Workflows, a crucial challenge for the future of HPC

43e Forum Orap

March 19, 2019

F. Bodin

Introduction

- HPC vision as numerical models running on a supercomputer is not representative anymore

- Data deluge / revolution is changing the game*
  - Atmospheric Radiation Measurement
  - Personalized Medicine
  - Seismic imaging of hydrocarbon reservoirs
  - Light Sources
  - Smart Cities
  - Precision agriculture
  - Digital Manufacturing/Industry 4.0
  - ...

- HPC included in a larger ecosystem
  - IoT, Big Data, Cloud technology

*Big Data and Extreme-scale Computing (BDEC) Project
A serie of international workshops to foster community convergence on a next generation cyberinfrastructure for science

Outline of the Presentation

1. Applications

2. Resources

3. Workflow challenges

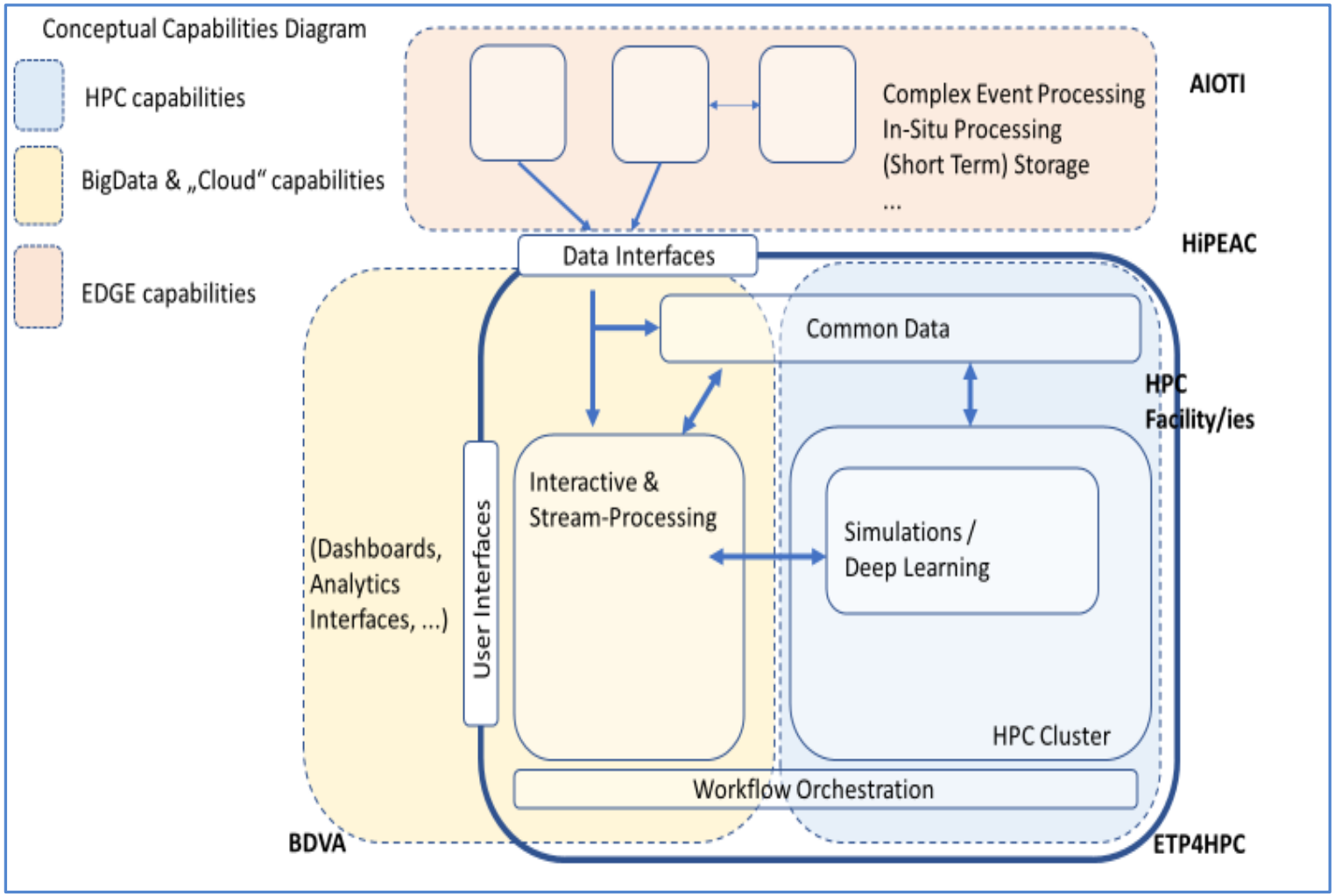
Applications

- Many new applications crossing the usual HPC frontiers
  - IoT
  - Urgent computing
  - Increasing use of AI techniques
  - Digital Twins

Digital Twins

- Numerical representation of a part of the real world
  - Capture the temporal-spatial relationships
  - Inform & predict
- Virtual testing platform
  - Allow experimentations
  - Combined with data from the IoT, ...
- Numerical and qualitative models
  - CFD, actors models, DL based models, ...

Digital Twins

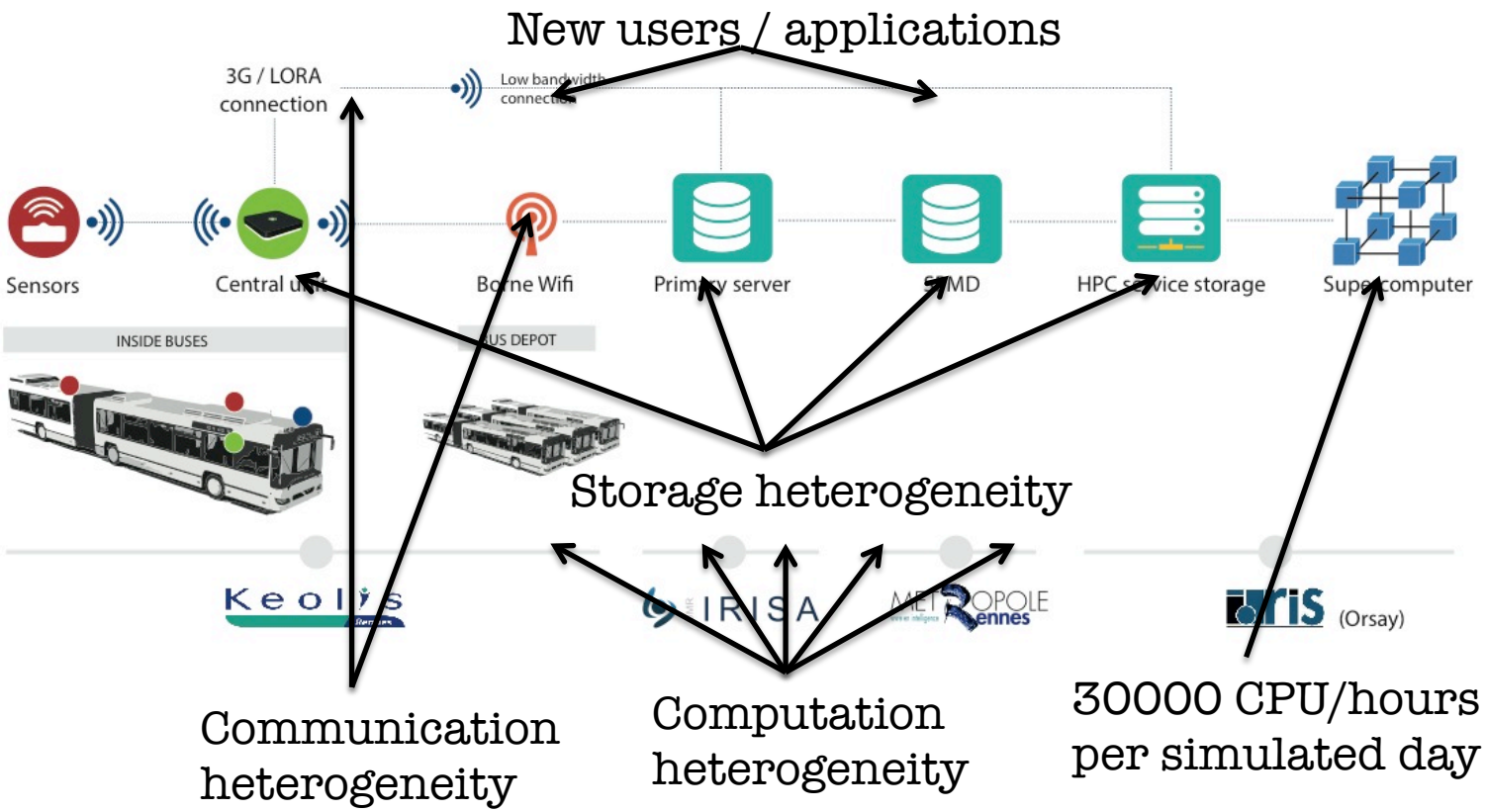


Src: Jens Krueger ITWM Faunhofer, BDVA/ETP4HPC/EXDCI2

Air Quality & Mobility

AQMO

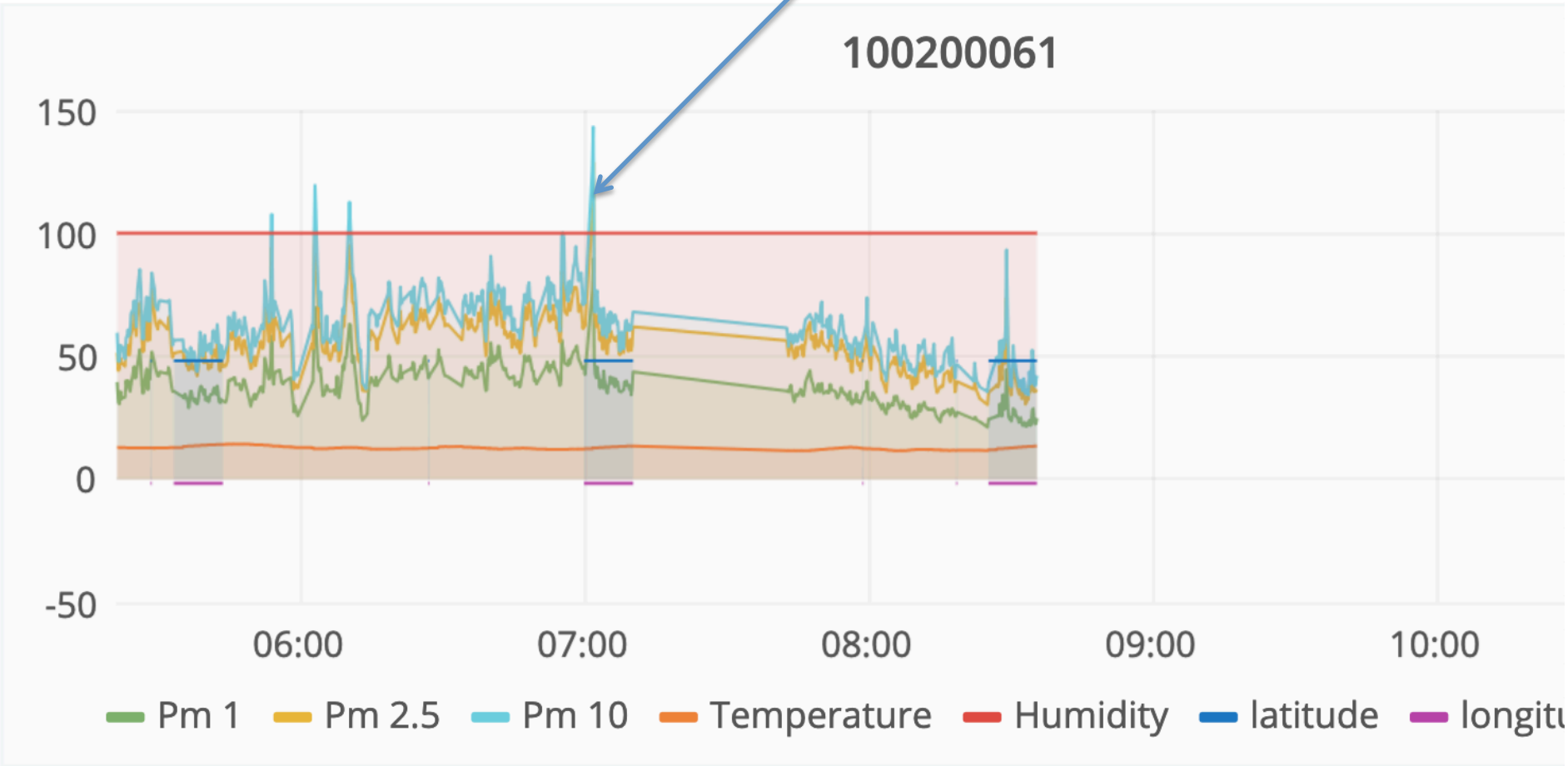
Co-financed by the Connecting Europe Facility of the European Union



http://www.aqmo.eu

Sensor Data from Bus

Use of machine learning to get measurement context



AQMO Technical Background

- Massive heterogeneity in compute, network and storage resources
  - At the technical level
  - At the governance level: multi-owners, multi-tenants
- Sensors, weather, topology and simulation data
  - Accumulation over time
  - Edge and HPC computing in the same workflow
- Multiple kinds of networks (LoRa, 4G, WLAN)
  - Some connectivity can be intermittent

Conclusion on Applications

- Different execution profile must be taken into account
  - Long-running jobs based on streaming data
- Involve many research groups that need to share storage and compute resources

➔ Data logistic can be very complex

Compute and Storage Resources

- Sensors and scientific instruments
- **Edges**, **Fog**, Data Centers, Cloud, HPC centers*
- New **agile resources** for storage and compute
  - Adapt to the fast changing landscape

*CI2030: Future Advanced Cyberinfrastructure
A report of the NSF Advisory Committee for Cyberinfrastructure

Edge Computing

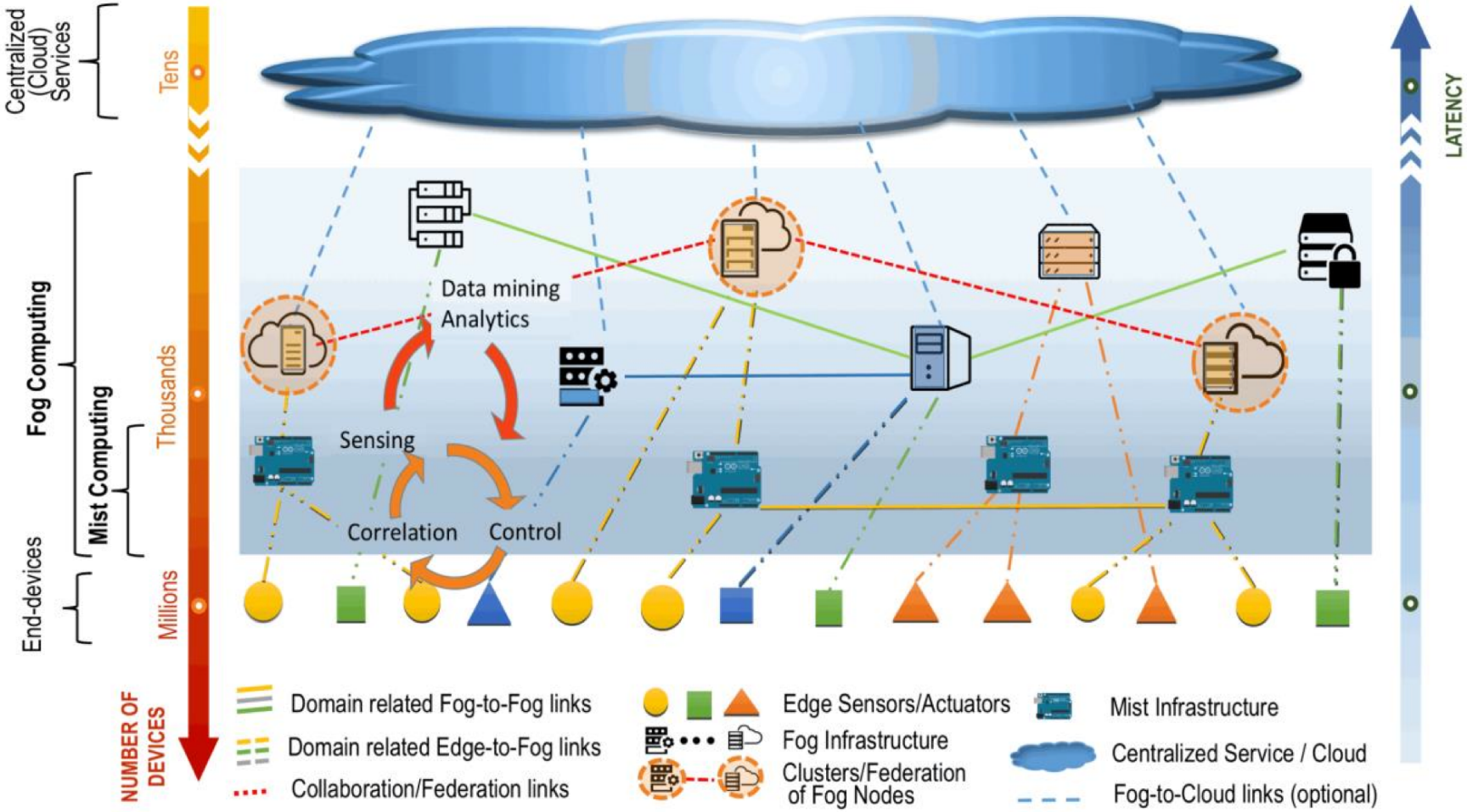
- Compute and store close to the sensors
  - To avoid moving data
  - For privacy motives
- Can use HPC technologies
  - e.g. GPU computing
- Need to be integrated in the data logistic and application workflows

FOG Computing-1

- Fog computing is a layered model for enabling ubiquitous access to a shared continuum of scalable computing resources
- Into the network itself decentralized
  - applications,
  - management,
  - data analytics
- Based on a distributed and federated compute model

Src: NIST Special Publication 500-325, Fog Computing Conceptual Model
Recommendations of the National Institute of Standards and Technology

FOG Computing-2



Src: NIST Special Publication 500-325, Fog Computing Conceptual Model
Recommendations of the National Institute of Standards and Technology

Agile Storage

- Agile access to storage infrastructure resources should be provided
  - Needed in a fast changing science
- Federation of infrastructures
  - Focus on sharing data
- A new initiative led by Alex Szalay
  - Aim at providing a design that is scalable and flexible
  - https://www.openstoragenetwork.org/

Conclusion on Resources-1

- Need to construct a federation of private and public infrastructures
  - Need to increase permeability of supercomputing systems to facilitate data injection
- Need a formulation of a vision for a future service-oriented architecture framework
  - For HPC compute and other compute services as well as storage and other data services/logistic
- Network topology obliviousness is an issue
  - Data logistic depends on network topology, links throughputs, storage placements, etc.

Conclusion on Resources-2

- Heterogeneity of systems/networks components is a great source of complexity
  - but we cannot impose homogenization!

- Multi-tenants, multi-owners adding governance issue

- Agile resources needed as well as larger more permanent infrastructures
  - But need sharing experience, specification, etc.
  - Better if federated

- Archiving data is under-estimated
  - 50K€ / petabytes / year (~AWS)

Workflows Challenges

- Providing tools and governance for deploying large scale, science-driven workflows
  - Orchestrating high-end data analysis, intensive computing stages
  - Increasing use of AI techniques
  - Across a continuum of resources
- Data logistics must be deployed across the continuum of compute and data resources

Conclusion

- **Dropping frontiers** between systems is needed to address current scientific challenges
- Workflows deployment across a continuum of resources is a **technical** and **governance issue**
- **Security/privacy** needs to be addressed in a way that it does not forbid moving forward