ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Challenges in Holistic Monitoring and Data Integration

Andrea Bartolini, Prof
University of Bologna – DEI, Italy

EU H2020 FETHPC
project ANTAREX
(g.a. 671623)

EU FP7 ERC Project
MULTITHERMAN
(g.a.291125)

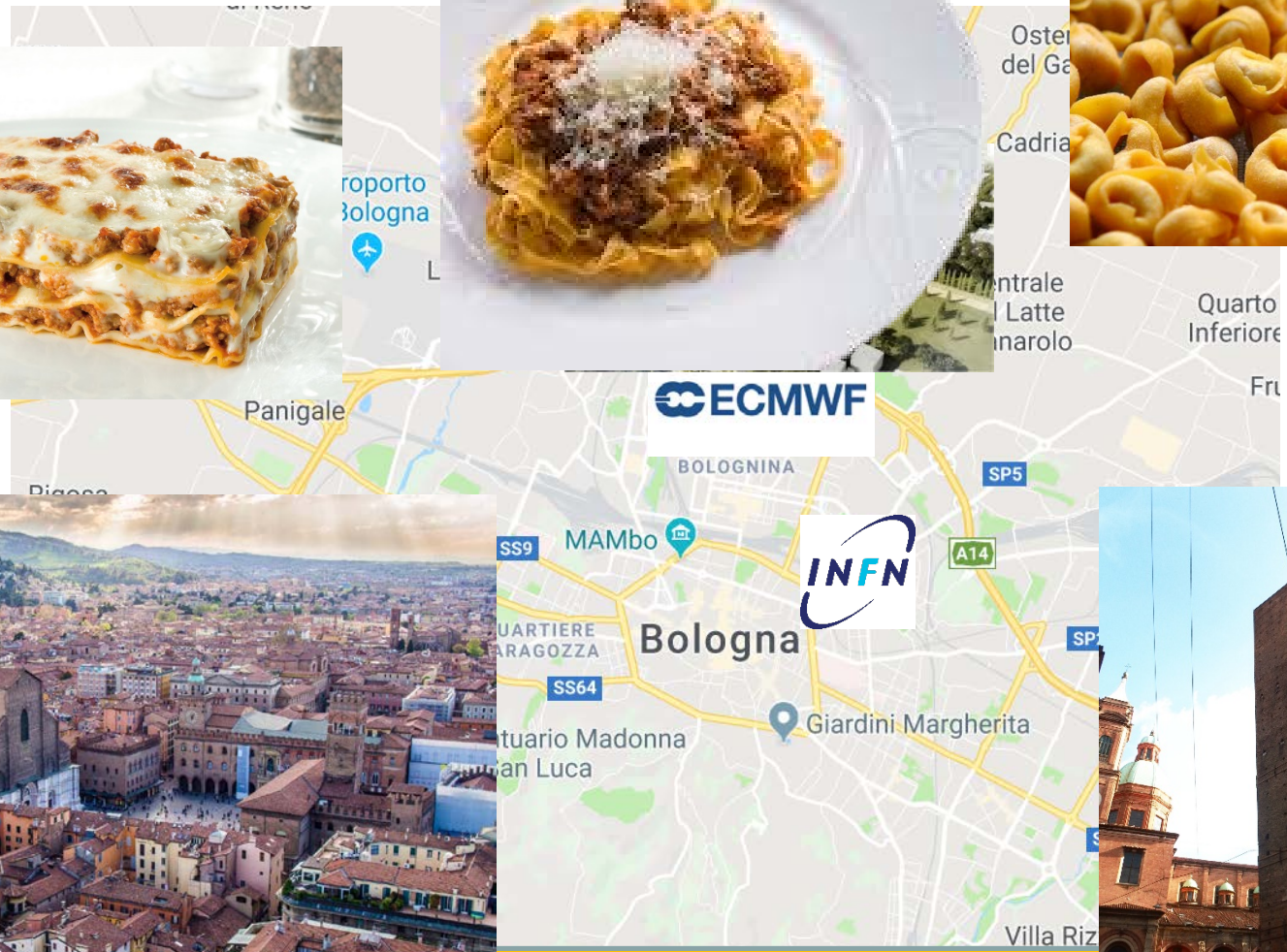
The University of Bologna

MULTI-CAMPUS UNIVERSITY

Alma Mater Studiorum
Università di Bologna
is a multi-campus
university based in Bologna,
Cesena, Forlì, Ravenna, and
Rimini.

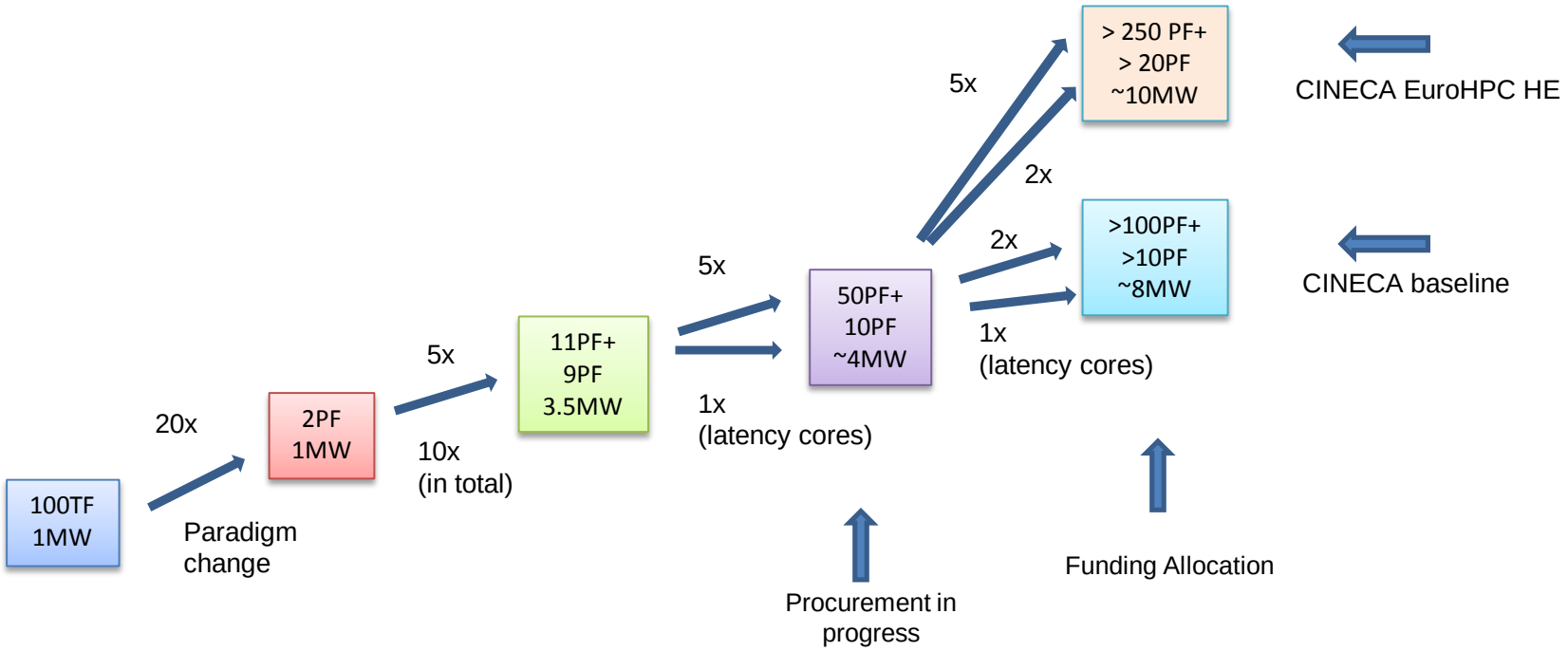
University Statute, Constituent
Principles, Art. 1 para. 2

Bologna the city of …



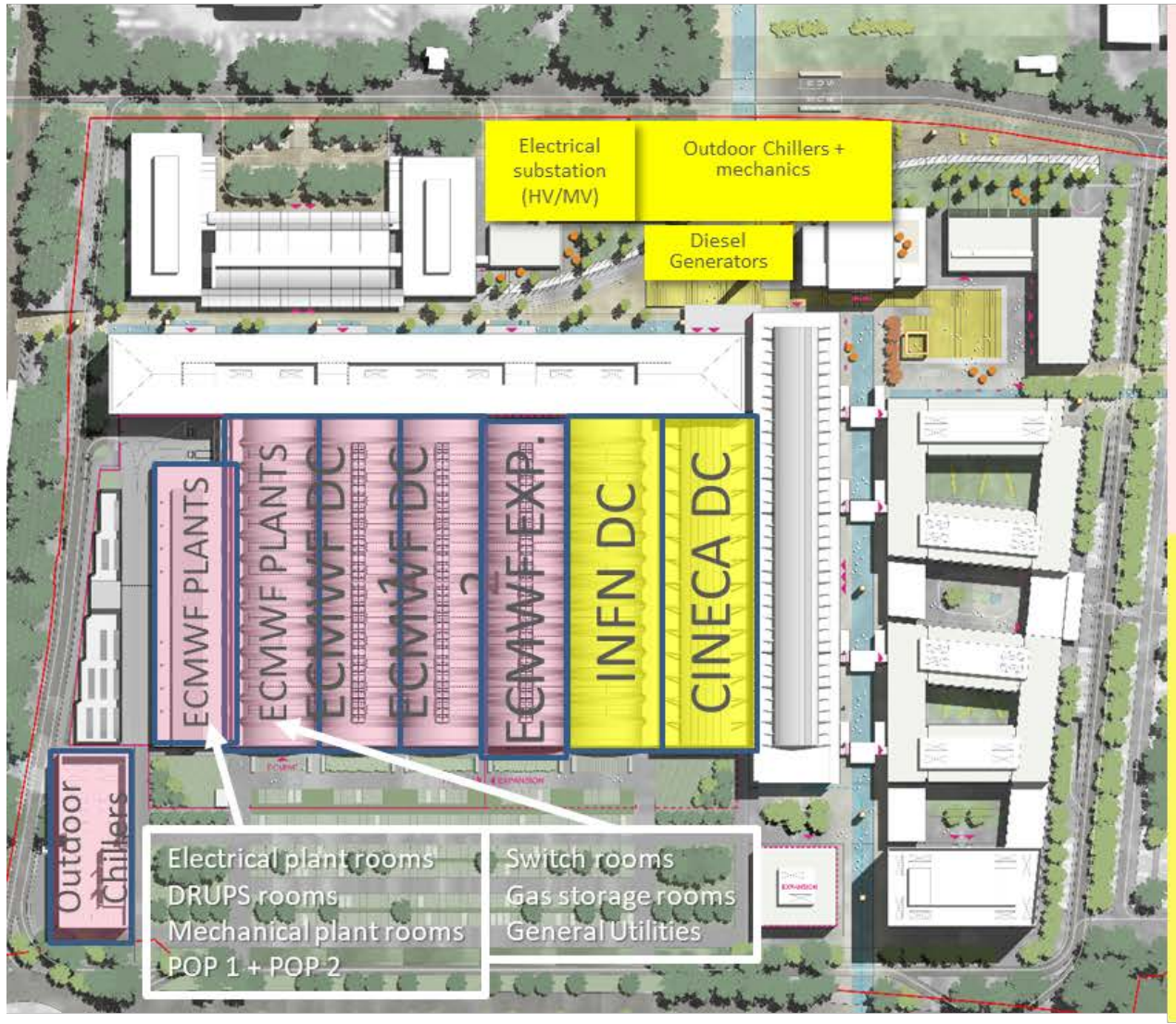
.Datacentres.

Cineca Roadmap



2009	2012	2016	2019	2021	2025
IBM SP6 Power6	Fermi IBM BGQ PowerA2	Marconi Lenovo Xeon+KNL	Marconi PPI4HPC ICEI - PPIHBP	Latency core consolidation + new generation scale out	Hybrid technology Digital computing and quantum computing

The data centers at the Science Park



ECMWF DC main characteristics

- 2 power line up to 10 MW (one bck up of the other)
- Expansion to 20 MW
- Photovoltaic cells on the roofs (500 MWh/year)
- Redundancy N+1 (mechanics and electrical)
- 5 x 2 MW DRUPS
- Cooling
  - 4 dry coolers (1850 kW each)
  - 4 groundwater welles
  - 5 refrigerator units (1400 kW each)

INFN – CINECA DC main characteristics

- up to 20 MW (one bck up of the other)
- Possible use of Combined Heat and Power Fuel Cells Technology
- Redundancy strategy under study
- Cooling, still under study
  - dry coolers
  - groundwater welles
  - refrigerator units
- PUE < 1.2 – 1.3

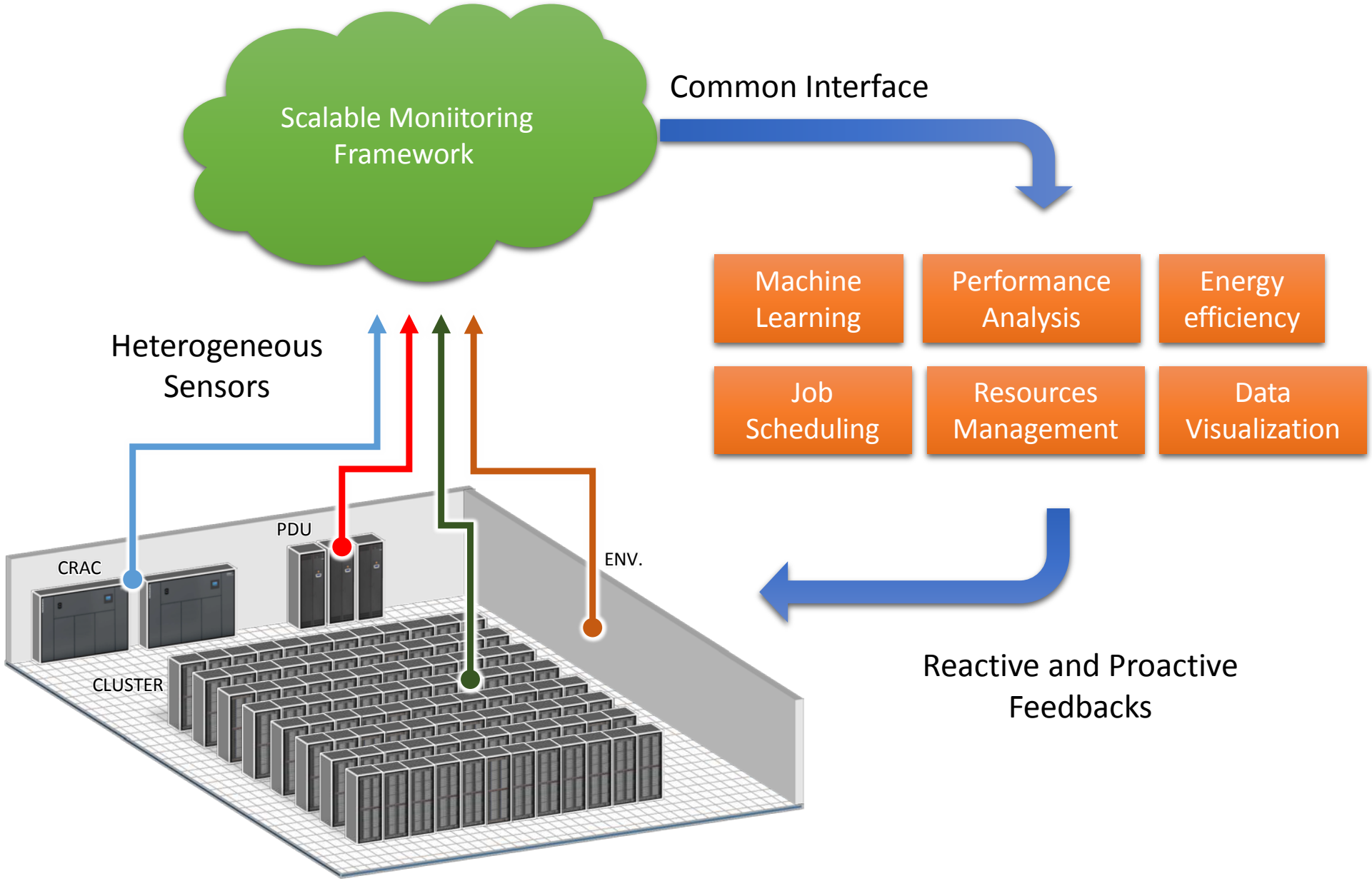
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Outline

- **Datacenter Automation**
- Holistic Monitoring @ CINECA with ExaMon
- Holistic Monitoring + Stream and Edge processing
- Conclusions & Future Works

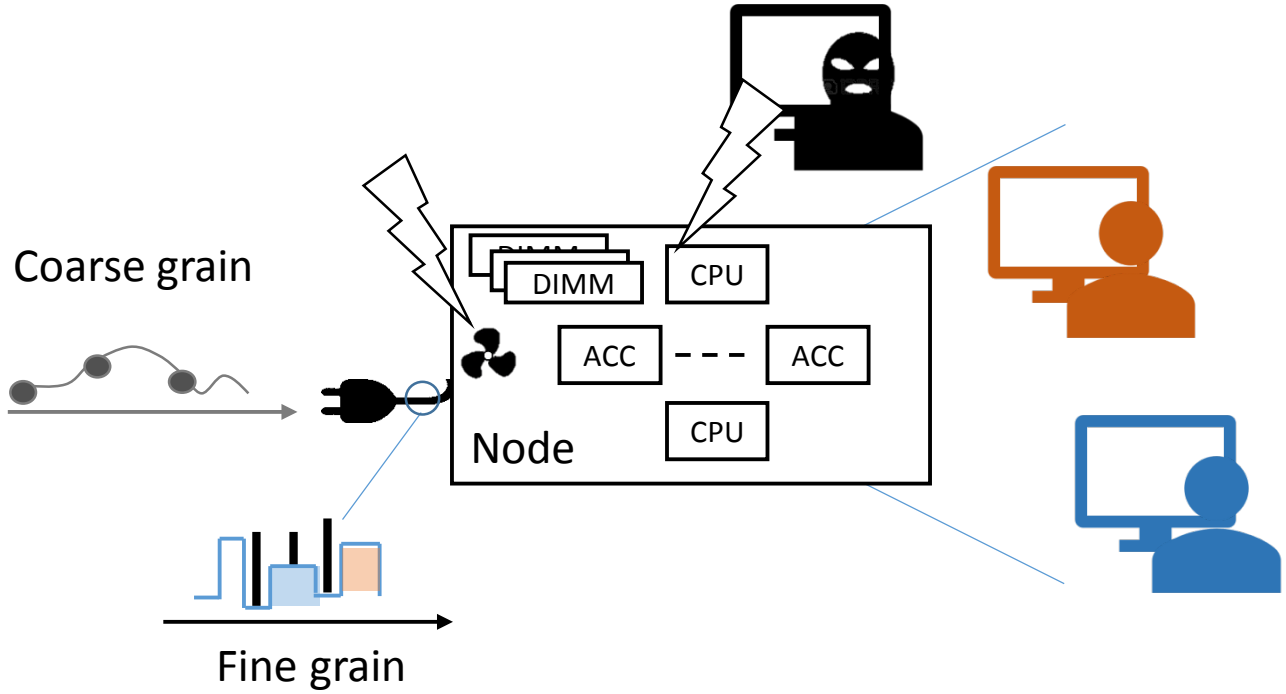
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

A New Trend: Datacentre Automation



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

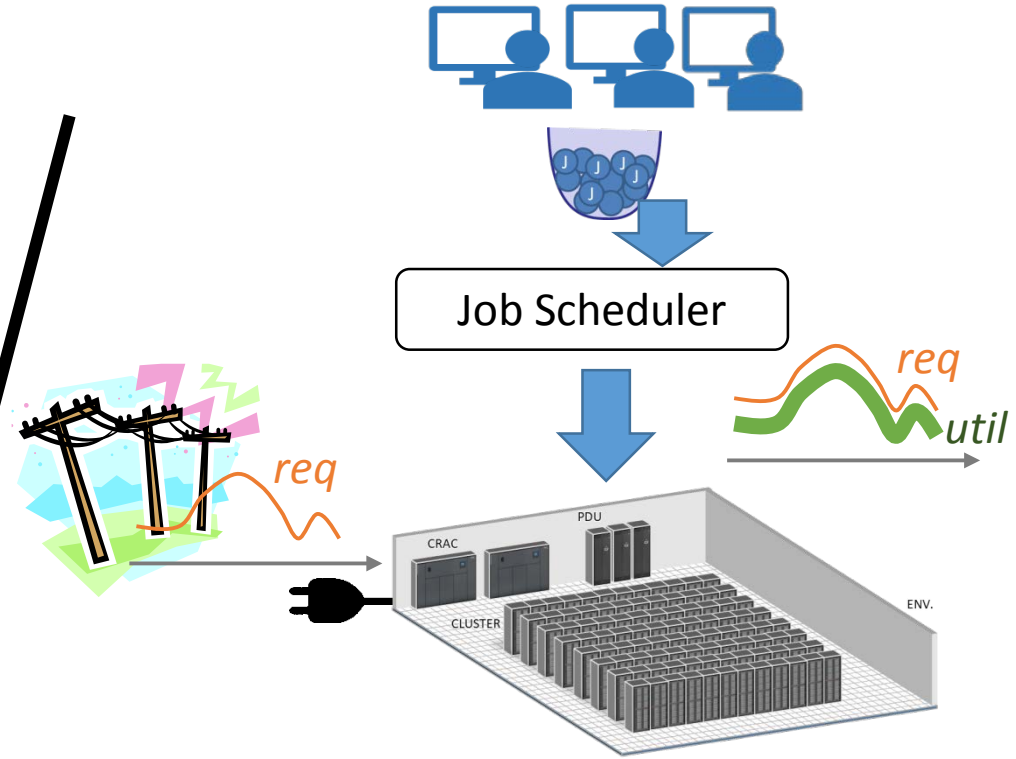
Usage Scenarios



Fine Grain Power and Performance Measurements:

- Verify and classify node performance
  - In spec / out of spec behaviour
  - Miss configuration
  - Aging and wear out
- Predictive maintenance

Fine Grain Required



System Power Capping

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

Coarse Grain Required

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

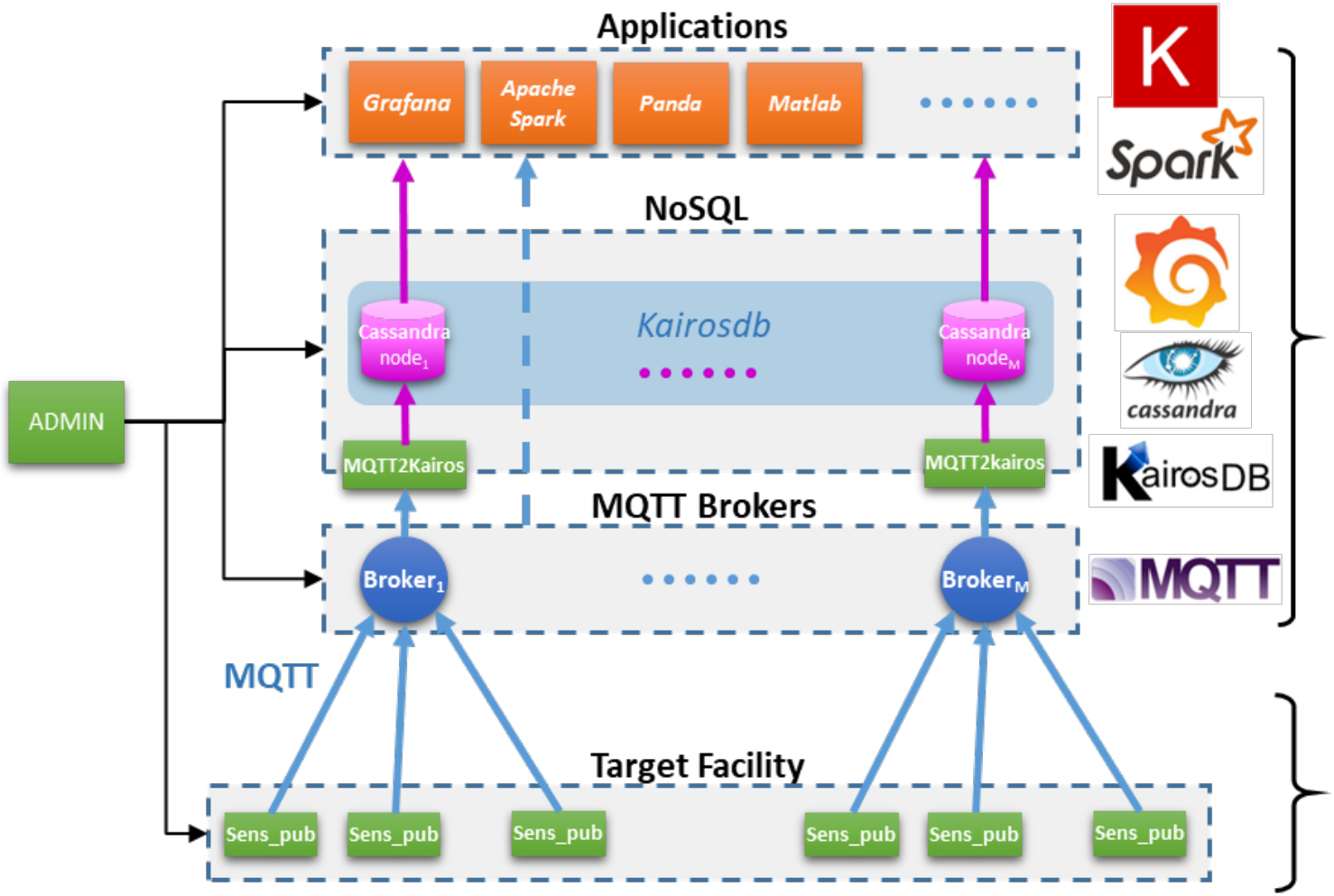
Outline

- Datacenter Automation
- **Holistic Monitoring @ CINECA with ExaMon**
- Holistic Monitoring + Stream and Edge processing
- Conclusions & Future Works

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Scalable Data Collection, Analytics

EXAMON



Front-end

- **MQTT** Brokers
- Data **Visualization**
- **NoSQL** Storage
- **Big Data** Analytics

Back-end

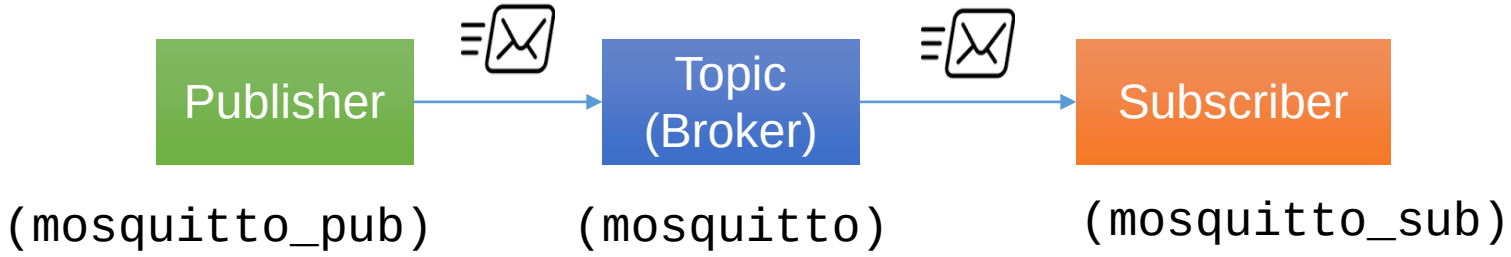
- MQTT–enabled sensor collectors

https://github.com/EEESlab/examon

F. Beneventi et al., "Continuous learning of HPC infrastructure models using big data analytics and in-memory processing tools"

MQTT: MQ Telemetry Transport

- **Lightweight** message queuing and transport protocol
- Developed by IBM and Eurotech
- Well suited for **low resource demanding** scenarios like M2M, WSN and **IoT** applications
- Basic features:
  - **PubSub** model
  - Async communication protocol (**messages**)
  - **Low overhead** packet (2 bytes header)
  - **QoS** (3 levels)
  - Open source implementation:
    - https://mosquitto.org/



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ExaMon MQTT Msg Formats

Topic format details:

<key>/<value>/<key>/<value>/ … **/plugin/**<plugin_name>**/chnl/**<data|cmd>/<key>/<value>/ … /<metric_name>

Sensor location
Free hierarchical sequence of key/value couples to locate sensors. At least one couple should be present.

Plugin name and channel type
(mandatory)
Channel type:
- *data*: metrics values sent by the plugin
- *cmd*: commands sent to the plugin

Specific plugin info
Additional free key/value couples to add custom plugin attributes (optional)

Plugin data

Metric name

Example:

Topic:

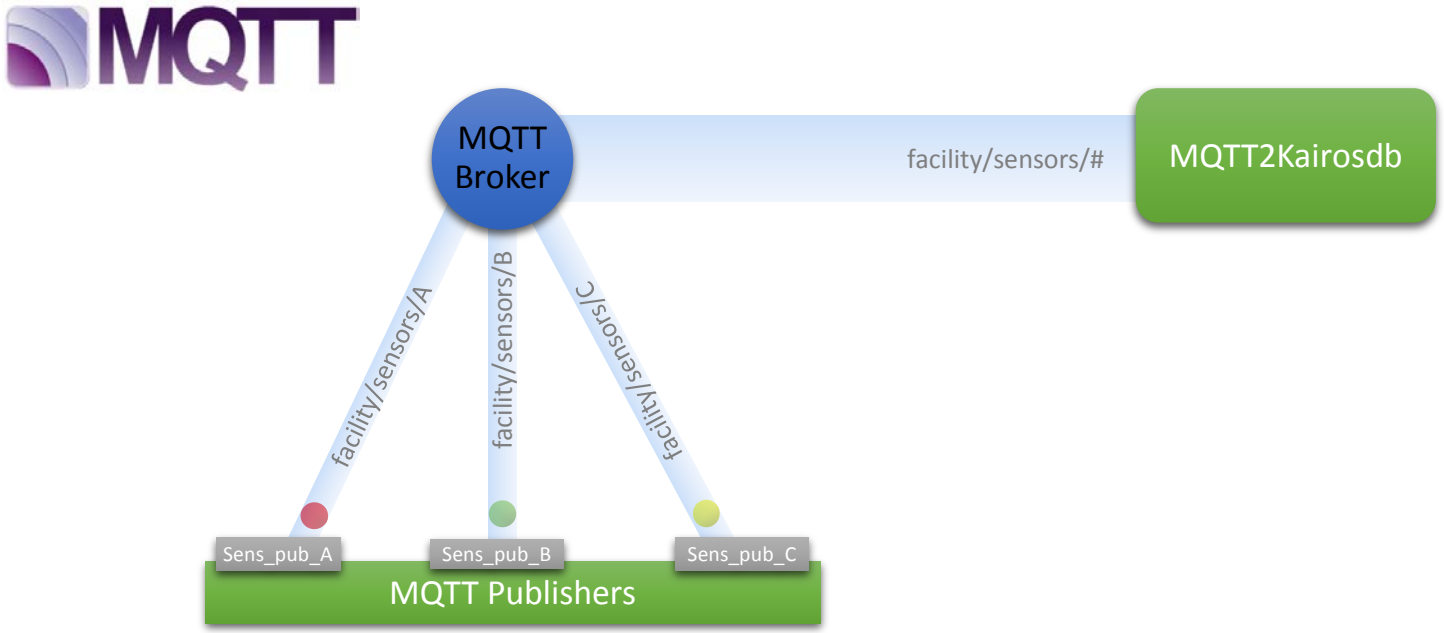
org/e4/cluster/davide/node/davide34/plugin/ame_pub/chnl/data/occ/1/cmp/CORE/id/10/unt/C/TEMP_P0

Payload:

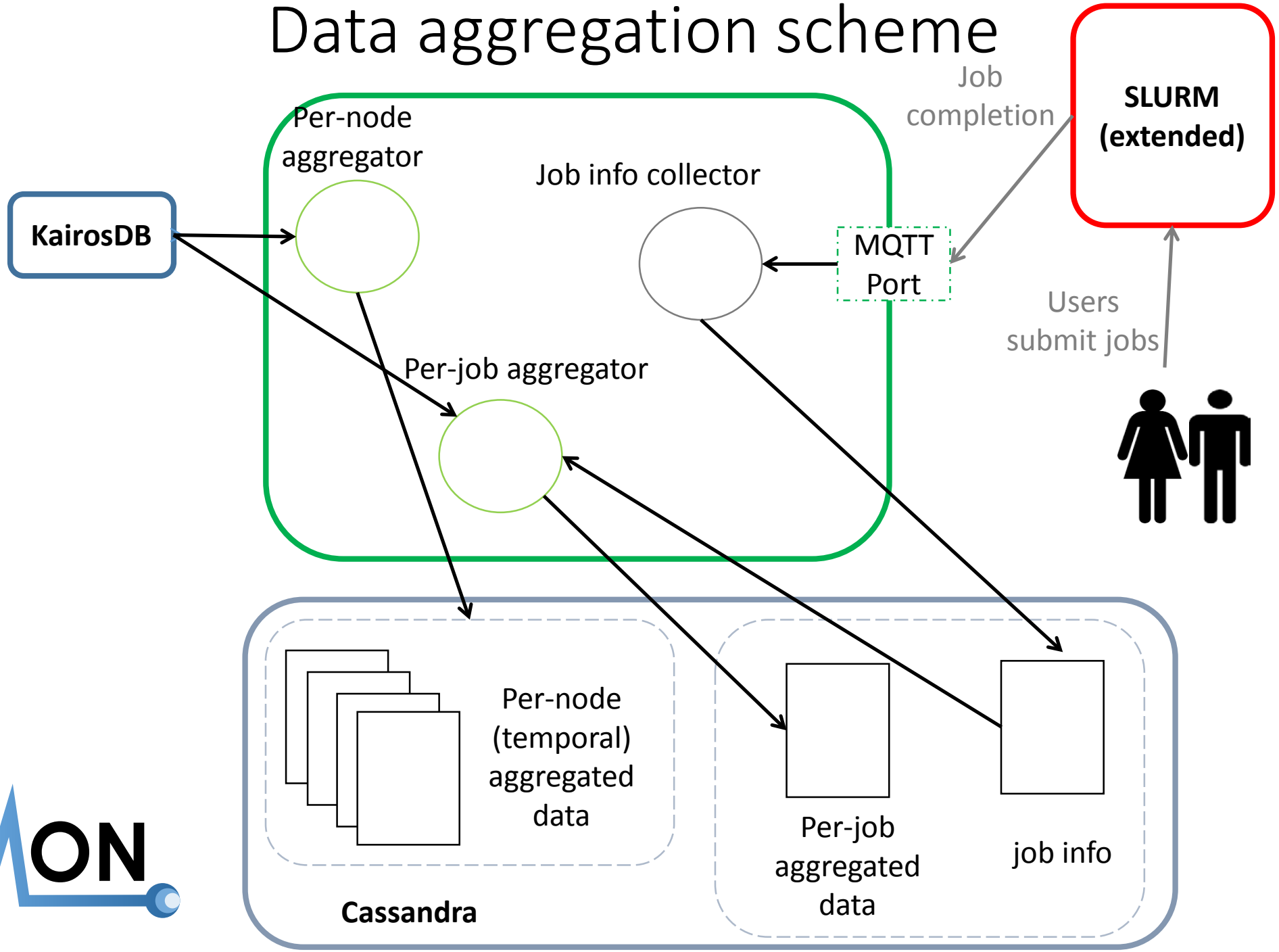
28.0;1511370841.62

MQTT to NoSQL Storage: MQTT2Kairosdb

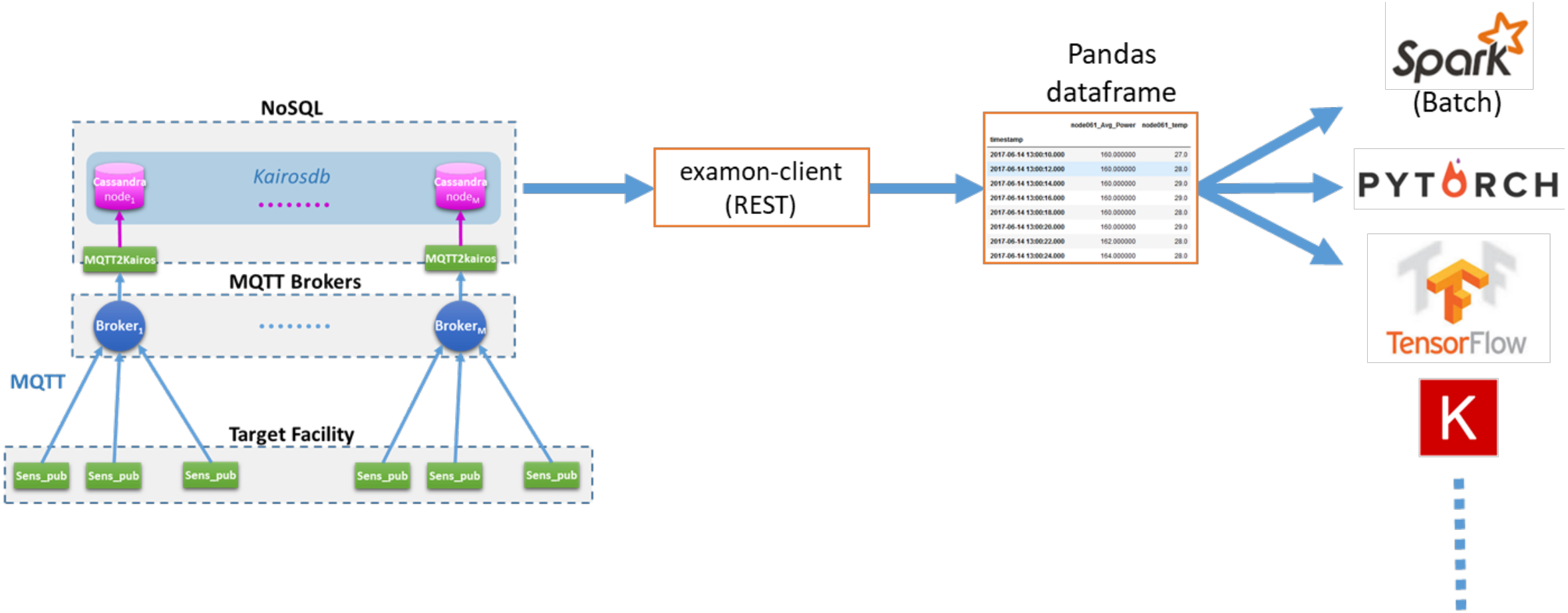
● = {Value;Timestamp}



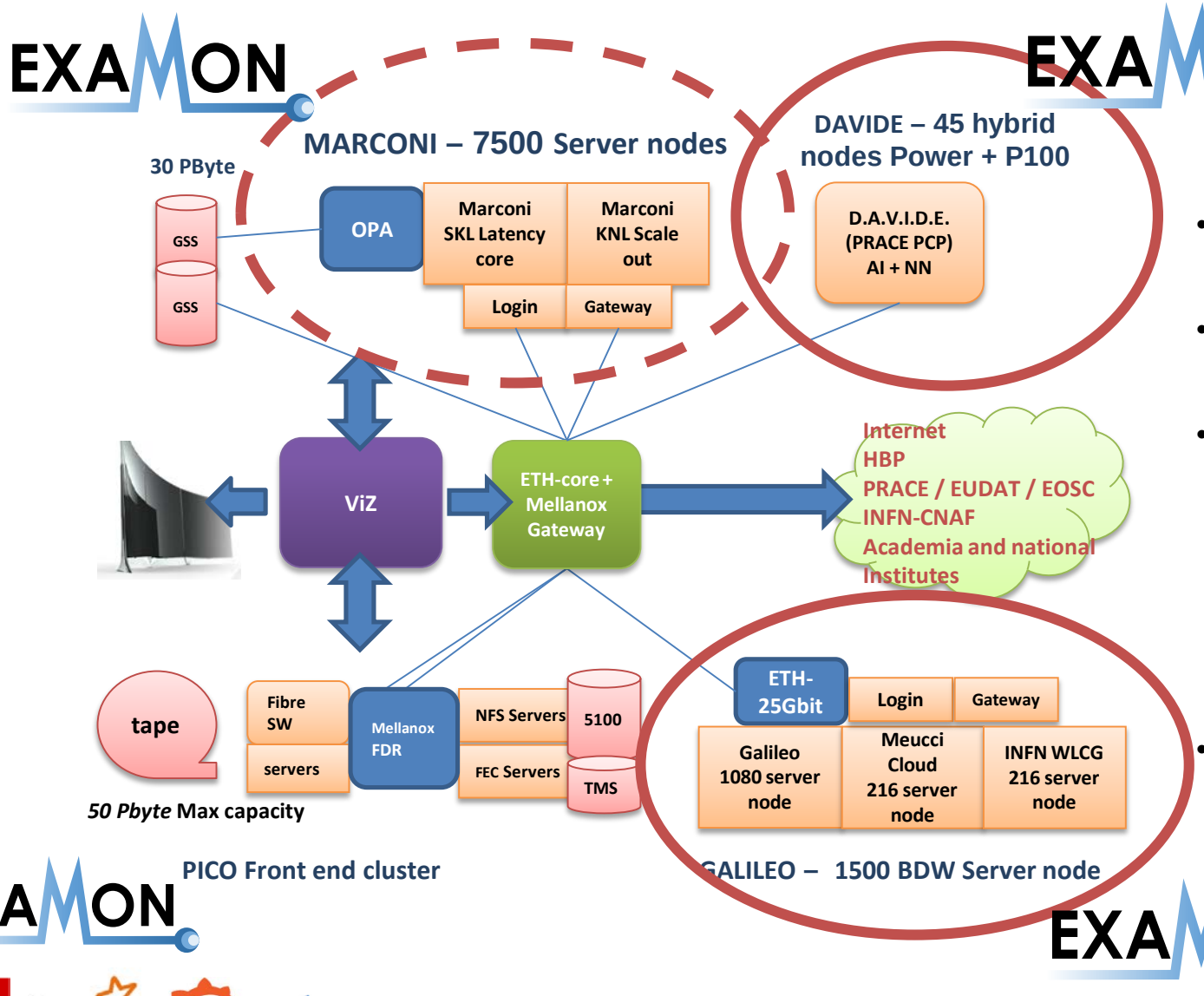
Data aggregation scheme



Examon Analytics: Batch training + Edge inference

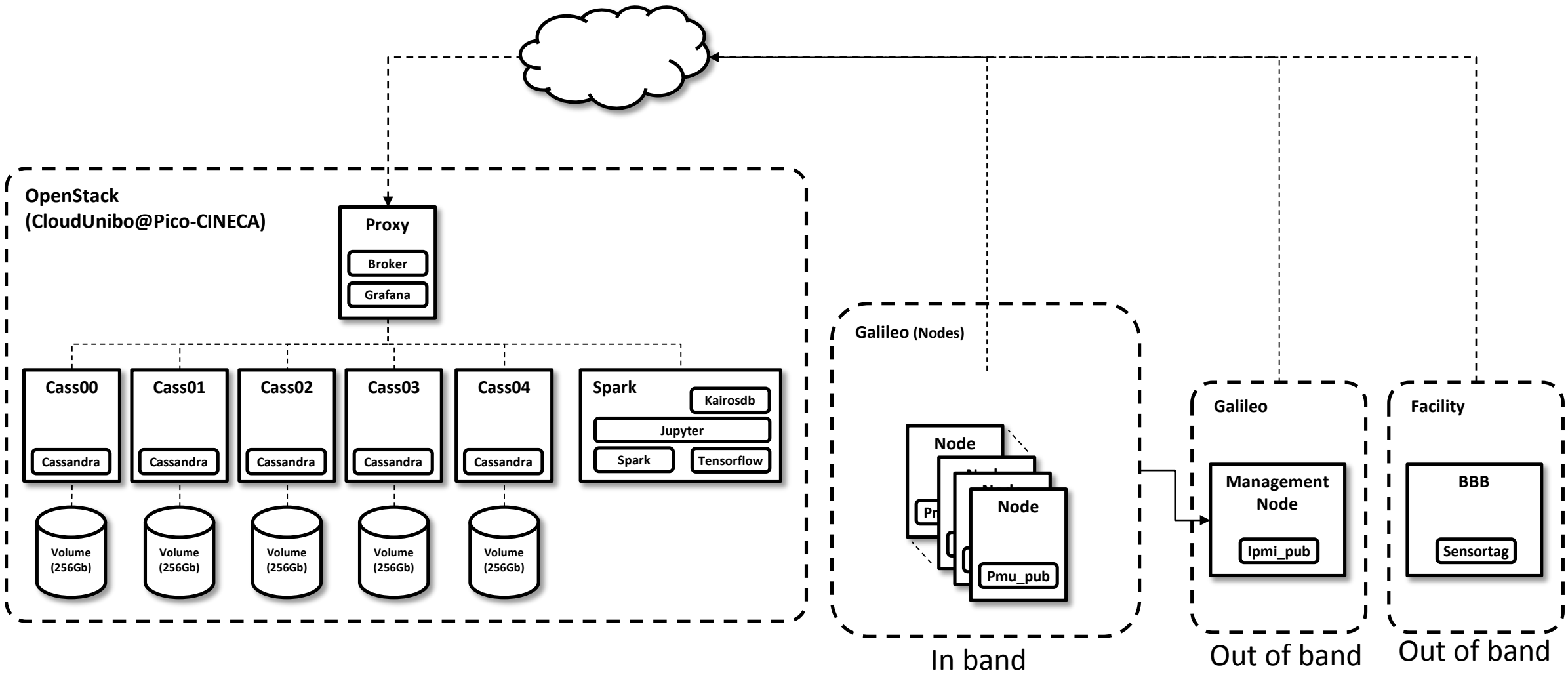


Current HPC architecture infrastructure



- Global peak performance in excess of 25 Pflops
- Global production capacity > 9000 servers nodes
- Modularity computing combining different microprocessors technology based on the same asics and different software stack enabling effective production work flows
- Architecture development driven by the scientific challenges and applications domains

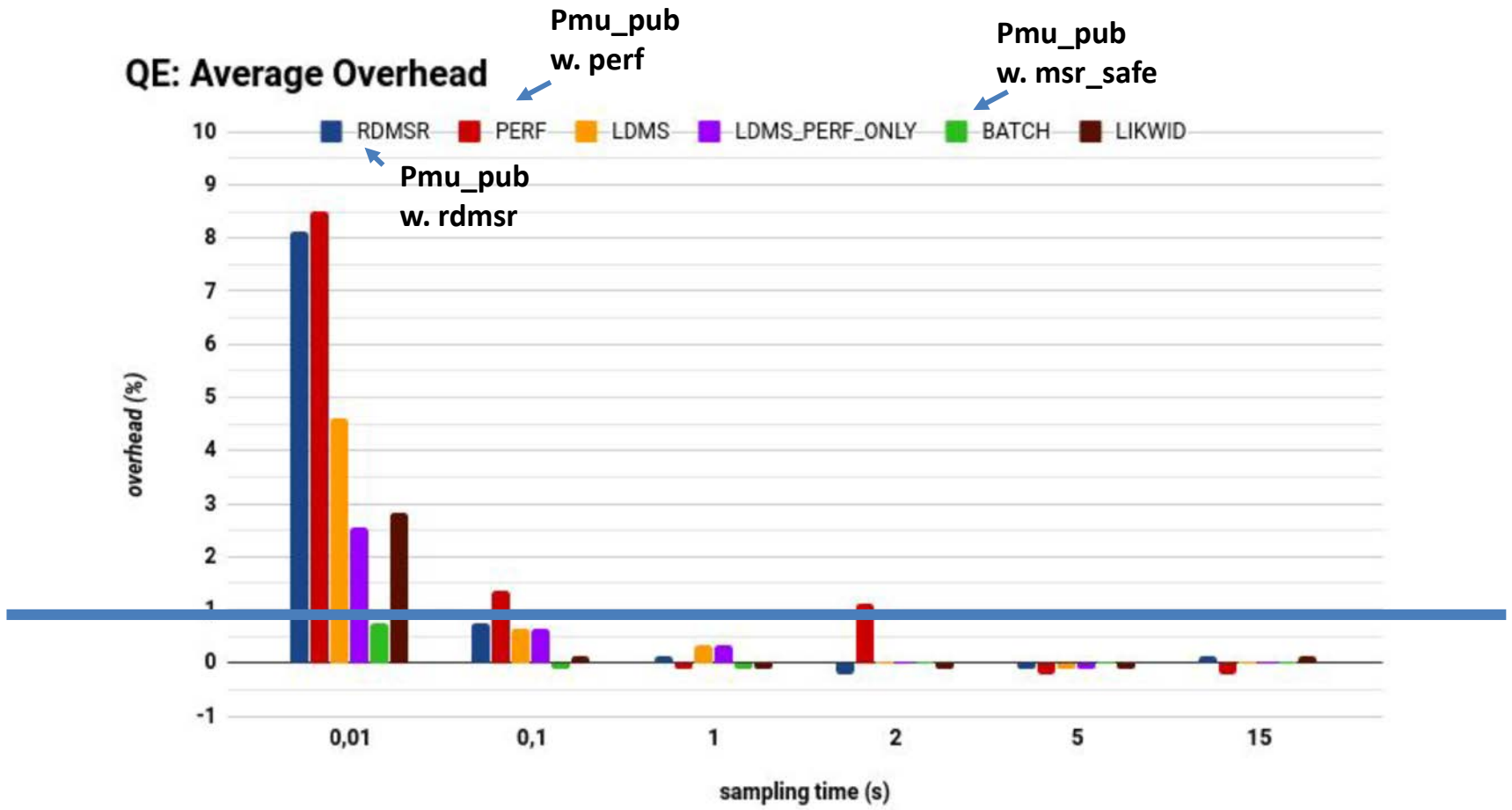
Galileo: Current setup



Pmu_pub (in band)

Low-overhead in-band always-on monitoring is feasible

- Overhead @ 236 counters/node



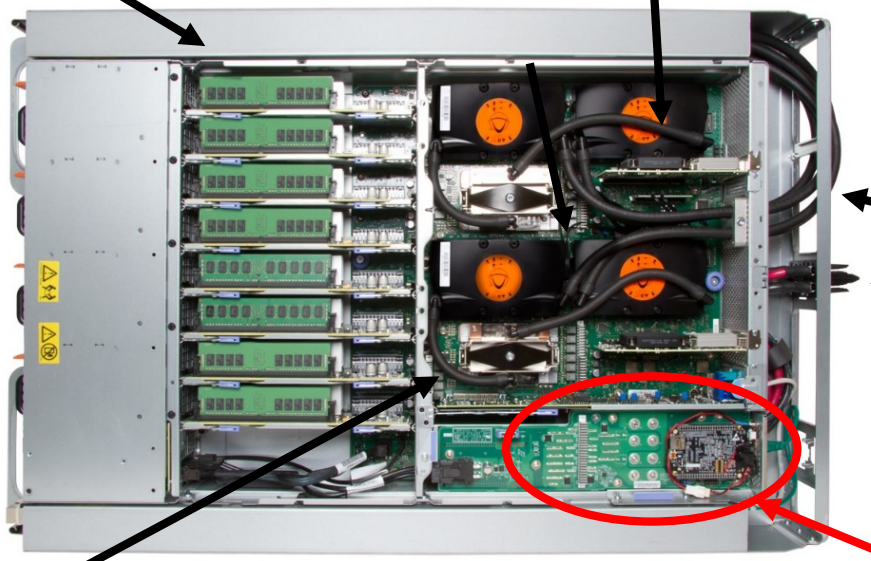
Pmu_pub =
MQTT + msr_safe + batch
Negligible overhead
down to Ts = 10ms

DiG = High Frequency Monitoring on D.A.V.I.D.E.

OCP form factor compute node based on IBM Minsky

4x NVIDIA Tesla P100 HSMX2

2 x IBM POWER8 with NVLink



2xIB EDR

BusBar

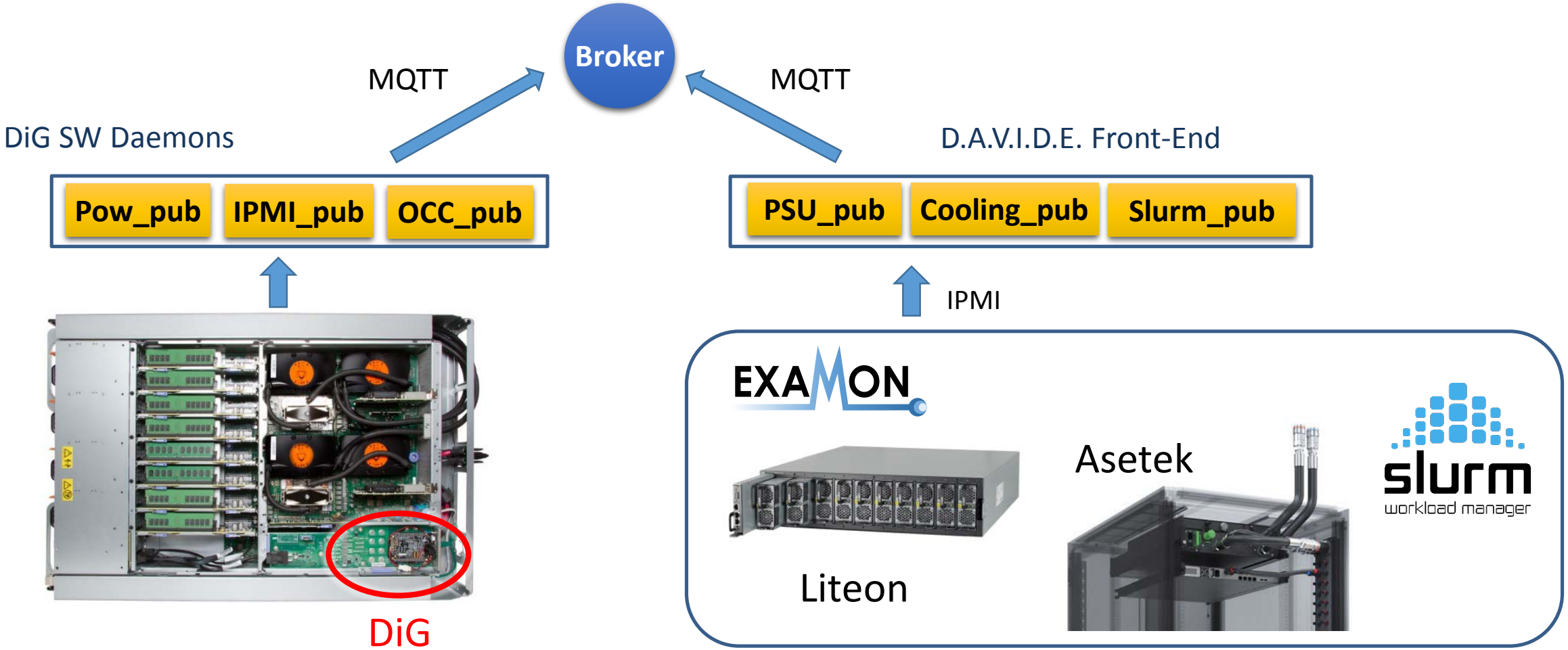
DiG

ETH Zurich / Univ. of Bologna
SoA out-of-band
High Resolution Power
and Performance Monitoring

LIQUID COOLING

E4 COMPUTER ENGINEERING
PRACE
wistron
CINECA
TOP 500 The List.
THE GREEN 500
OpenPOWER

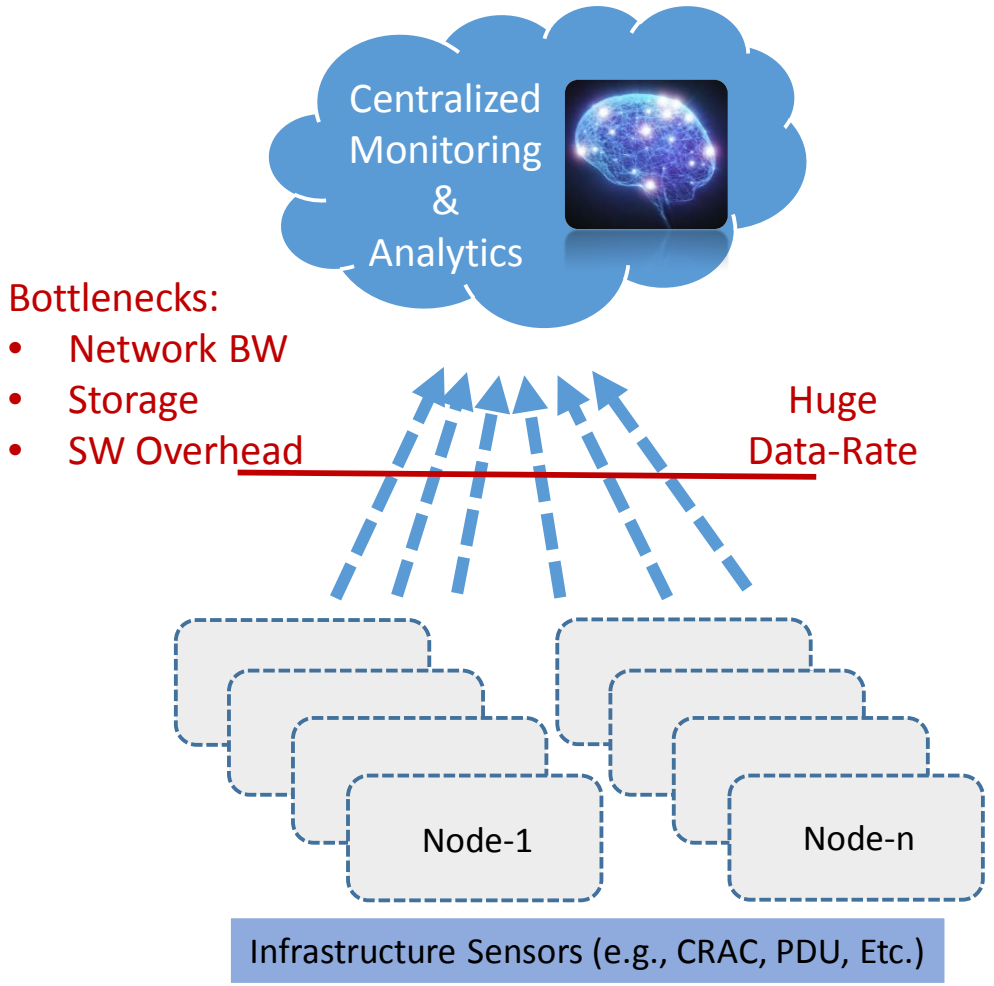
DiG = High Frequency Monitoring on D.A.V.I.D.E.

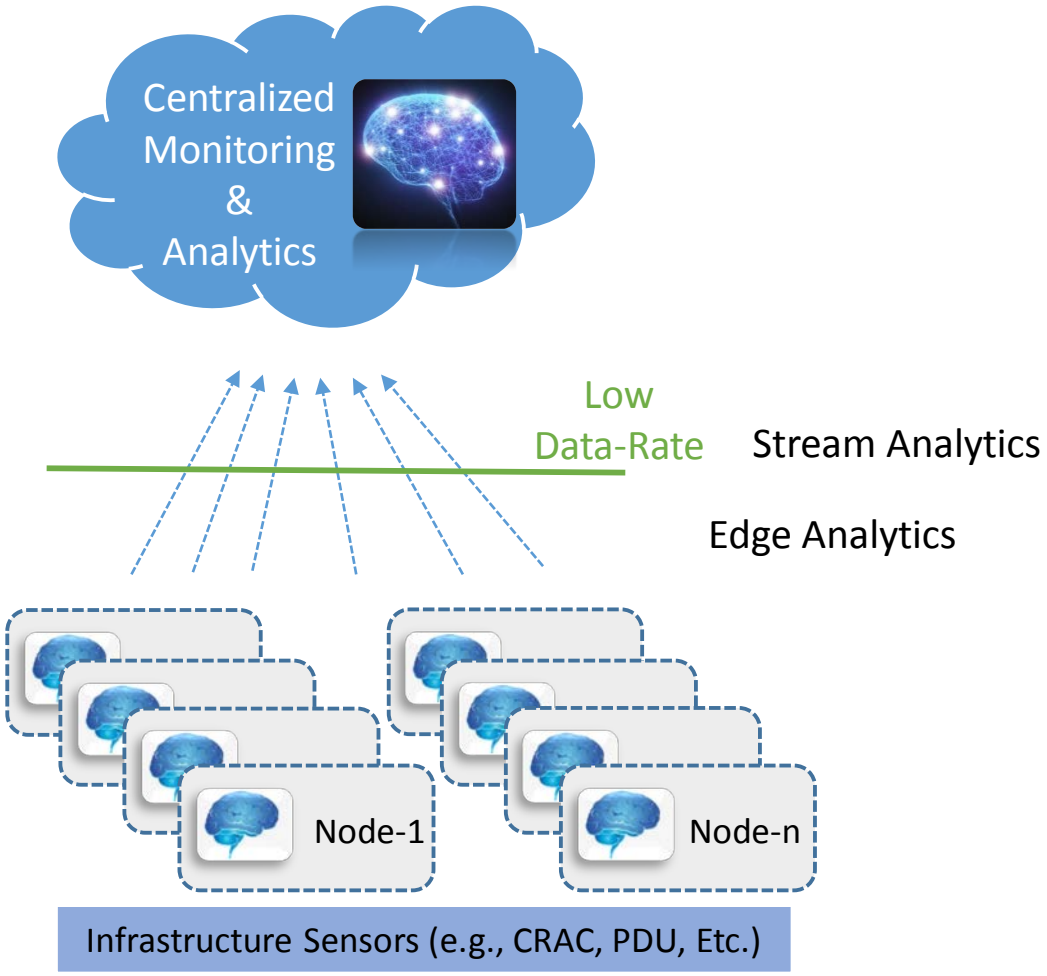


Outline

- Datacenter Automation
- Holistic Monitoring @ CINECA with ExaMon
- **Holistic Monitoring + Stream and Edge processing**
- Conclusions & Future Works

Datacenter Automation Design and Bottlenecks



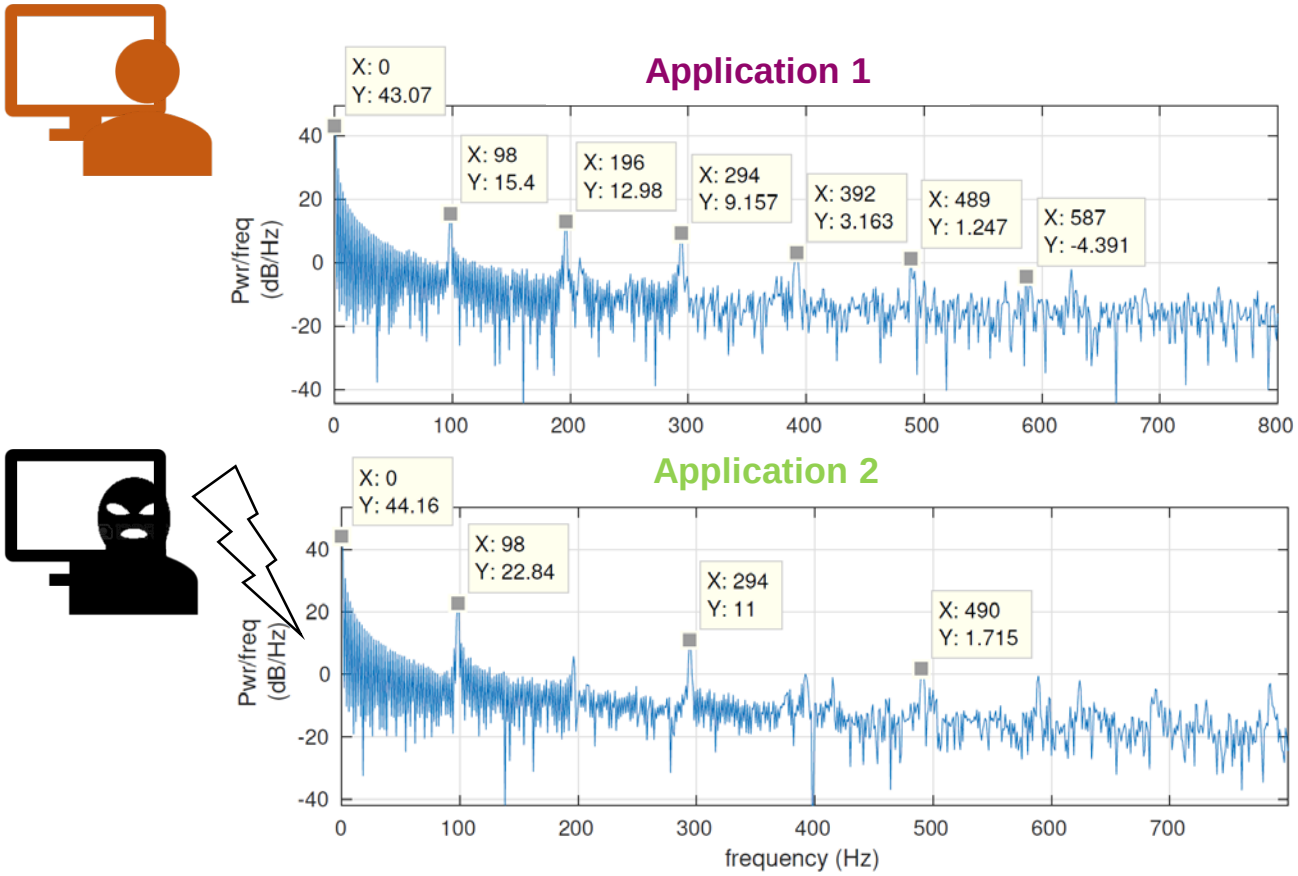


What to do with high frequency power monitoring

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Real-time Frequency analysis on power supply and more…a live oscilloscope

- For instance, using the FFT we plot the power spectral density of the power benchmark of two applications, and we can distinguish them by the harmonics present in each of the signals



Spectral signature of an application!

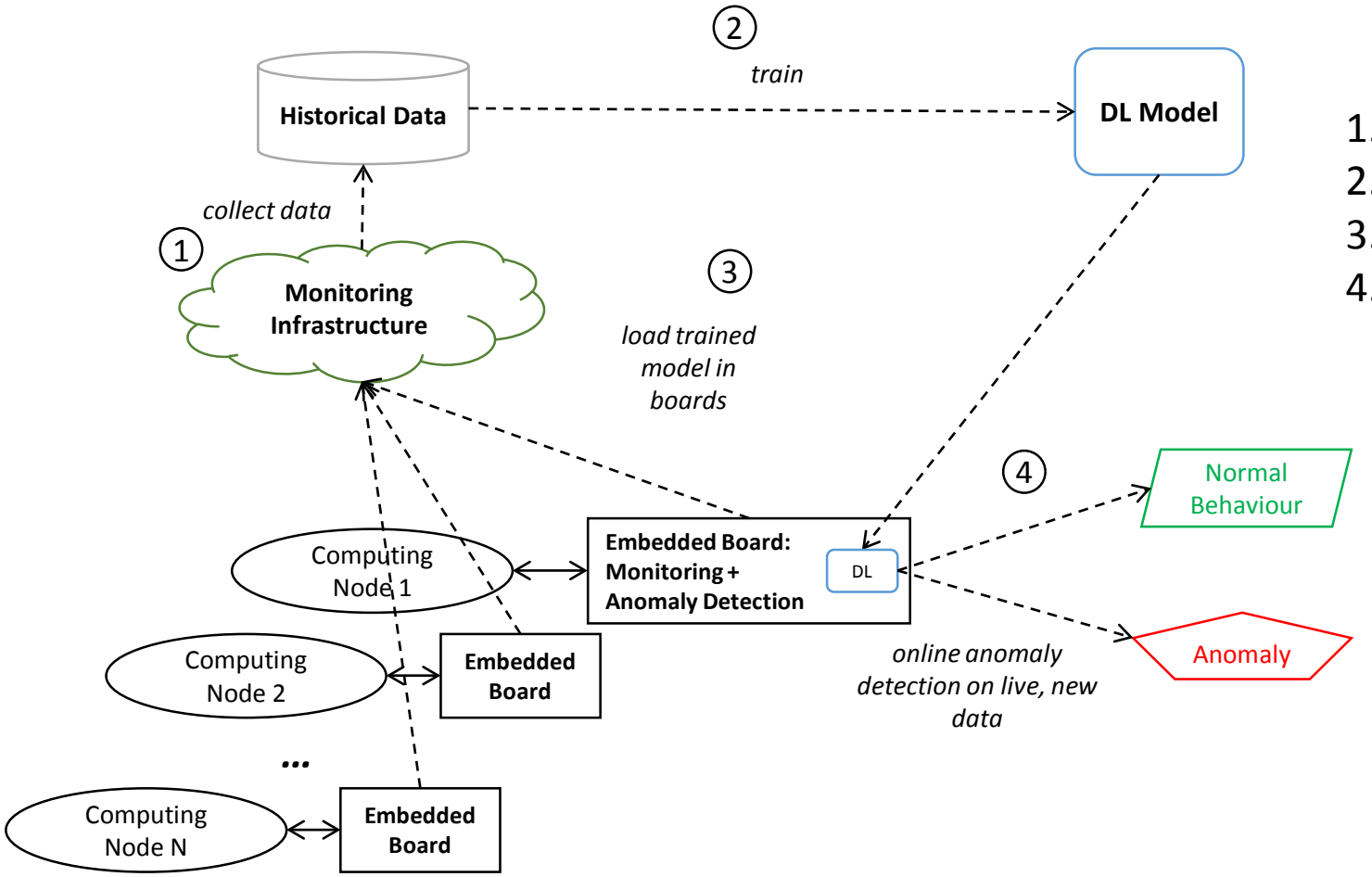
Interesting feature for node level and system level

Intrusion Detection

Can be done online on the Edge w. DiG !

AI+Big Data on D.A.V.I.D.E.: Anomaly detection

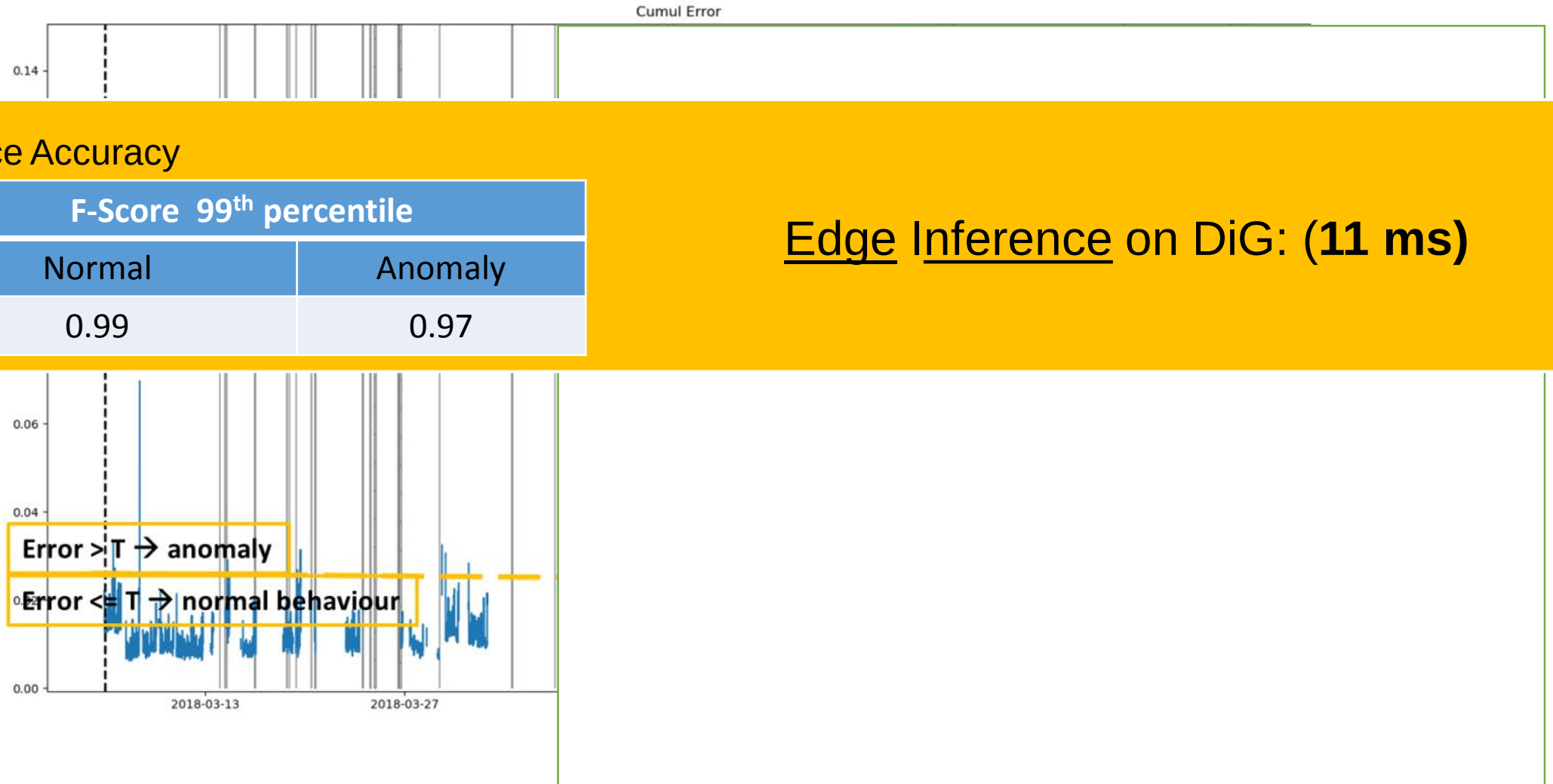
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



1. Collect Data
2. AI Train
3. Edge
4. AI Inference

AI+Big Data on D.A.V.I.D.E.: Anomaly detection

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

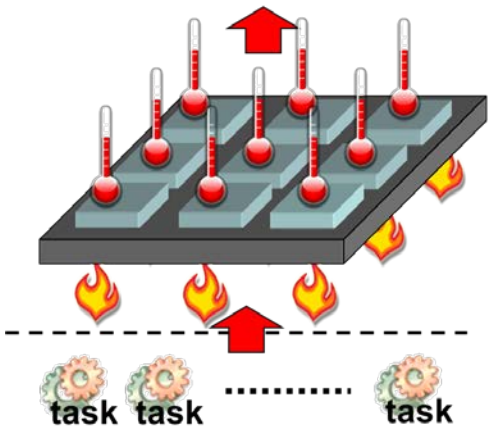


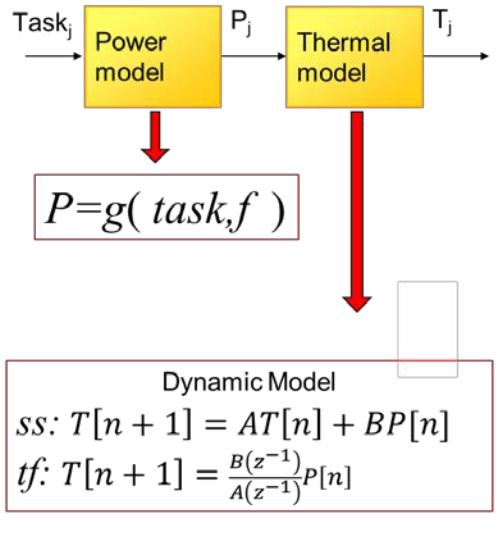
Inference Accuracy

F-Score 99th percentile	
Normal	Anomaly
0.99	0.97

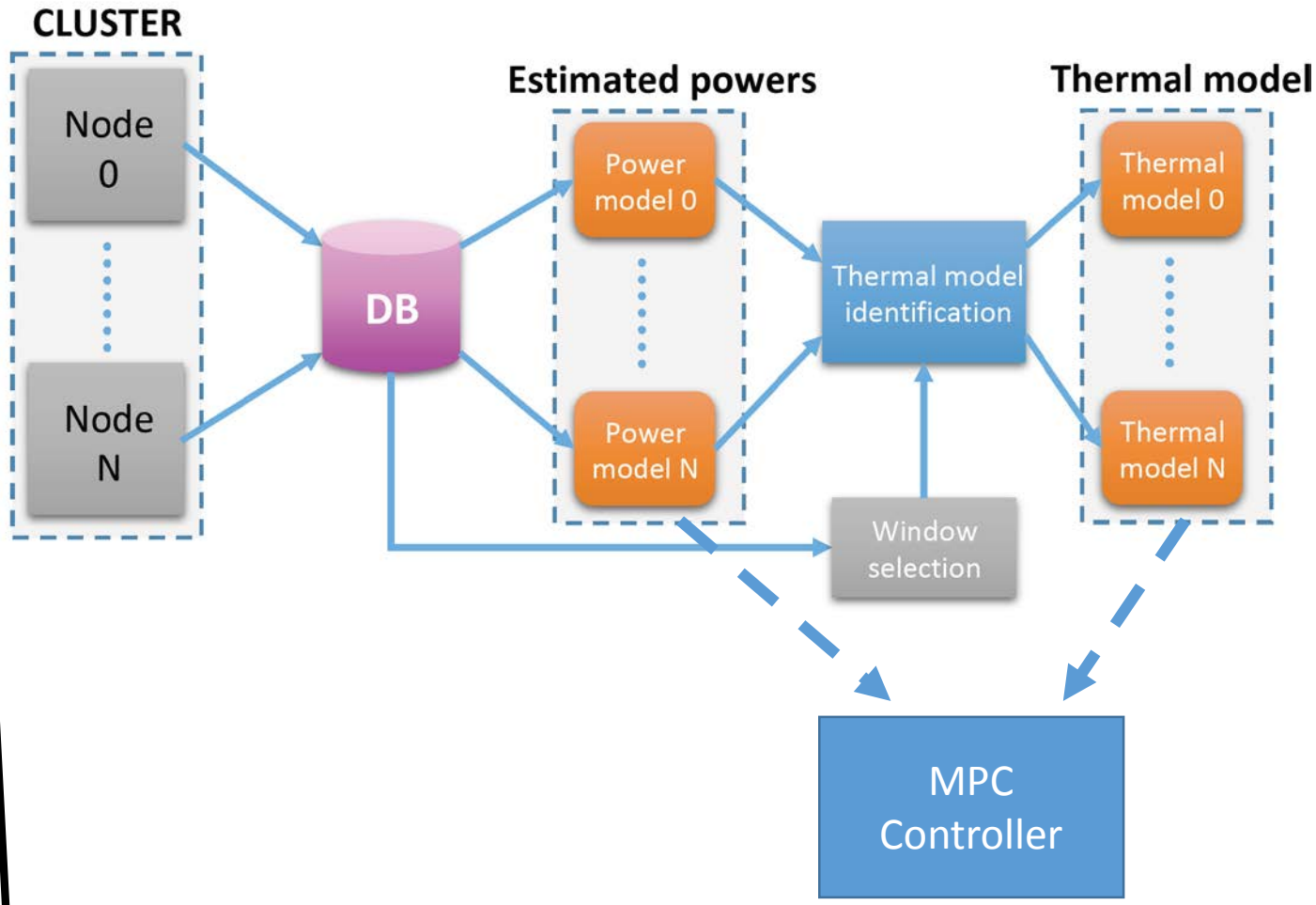
Edge Inference on DiG: (**11 ms)**

Predictive Thermal & Power Modeling at scale

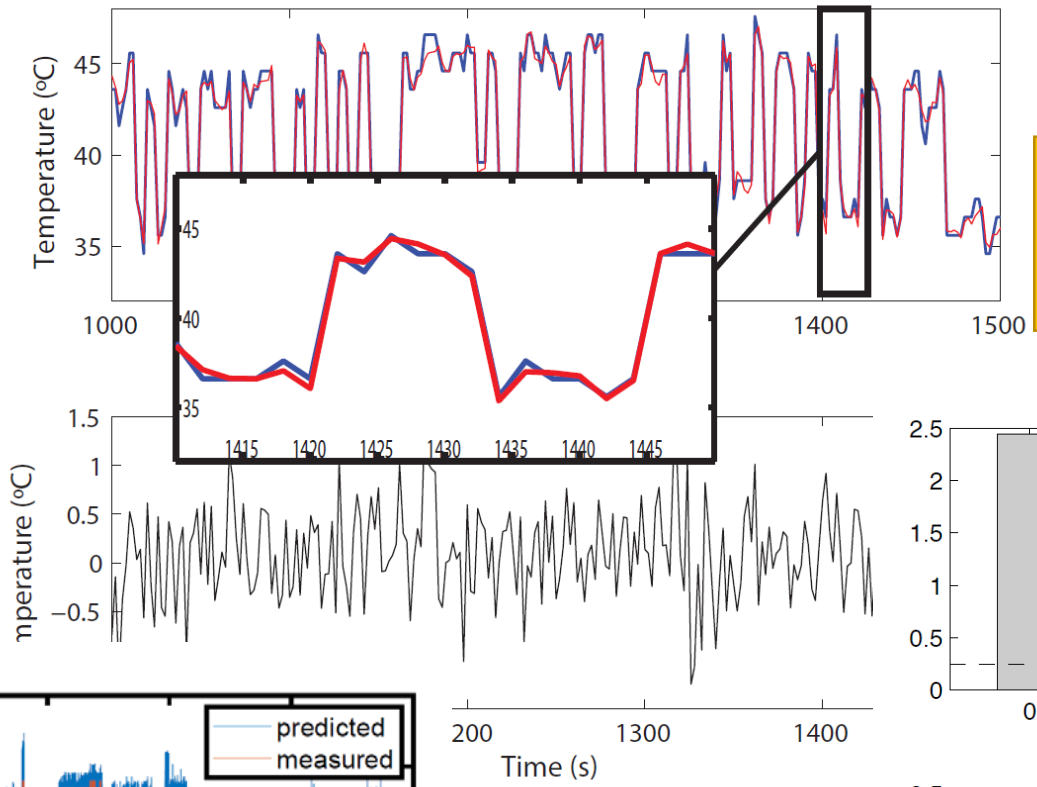




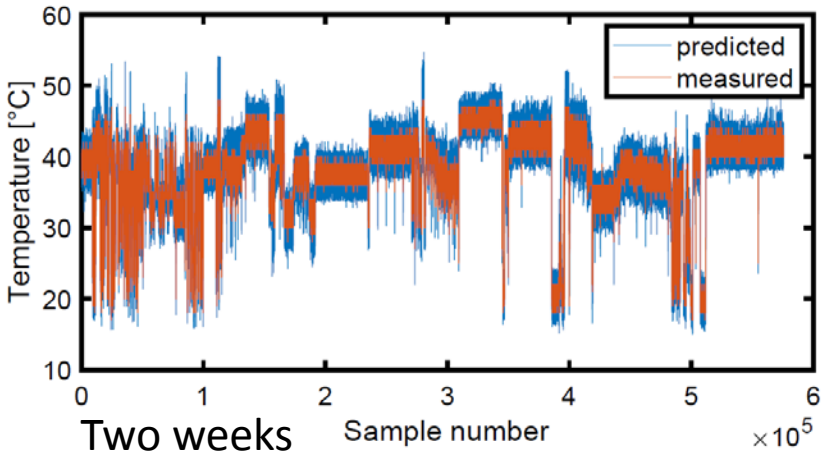
Thermal model



Galileo Modelling



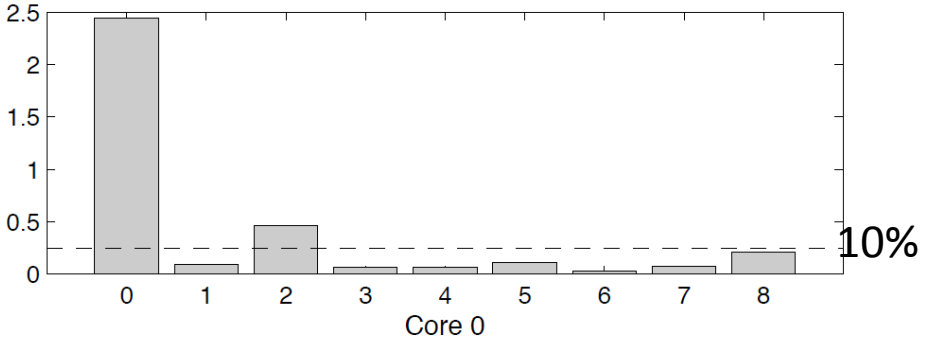
Very accurate prediction <1°C also with quantization noise (±1°C)

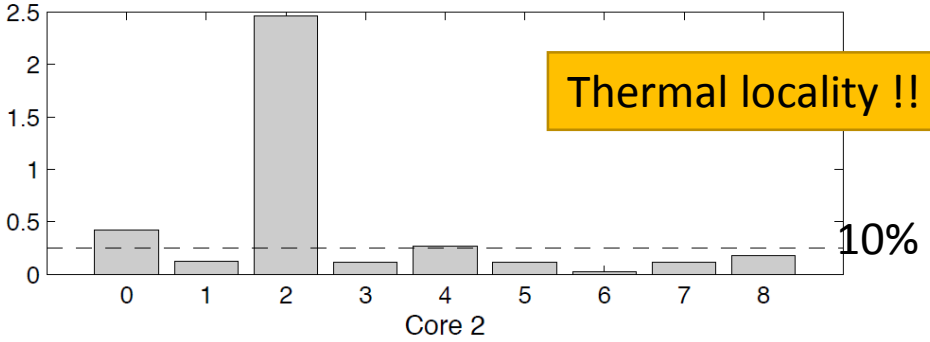


Two weeks

(b) window 22

Gain w.r.t. core Id





Thermal locality !!

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Conclusion & Future Works

- Holistic and Fine Grain Monitoring feasible w. Open Source and Scalable of the shelf tools
- Challenge in the effective usage and knowledge extraction from the monitored data
- AI, ML and Predictive Control can leverage the infrastructure toward datacentre automation.
- Future Works:
  - Scale anomaly detection at the full datacentre level
  - Integrate thermal modelling with model predictive control/AI and emerging job level power manager (a.k.a Intel GEOPM)

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Multitherman

EU FP7 ERC Project
MULTITHERMAN
(g.a.291125)

ACKNOWLEDGE

The UNIBO/ETHZ Datacenter Automation TEAM

- Luca Benini, Michela Milano, Andrea Tilli, Roberto Diversi, Cristian Conficoni, Andrea Borghesi, Daniele Cesarini Antonio Libri, Francesco Beneventi, Federico Pittino

CINECA:

- Sanzio Bassini, Carlo Cavazzoni, Elda Rossi, Daniela Galletti, Alessandro De Federico, Isabella Baccarelli, Marco Sbrighi, Giuseppa Muscianisi

EX 10^18

EU H2020 FETHPC
project ANTAREX
(g.a. 671623)

ETH zürich

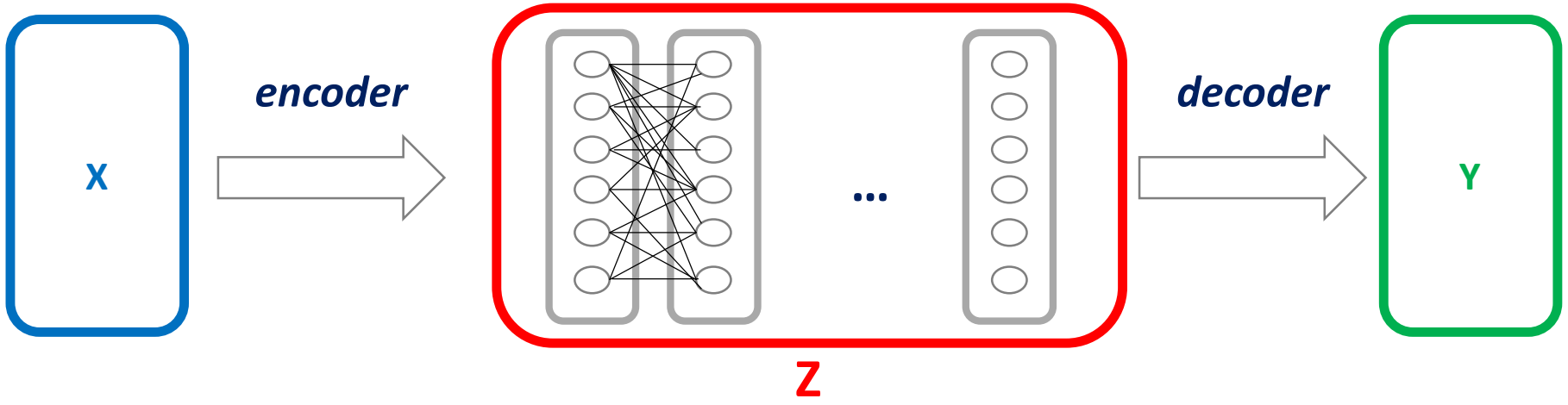
PRACE

CINECA

E4
COMPUTER
ENGINEERING

Anomaly Detection

AUTOENCODER



- An autoencoder tries to copy its input (X) into its output (Y)
- An autoencoder learns to represent its input in the latent space, extracting the important characteristics of the input set X

IDEA: train an autoencoder with the normal behavior of a HPC system and use its reconstruction error to detect anomalies