



CYBELE

BDVA: HPC, Big Data, IoT and AI future industry-driven collaborative strategic topics (part 2)

Dr. Sophia Karagiorgou, UBITECH

03/07/2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825355.

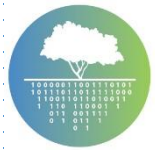




Societal challenges to address

- One third of food produced is lost or wasted every year;
- This loss is due to inefficiencies in planting, harvesting, feeding, water use, and uncertainty about weather;
- Global food waste and loss cost \$940 billion a year and have a carbon footprint contributing in more than 8% of global greenhouse-gas emissions;
- At the same time, the need for more and better-quality food increases.

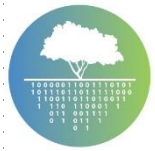




Technical challenges to address

- Large volumes of data request diverse and online computing modalities for collection, processing and analysis;
- When data converge at the testbeds require efficient and distributed data services (curation, anonymization, enrichment);
- Upon data analysis, complex and dynamic workflows require intelligent mechanisms bridging the *Big Data and HPC worlds*;
- Voluminous analysis results require adaptable and non-blocking visualization services.





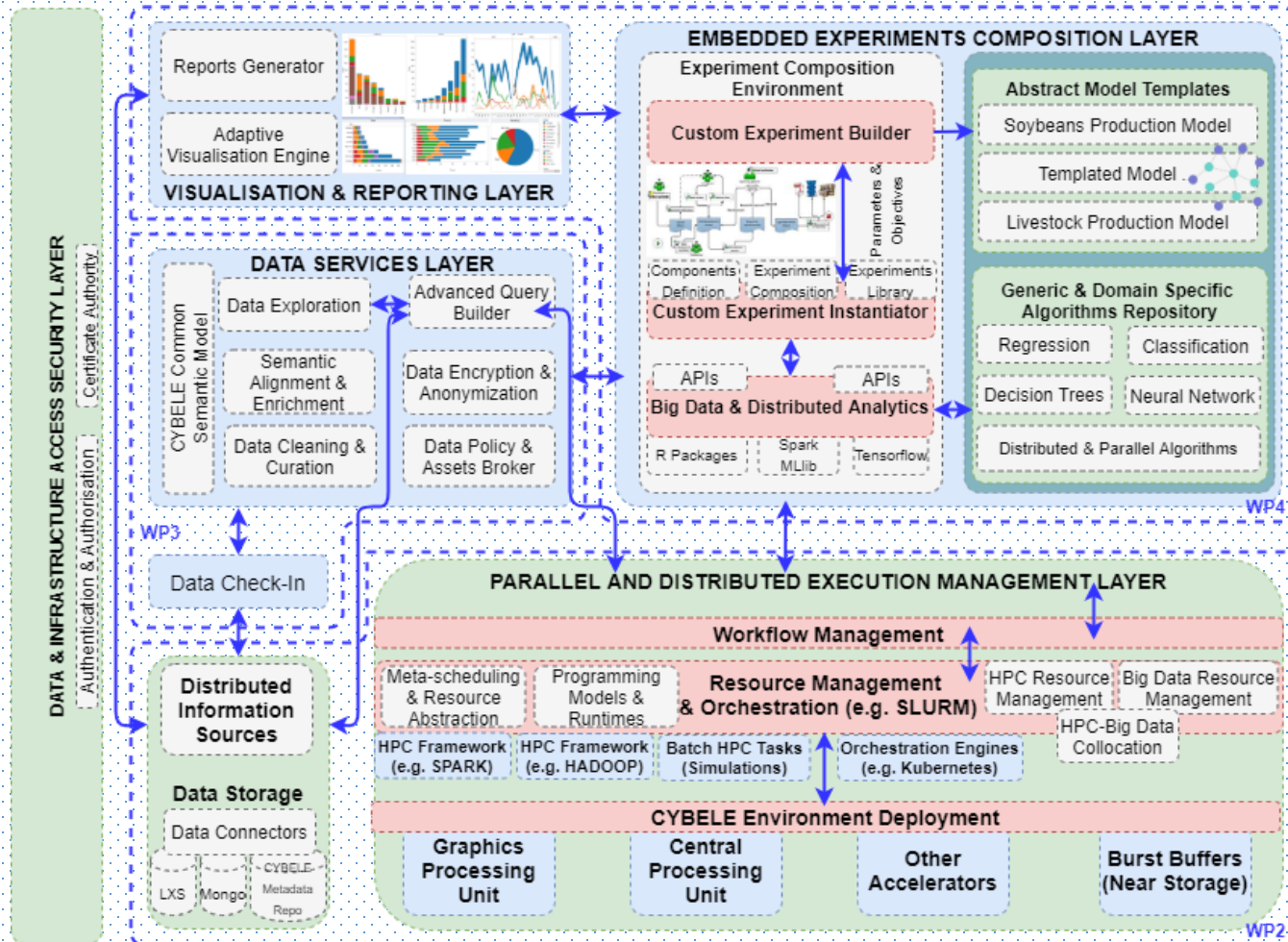
CYBELE Current Status

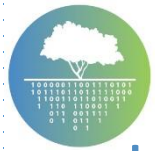
- Harvests huge amounts of images, time-series and textual data to deliver a bouquet of AI-fueled generic and domain specific data analytic applications;
- Provides an HPC-Big Data e-infrastructure with parallel and distributed computing capabilities;
- Builds over big data technologies, distributed machine learning and deep learning methods;
- Creates for re-use common repositories w.r.t. the CYBELE trained models able to be easily onboarded and deployed;
- Delivers a resource abstraction layer translating application level configurations directly to HPC-Big Data workloads;
- Generates innovation and creates value in the field of Precision Agriculture (PA) and Precision Livestock Farming (PLF).





CYBELE Conceptual Architecture

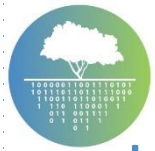




How AI, HPC & Big Data co-exist in CYBELE

- AI, HPC and Big Data convergence lies at several cases of CYBELE ecosystem:
 - **Pilot 1** (organic Soya yield and protein-content prediction): tasks parallelization/execution speed up;
 - **Pilot 2** (food safety), **Pilot 9** (aquaculture monitoring and feeding optimization): hyperparameter tuning adapted for Spark;
 - **Pilot 5** (optimizing computations for crop yield forecasting), **Pilot 8** (open sea fishing): distributed execution over Spark & Big Data partition;
 - **Pilot 4** (autonomous robotic systems within arable frameworks), **Pilot 6** (pig weighing optimization), **Pilot 7** (sustainable pig production): multi-nodes and multi-GPUs deployment by combining PyTorch & MPI;
 - **Pilot 3** (climate services for organic fruit production): parallelisation over HPC partition.

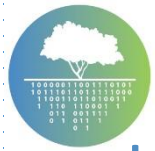




Unique AI, HPC & Big Data needs from the industry

- Huge data volumes collected from geographically distributed locations;
- Added value services for food safety are being developed exploiting distributed deep learning algorithms;
- Need for global and local learning preserving privacy and contributing in advanced decision making at strategic level;
- Need for distributed processing and speed up of time demanding simulations, complex computations, etc.





How CYBELE provides solutions to these challenges

- Seamless HPC resource management over diverse frameworks, systems and testbeds;
- **AI-HPC-Big Data collocation** exploiting Slurm HPC resource manager with Kubernetes enabled Big Data resource manager;
- **Resource abstraction layer** (middleware) leverages and efficiently orchestrates both HPC-Big Data partitions.





CYBELE

www.cybele-project.eu

Thank you!



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.