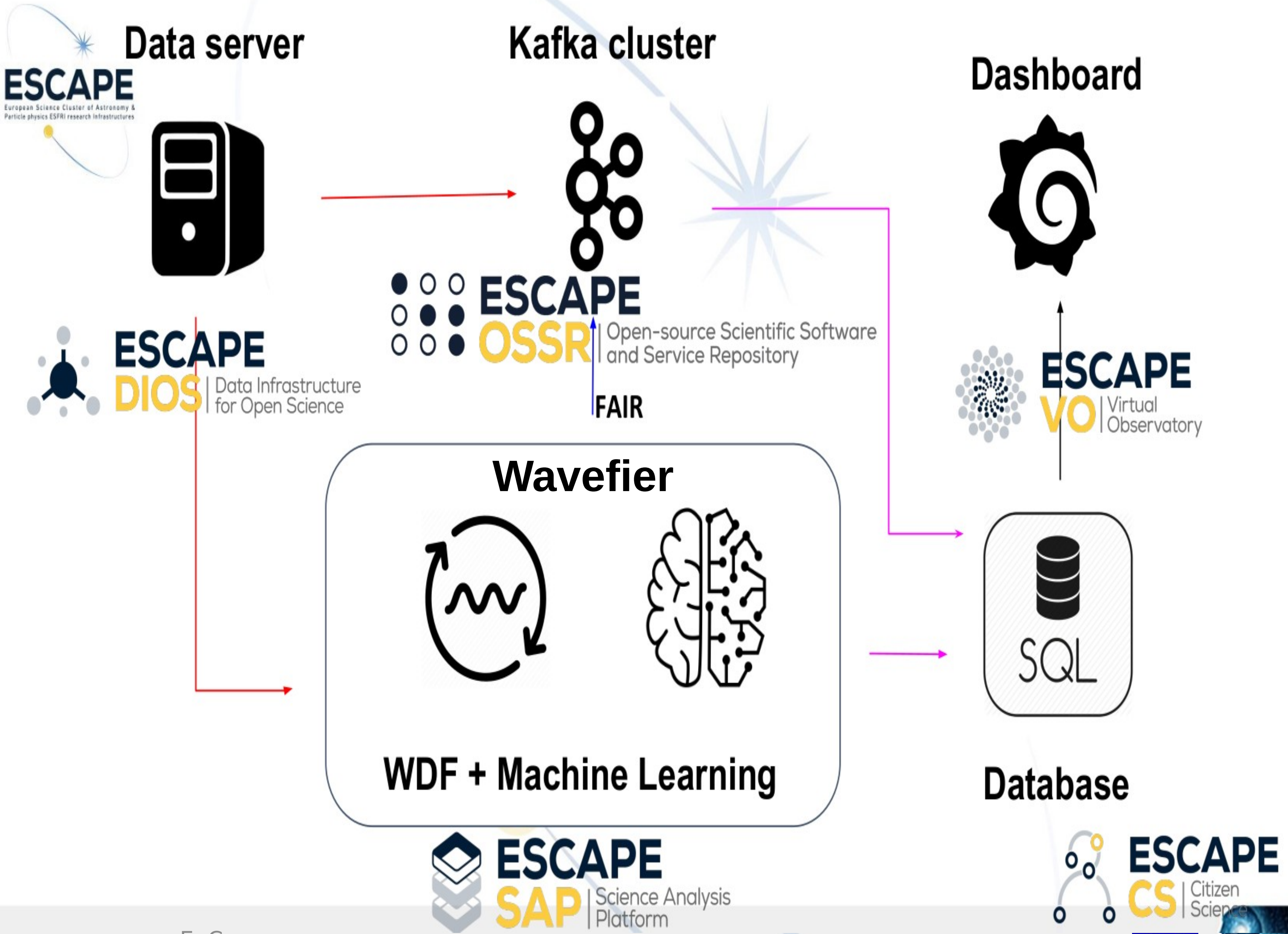
Wavefier status
EGO / Virgo perspective

Pierre Chanial, Sara Vallero, Elena Cuoco, Filip Morawski

Funded by the European Union's
Horizon 2020 - Grant N° 824064



Data server

Kafka cluster

Dashboard

ESCAPE OSSR Open-source Scientific Software and Service Repository

FAIR

ESCAPE DIOS Data Infrastructure for Open Science

ESCAPE VO Virtual Observatory

Wavefier

WDF + Machine Learning

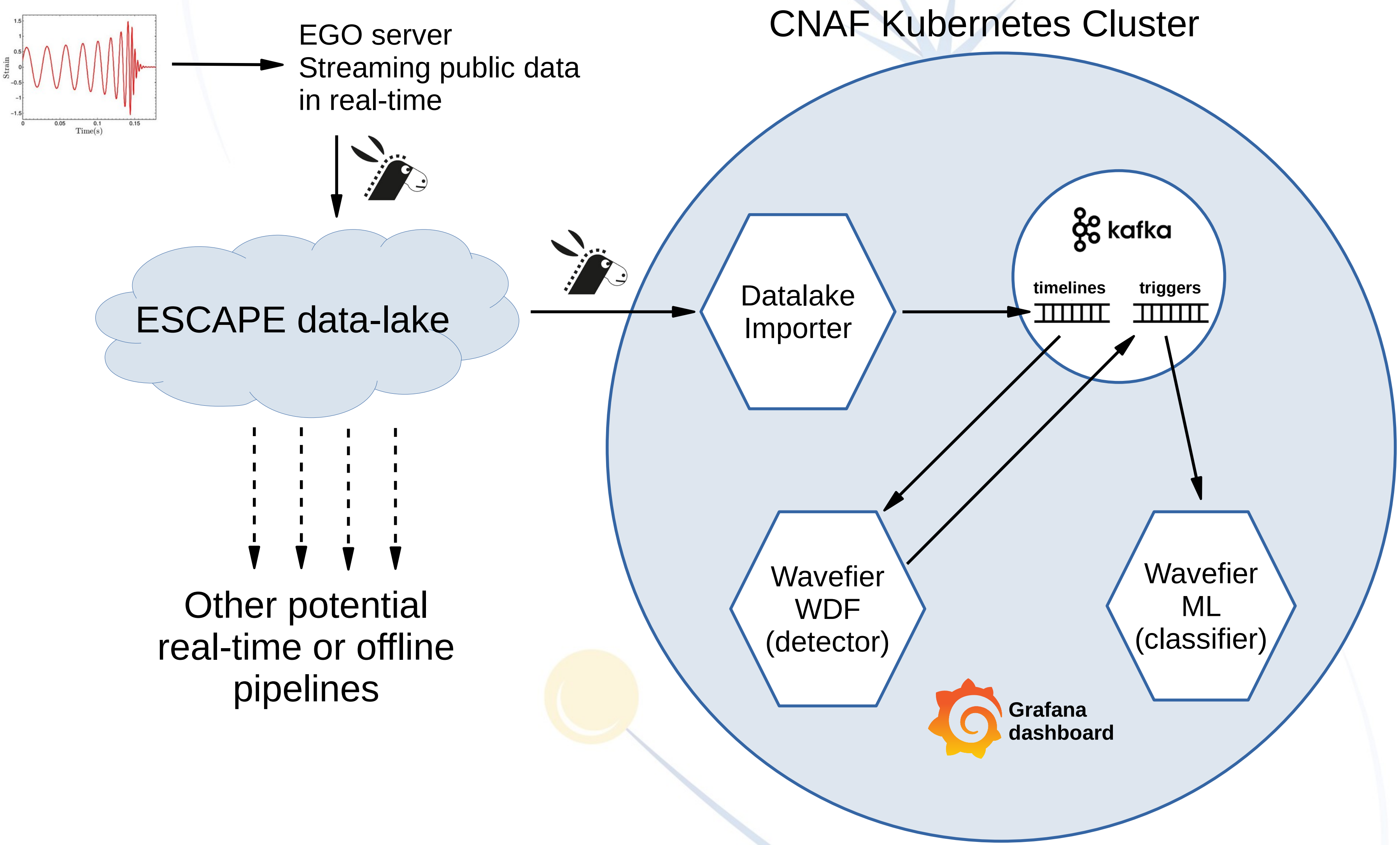
SQL

Database

ESCAPE SAP Science Analysis Platform

ESCAPE CS Citizen Science

Wavefier Online / Offline Architecture



EGO server
Streaming public data
in real-time

CNAF Kubernetes Cluster

ESCAPE data-lake

Datalake Importer

timelines

triggers

Wavefier WDF (detector)

Wavefier ML (classifier)

Grafana dashboard

Other potential real-time or offline pipelines

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

First Data-lake Injection Dress Rehearsal

- EGO : 4h test of upload to and download from the ESCAPE data-lake prototype
  - ➔ real-time
  - ➔ public Virgo h(t) drawn randomly from an O2 h5f file
    - O2 llhoft have not been made public
    - ESCAPE data-lake prototype not yet secured for proprietary data
  - ➔ chunks : 1 second, 4kHz
  - ➔ Data rate : 85kB / s

- Goal is to test functionalities, not yet performances. Latencies have been measured, but they should be taken as the baseline we will improve on, not definitive numbers.

- Uploader: Celery application to pace the uploads
  + Rucio Python non-docker client

- Downloader: Multiprocessing Python
  + Rucio Python docker client

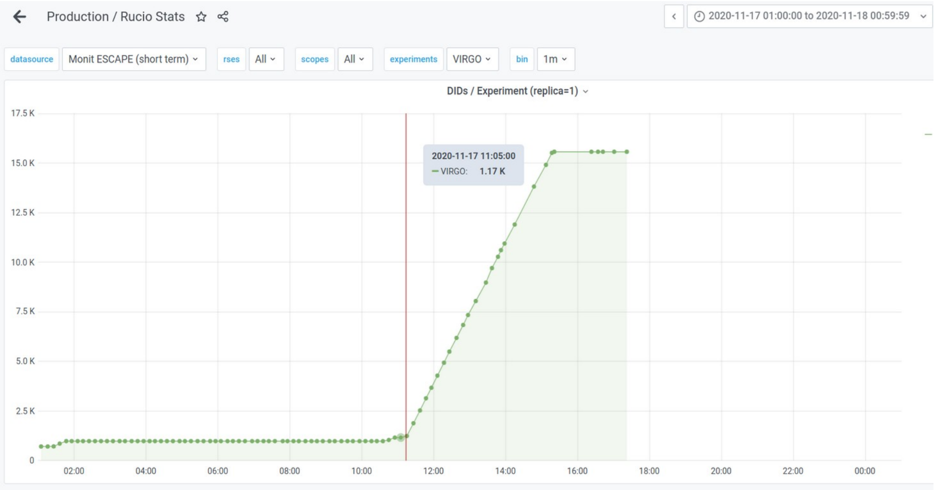
  Layman's approach : download the dataset content metadata at regular intervals to poll new entries in the dataset

Funded by the European Union's
Horizon 2020 - Grant N° 824064

First Data-lake Injection Dress Rehearsal

- Great success

- The 4 x 3600 = 14400 data chunks have been sent
    - All samples uploaded
    - All samples downloaded
    - None corrupted



Production / Rucio Stats

2020-11-17 01:00:00 to 2020-11-18 00:59:59

datasource: Monit ESCAPE (short term) · rses: All · scopes: All · experiments: VIRGO · bin: 1m

DIDs / Experiment (replica=1)

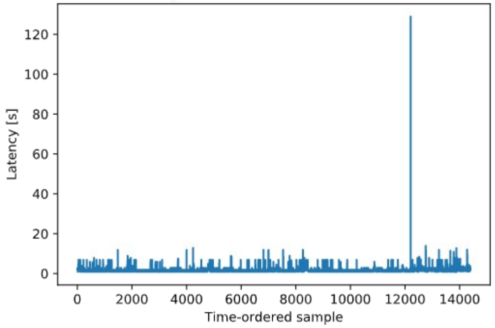
2020-11-17 11:05:00

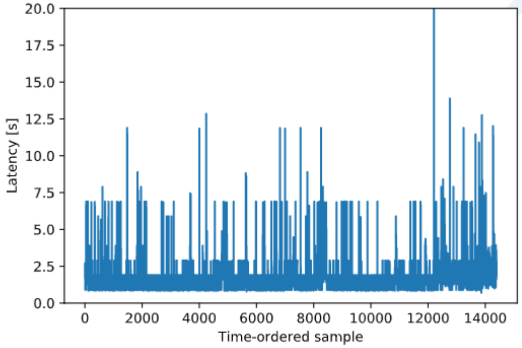
VIRGO: 1.17 K

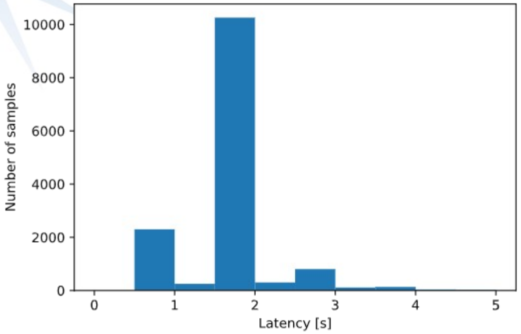
Funded by the European Union's
Horizon 2020 - Grant N° 824064

First Data-lake Injection Dress Rehearsal

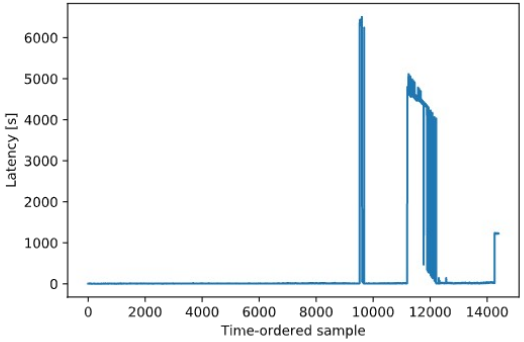
- Upload (EGO ➡ CERN) latency analysis (mean : $1.9\text{s}_{-0.5}^{+0.3}$ s)

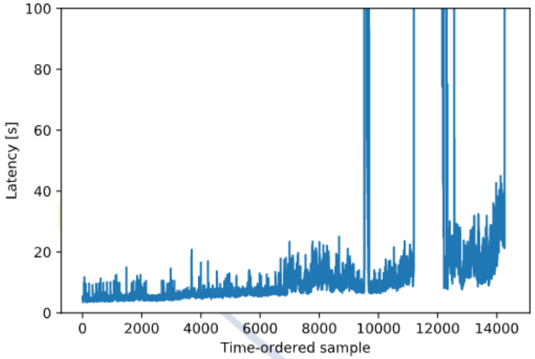


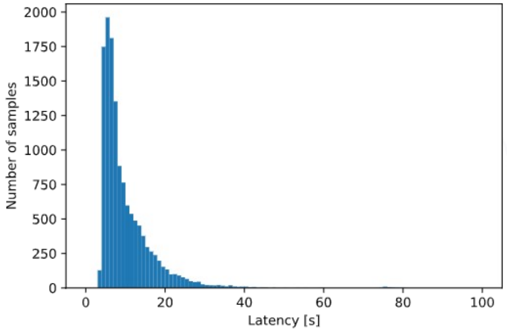




- Total (EGO ➡ CERN ➡ CNAF) median : 8.2 s







Funded by the European Union's
Horizon 2020 - Grant N° 824064

Status & Plans

- The quest for GPUs
  - CNAF local k8s cluster has no GPU
  - INFN-ML (6 Tesla T4) still waiting for an account
  - In the future, we may use CNAF Corporate Cloud
- Source of GW data :
  - GW interformeters are not in science mode : upgrade for O4 (not before June 2022)
  - Data Replay stream is being put in place (ETA : couple of weeks)
  - But even if some of the raw data is public, the processed data is not.
  - Create a simulated stream that will be integrated with ESCAPE datalake
  - Use Replay data stream @ CNAF using private storage.

Funded by the European Union's
Horizon 2020 - Grant N° 824064