



Machine Learning as a Service (for the ALICE collaboration)

Sara Vallero
INFN Torino

Introducing myself

Within the ALICE collaboration

- Data analysis
- Detector testing/commissioning
- Data-taking operations

Building competences

- Machine Learning (ML) algorithms
- Distributed ML systems
- Integration of non standard resources (GPUs) with Linux containers and orchestrators (Mesos, OpenNebula)

PRIN Project

(optimization of access to LHC data using the grid and cloud computing)

INDIGO-DC

THIS FELLOWSHIP

2013

2015

March 1, 2018

2007

2012

MD in Physics
(University of Torino)

PhD in Physics
(University of Heidelberg)

Cloud Computing

(full virtualization)

- OpenNebula and KVM
- Contextualization of complex services
- Auto-scaling
- FairShare scheduling

Computing Model as a Service

(lightweight virtualization)

- runtime/application packetization (Docker)
- Distributed scheduling (Mesos and its frameworks)
- Network virtualization (Calico)
- HPC: MPI over InfiniBand in Docker

Batch System as a Service employed at the INFN and UNITO's HPC Cluster (OCCAM)

Machine Learning as a Service



CHALLENGES

- Reconstruction
- Analysis
- Trigger
- Data quality
- Detector monitoring
- Computing operations
- Monte Carlo tuning
- ...

REQUIREMENTS

- Workflow definition
 - Results reproducibility
- Multi-tenancy (scheduling, authentication...)
- Parallel execution and scaling
- Data handling
- Ease of use and management
- ...

IMPLEMENTATION

- Lightweight virtualization
- Modularity
- Flexibility
- Heterogeneous back-end infrastructures
- ...

LEVERAGING

Existing OpenSource software (mature and maintained)

INDIGO-DataCloud products

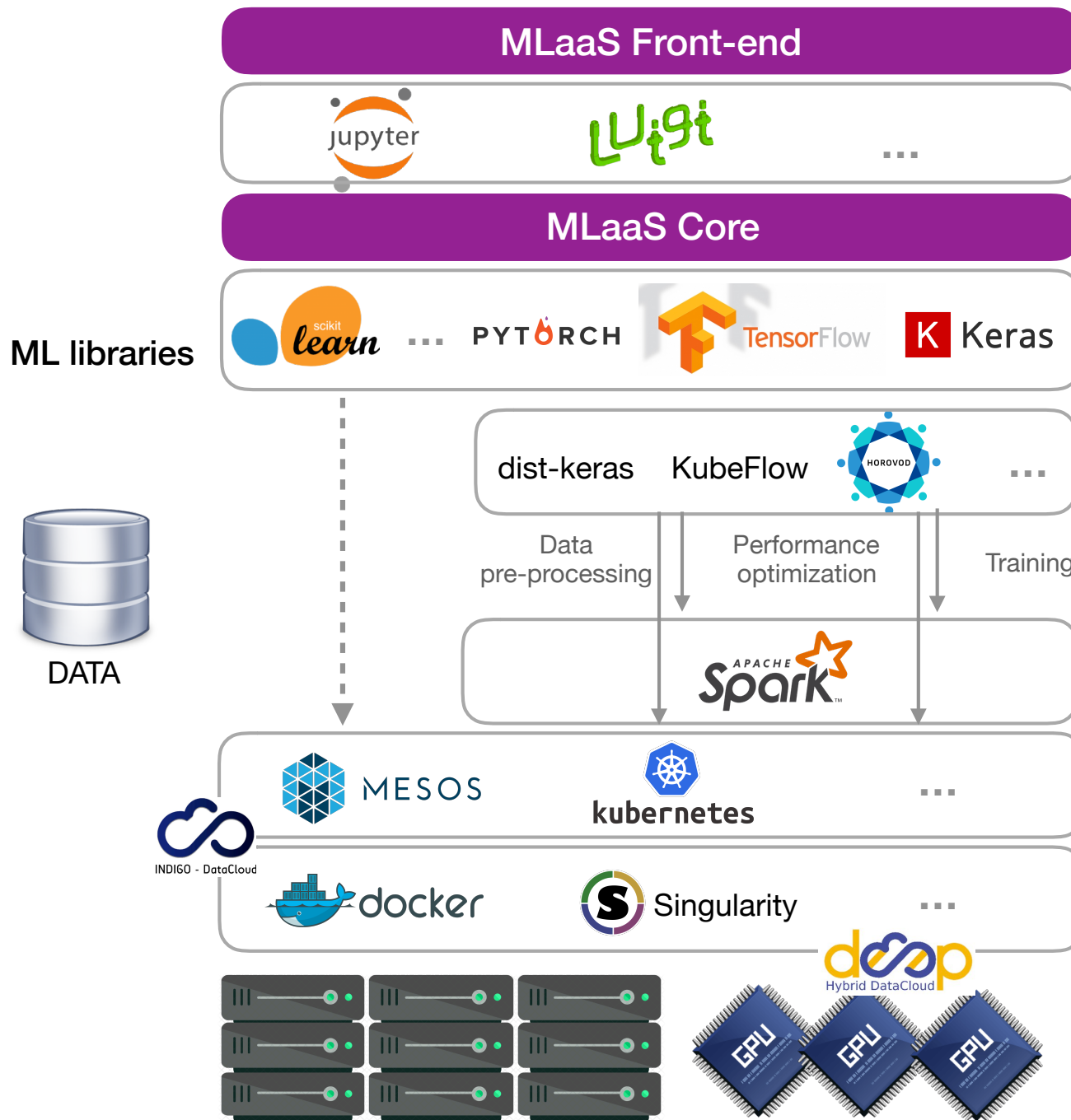


DEEP Hybrid DataCloud products



Brainstorming

↑ Test use-case
Implementation
Technologies
Architecture
Goals



- Workflow definition
- Process Monitoring
- Authentication



Deep Learning framework

Distributed DL libraries

Cluster framework
(parallelize task)

Orchestrator
(schedule on resources)

Packetization and
virtualization

Resources:

- Bare metal
- Infrastructure as a Service

From Data to Results

