat exchanger	Liquid-liquid, redundant pumps



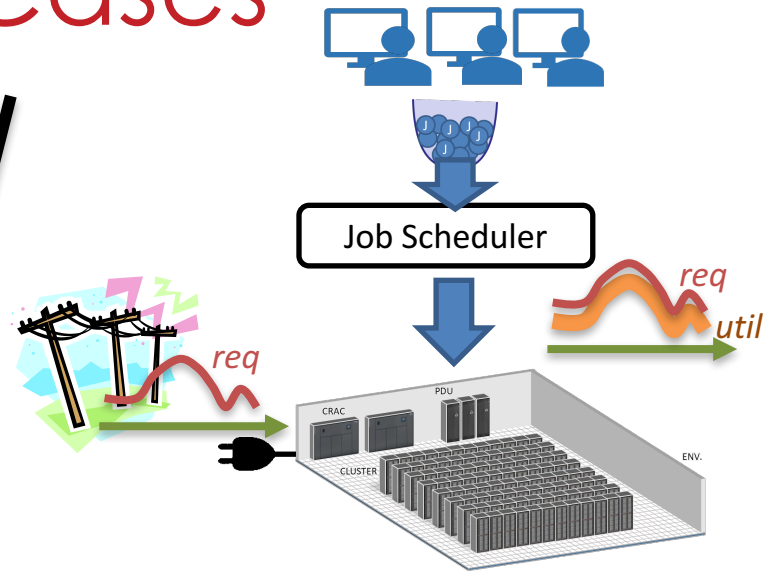
E4 Power Performance Black Box

Target Use Cases



Fine Grain Power and Performance Measurements:

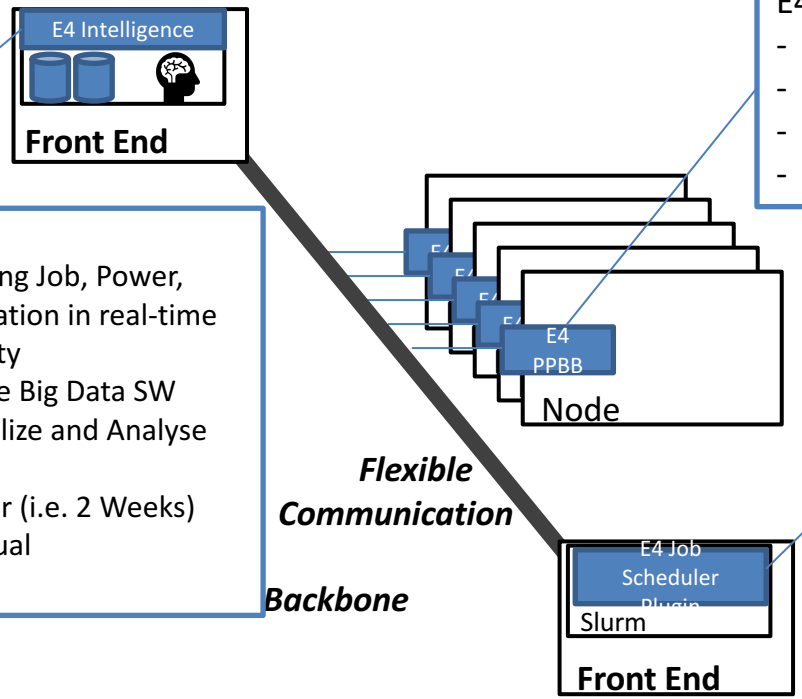
- Verify and classify node performance
  - In spec / out of spec behavior
  - Aging and wear out
- Predictive maintenance
- Per user - Energy / Performance – accounting



System Power Capping

- New Installations, Grid SLA, Power Shortage, Natural Disasters
- Ensures operating power below a maximum power consumption level

E4 KEY ENABLING TECHNOLOGIES



E4 PPBB (Power and Performance «Black Box»)
- Out-of-band => Zero overhead
- Fine grain => down to ms scale
- IoT communication technology => scalable
- Time synchronous (NTP, PTP)

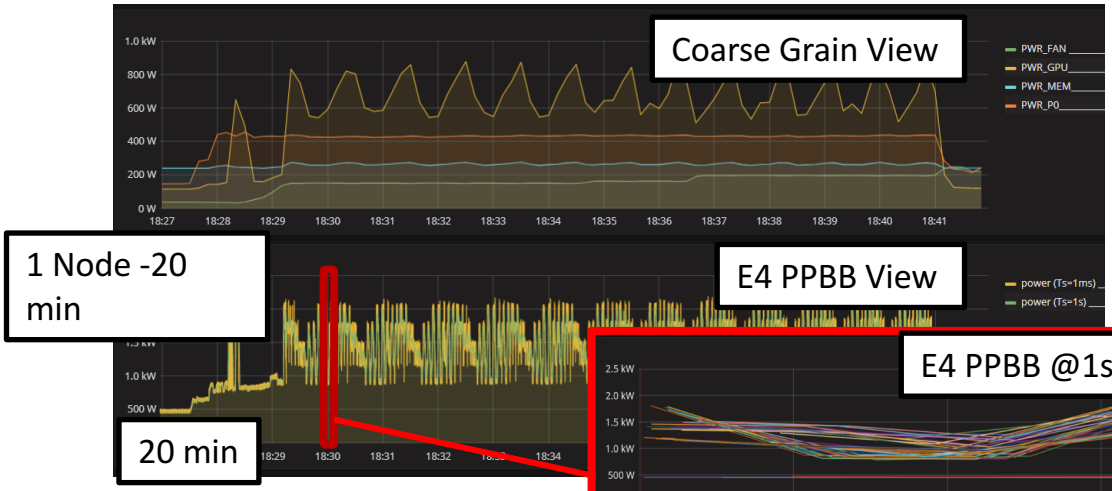
E4 Intelligence
- Capable of aggregating Job, Power, Performance information in real-time and at fine-granularity
- Based on opensource Big Data SW
- Store, Process, Visualize and Analyse Monitored Data
  - Historical Buffer (i.e. 2 Weeks)
  - Per Job Perpetual
- Web frontend

E4 Job Scheduler Plugin
- Monitors Job submission and scheduling
- Estimates Job Power Consumption
- Controls Power Consumption

E4 Key Enabling Technologies

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

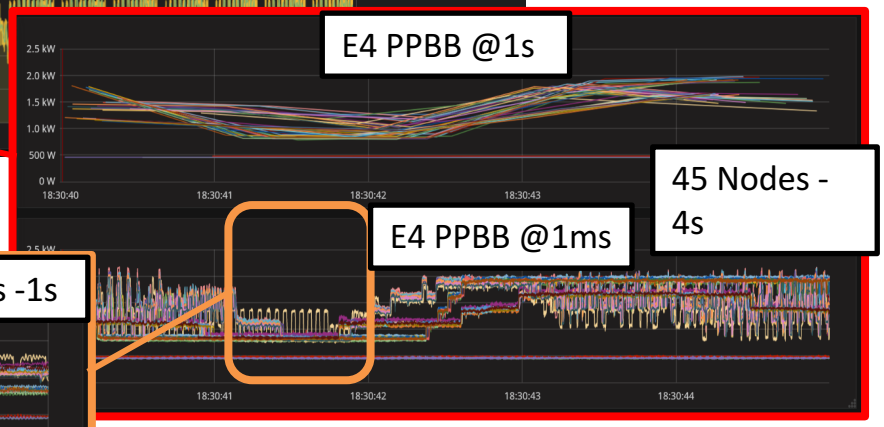
E4 PPBB Node



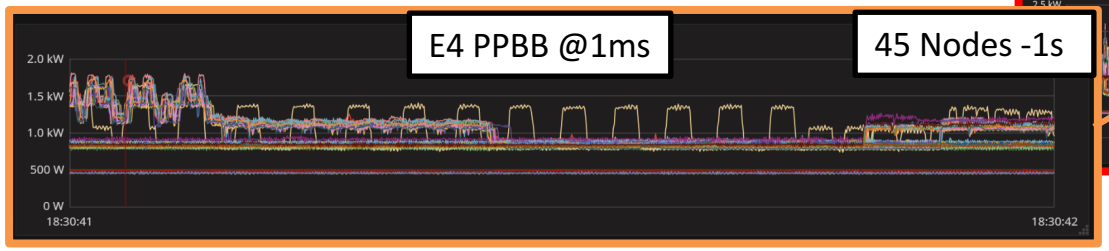
1 Node -20 min

E4 PPBB View

20 min

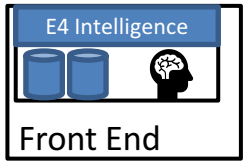


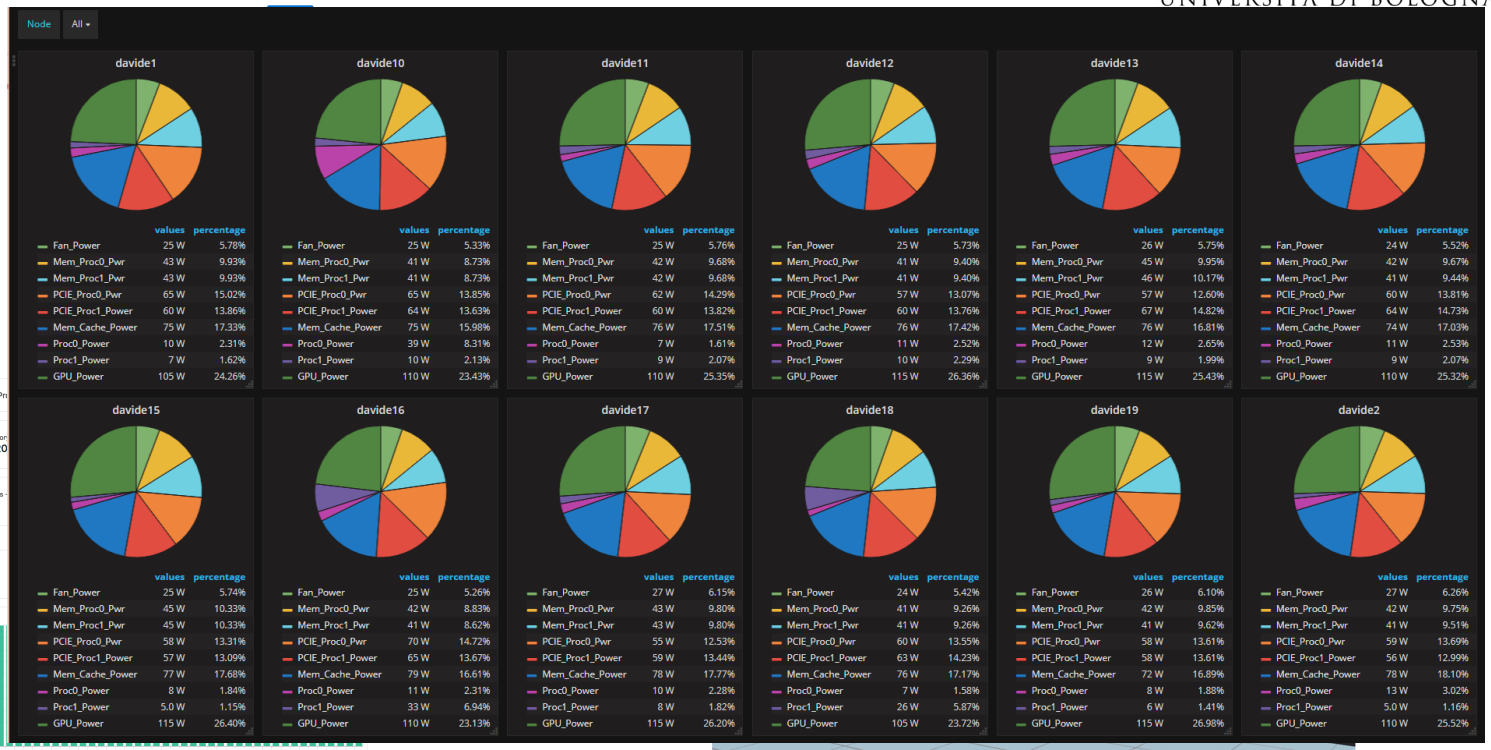
45 Nodes - 4s

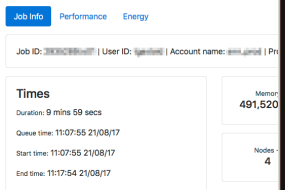


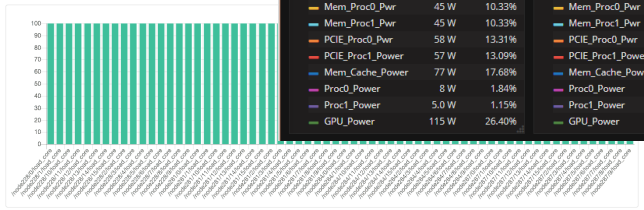
E4 Key Enabling Technologies

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



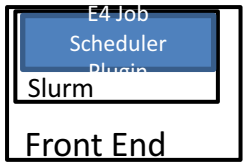




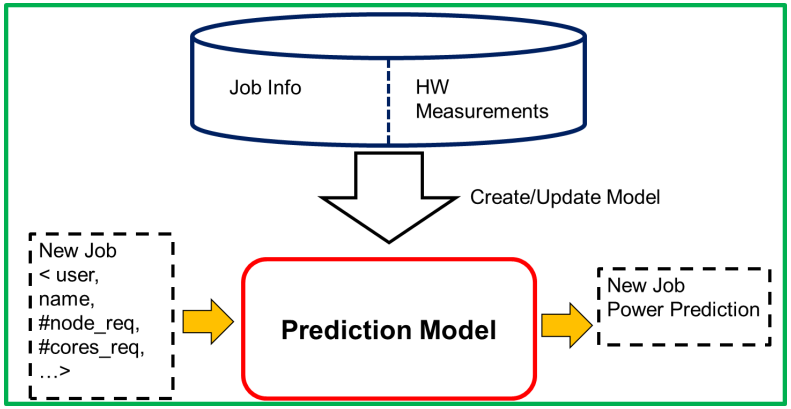


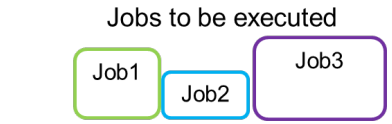
E4 Key Enabling Technologies

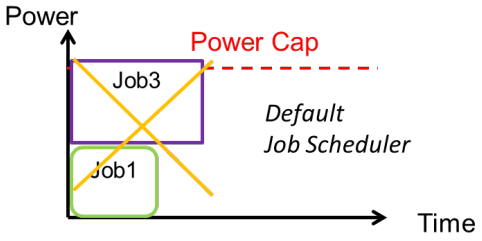
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

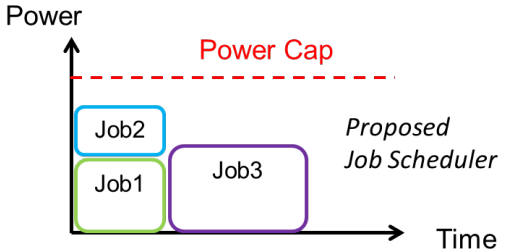


1. Machine Learning models to predict the power consumption of HPC applications
2. Slurm Custom Extensions to schedule jobs based on their power
3. Interacts with power management
   - Frequency scaling/RAPL-like mechanism









D.A.V.I.D.E Outreach

User-space APIs for Dynamic Power Management in Many-core ARMv8 Computing Nodes

Daniele Bortolotti
EEES - DEI
University of Bologna
Bologna, Italy
daniele.bortolotti@unibo.it

Simone Tinti, F
E4 Computer E
Scandiano,
simone.tinti@e4c
piero.altoe@e4c

***Abstract*—The push for energy-efficient and energy-proportional computing nodes, together with the increasing number of cores integrated in the same silicon die has lead to computing nodes with fine grained power management capabilities. To unleash the potential of this HW design a novel user-space power management APIs is needed to bring fine-grain power management in the hands of the programmer. In this work we present a novel programming mechanism for energy efficiency which is build around novel user-space power management APIs suitable to be embedded in user-space applications. We evaluated its timing and power saving performance on a novel computing node based on Cavium**

Design of an Energy Aware peta-flops Class High Performance Cluster Based on Power Architecture

Wissam Abu Ahmad[1], Andrea Bartolini[2,3], Francesco Beneventi[2], Luca Benini[2,3], Andrea Borghesi[2], Marco Cicala[1], Privato Forestieri[1], Cosimo Gianfreda[1], Daniele Gregori[1], Antonio Libri[3], Filippo Spiga[4,5], Simone Tinti[1]

[1] E4 Computer Engineering, Scandiano (RE), Italy.
[2] DISI, DEI, University of Bologna, Bologna, Italy.
[3] Department of Information Technology and Electrical Engineering, ETH, Zurich, Switzerland.
[4] Quantum ESPRESSO Foundation, UK
[5] University of Cambridge, Cambridge, UK

wissam.abuahmad@e4company.com, a.bartolini@unibo.it, francesco.beneventi@unibo.it, luca.benini@unibo.it, andrea.borghesi@unibo.it, marco.cicala@e4company.com, tino.forestieri@e4company.com, cosimo.gianfreda@e4company.com, daniele.gregori@e4company.com, a.libri@iis.ee.ethz.ch, filippo.spiga@quantum-espresso.org, simone.tinti@e4company.com

***Abstract*—In this paper we present D.A.V.I.D.E. (Development for an Added Value Infrastructure Designed in Europe), an innovative and energy efficient High Performance Computing cluster designed by E4 Computer Engineering for PRACE (Partnership for Advanced Computing in Europe). D.A.V.I.D.E. is built using best-in-class components (IBM's POWER8-NVLink CPUs, NVIDIA TESLA P100 GPUs, Mellanox InfiniBand EDR**

has caused an increment of the total power consumption. This was true till Tianhe-2 (the former most powerful supercomputer, 1st from 06/2013 to 11/2015 Top500 lists), where the IT power consumption reached the practical limit of 17.8 MW for 33.8 PFlops.The current most powerful supercomputer Taihu Light reaches 93 PFlops with a power envelope of only

AI@E4

OUR PARTNERS

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

NVIDIA

IBM

TIM

UNIVERSITÀ
DEGLI STUDI
FIRENZE

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Addfor

AXYON.AI

CINECA

energy way

ENGINSOFT

OROBIX
TOOLS WITH A MIND

AI@E4:
SELECTING THE BEST TECHNOLOGY AND SOLUTION

FLEXIBLE SOLUTION BY E4	IBM PowerAI E4 OP 206 Gold	NVIDIA DGX-1
PURPOSE Flexible Deep Learning solution for testing, development, benchmarks and early production	PURPOSE Fully integrated Deep Learning solution with hardware, software and development tools to run accelerated analytics applications	PURPOSE Fully integrated Deep Learning solution with hardware, software and development tools to run accelerated analytics applications
AI SOFTWARE Base Libraries, OpenSource Deep Learning framework	AI SOFTWARE Deep Learning framework (optimized)	AI SOFTWARE NV Docker, Deep Learning Framework (optimized), Monitoring software
NVLINK GPU – GPU	NVLINK GPU-GPU, GPU -CPU	NVLINK GPU – GPU
GPU From 1 to 8 NVIDIA® GPUs	GPU Up to 4 NVIDIA® Tesla® P100 (with NVIDIA® NVLink™)	GPU 8 NVIDIA® Tesla® P100 (with NVIDIA® NVLink™)
CPU Intel® Xeon® Processors	CPU IBM Power8™ Processor	CPU Intel® Xeon® Processors

SUCCESS STORY

Customer	UNICREDIT S.p.A.
Industry	Finance, Banking
Contact	Riccardo Prodam – Head of R&D

REQUIREMENTS
- Finance market simulation
- Risk assessment
- Transaction security
- Low latency trading

CHALLENGES Real time analysis and forecasting

SOLUTION NVIDIA DGX-1

APPLICATION Proprietary

KEY FACTORS Speed & Performance

BENEFITS
- Increased accuracy
- increased security
- Reduced transaction timing

Our Idea of HPC through E4 HPC Open Suite

Best Hardware:

- ✓ Low Failuere Rate
- ✓ Performed Benchmark in our lab
- ✓ Designe solution based on Custormer requirement

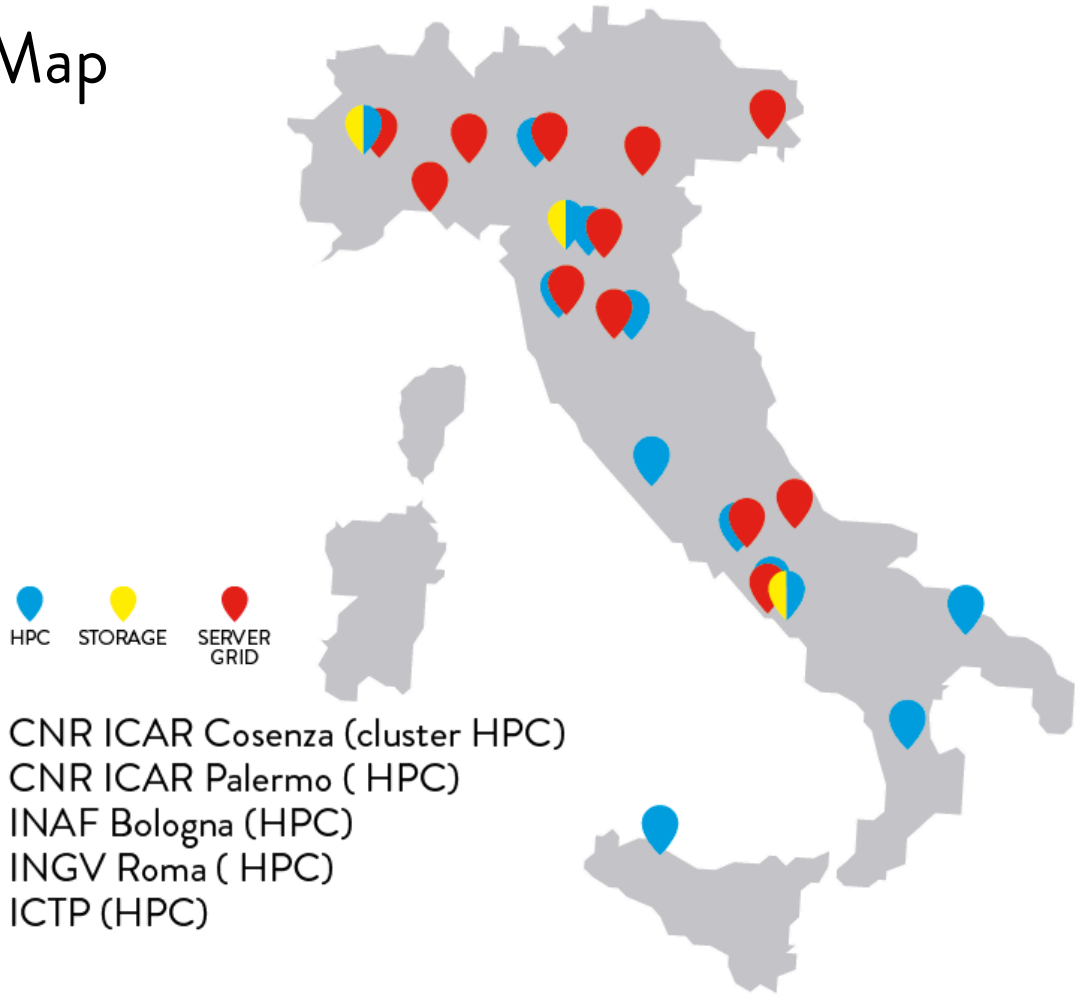
+

Best Skills: We are able to configure each Software Components, define the Modules Environment and customize the Cluster for a **Ready to Use Solution**

Case Studies Map

ITALIA

Università degli studi di Torino
Pirelli (Milano)
Scuola Normale Superiore di Pisa
Stazione Zoologica Anton Dohrn Napoli
Università di Modena e Reggio Emilia
Università di Bologna
E-GEOS (Roma)
INGV NAPOLI
PARTHENOPE NAPOLI
CNAF (Bologna)
INFN
Azienda Ospedaliera Perugia
Politecnico di Bari (cluster HPC)

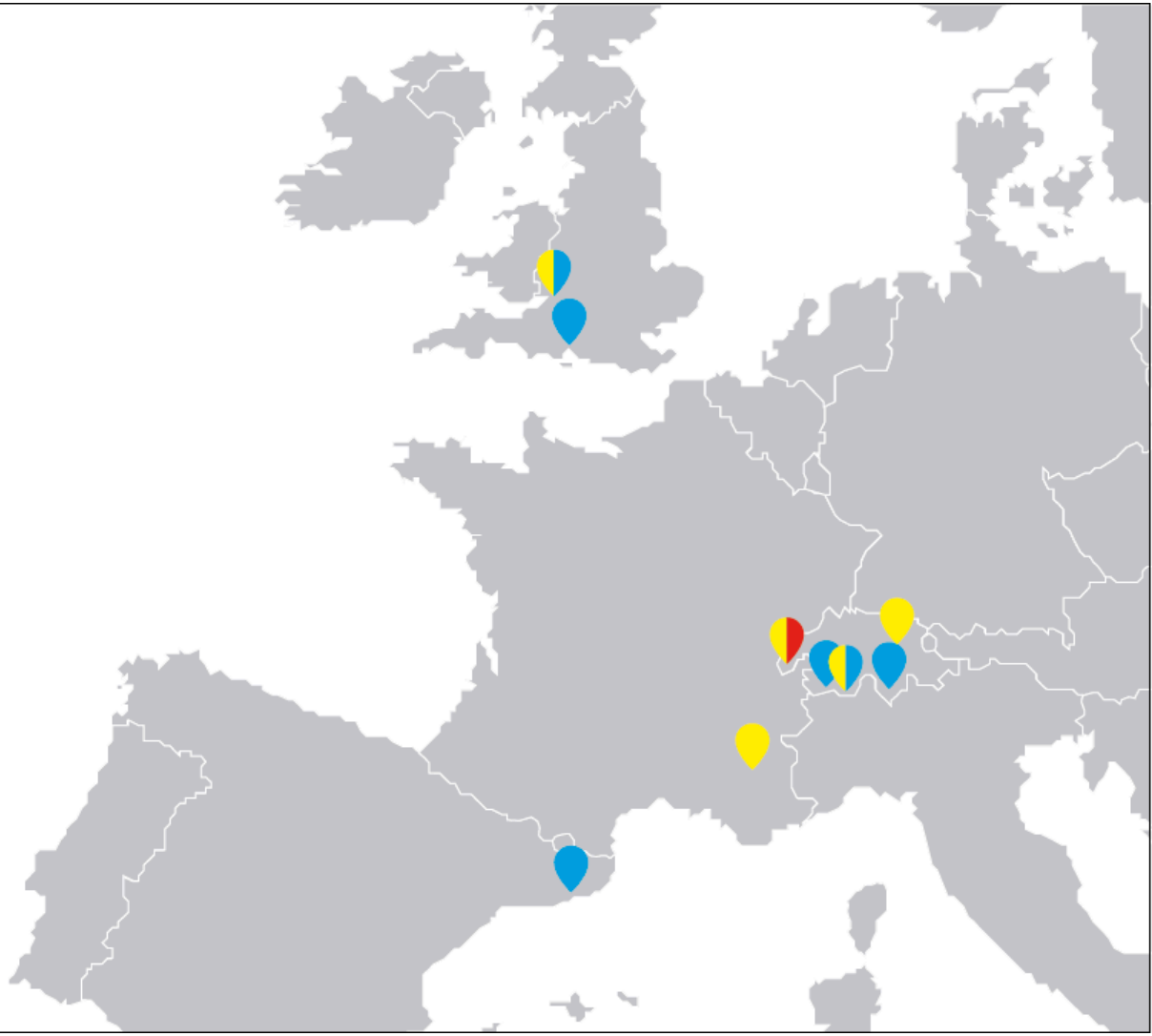


CNR ICAR Cosenza (cluster HPC)
CNR ICAR Palermo (HPC)
INAF Bologna (HPC)
INGV Roma (HPC)
ICTP (HPC)

E4 COMPUTER ENGINEERING

EUROPE

HYPERSTEM SA - Lugano (CH)
BSC - Barcelona (E)
EPFL - Losanna (CH)
ETH - Zurich (CH)
Swiss Insitute of Bioinformatics (CH)
CERN - Geneva (CH)
University of Southampton (UK)
British Aerospace - Bristol (UK)
ESRF - Grenoble (FR)

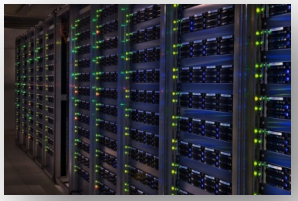


HPC ARM Success Story

Customer	BSC PEDRAFORCA CLUSTER
Industry	Supercomputing National Centre

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

REQUIREMENTS	Custom solution based on low power CPU and GPU accelerators
CHALLENGES	Creating an unique prototype with mobile SoC connected to high-end computing GPUs
SOLUTION	78 compute nodes equipped with Tegra 3 SoC, Nvidia K20, Mellanox Infiniband QDR
APPLICATION	GPU boosting
KEY FACTORS	Low power SoC Prototyping ability
BENEFITS	Accelerated computing at minimum power footprint First worldwide ARM+GPU prototype Disruptive innovation



REQUIREMENTS	High density computational nodes Big data storage
CHALLENGES	Delivering standard commodity hardware Providing high performances combined with energy efficiency Ensuring very low failure rate
SOLUTION	5.600+ dual socket mainboards (61.000+ cores) 35.000+ enterprise class hard disks (100PB Storage)
APPLICATION	Grid Computing

CHALLENGES	Delivering standard commodity hardware Providing combo of high performances & energy efficiency Ensuring very low failure rate
SOLUTION	12PB high performance storage (CNAF) 5PB direct attached storage (Alice – CMS) 4.500 server dual socket (~ 40k computing cores) Several GPU systems 4h intervention times
APPLICATION	Grid Computing

Email contacts

info@e4company.com

sales@e4company.com

luciano.fontana@e4company.com

daniele.gregori@e4company.com

E4 Computer Engineering SpA

Via Martiri della Libertà, 66
42019 Scandiano (RE) - Italy
Tel. 0039 0522 991811

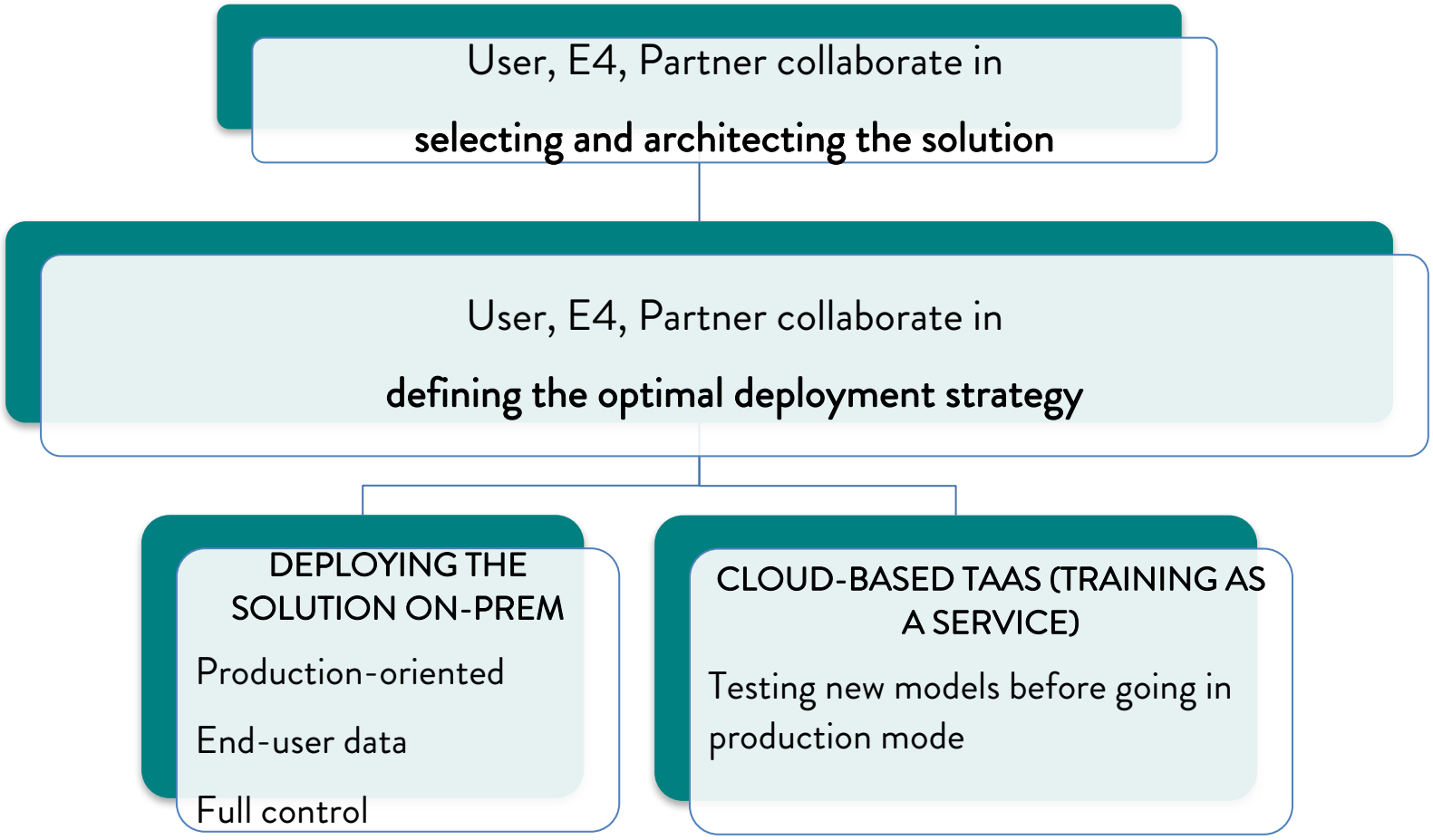
Text for acknowledgement Slide

Acknowledgement

- H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

BACKUP

AI@E4: ENGAGEMENT MODEL



AI@E4: DEPLOYMENT STRATEGY