3rd ASTERICS-OBELICS Workshop

23-25 October 2018, Cambridge, UK.

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Real-Time Data Streaming Architecture Survey

C. Bigongiari & L.A. Antonelli
INAF-OAR
on behalf WP 3

Introduction

ASTERICS will benefit ESFRI projects and other related major research infrastructures, including ESFRI-precursor experiments, enabling interoperability, software re-use and the use of open standards and software libraries.

The development of common solutions is a key point of this process.

We present a survey focused on the architectures for the real-time data streaming, already applied or envisaged by the ESFRI projects, seeking for synergies and common developments.

Project	ESFRI	Field	Type of data
CTA	Yes	Cherenkov observatories for gamma-ray astronomy	Events
H.E.S.S.	Pathfinder for CTA	Cherenkov telescope array for gamma-ray astronomy	Events
MAGIC	Pathfinder for CTA	Cherenkov telescope array for gamma-ray astronomy	Events
KM3NeT	Yes	Neutrino telescope	Events
IceCube	Pathfinder for KM3NeT	Neutrino telescope	Events
ANTARES	Pathfinder for KM3NeT	Neutrino telescope	Events
ELT	Yes	Ground-based optical/near-infrared telescope	Images
LSST	No	Optical telescope	Images
EUCLID	No	Satellite mission to map the dark Universe	Images
SKA	Yes	Radio telescope arrays	Signals
e-EVN	Pathfinder for SKA	e-VLBI network for radio astronomy	Signals
LOFAR	Pathfinder for SKA	Radio interferometric array	Signals
Advanced LIGO	No	Gravitational wave detectors	Signals
Advanced Virgo	No	Gravitational wave detector	Signals

Experiments are classified according to the type of data that they produce. Although final science products are similar for all the Astro(particle)physics experiments (e.g., skymaps, catalogues, spectra), raw and processed data below the science data level have fundamental differences depending on the experimental technique.

Image-based experiments

- **LSST** observing program will produce approximately 15 TB of raw imaging data and millions of alerts every night. Alerts will be delivered within 60''.
- **ELT** requirements is that the raw scientific data, including calibrations, shall arrive at the ESO Science Archive Facility not later than 1 h after they have been acquired. The average foreseen nightly data production is at the level of 1–2 TB (uncompressed) per night, with large variations depending on the instruments and modes actually used on a given night.
- **Euclid** mission will deliver a lower data rate but still unprecedented for a space mission: about 110 GB of compressed data per day. On-board data processing is required to reduce the data stream generated by the 4 Megapixel detectors by a factor over 100, since it is impossible to deliver to the ground all the raw detector data.

Event-based experiments

- HESS: The data rates of cameras peak at 46MB/s for the primary scientific data during routine operation. The required maximum data rate is of the order of 80MB/s for fast transients.
- MAGIC: raw data files have a typical size of 2 GByte while the data volume nightly generated and processed in real time at a 600Hz acquisition rate might exceed 1 TByte.
- ASTRI: the expected over-all stream data rate for the typical event trigger rate of 600 kHz is of the order of 6-7 MB/sec.
- CTA: camera readout, the buffering of the read-out data, the processing of array trigger decisions, the building of camera-dependent events and filtering of interesting events to reduce the overall data volume (about 20 PB/y)
- Antares & KM3Net: real-time processing will be performed for all the triggers passing a certain threshold.

Signal-based experiments

- EVN: European VLBI Network has been designed to observe at a bandwidth of 128 MHz, resulting in data flows of up to 1 Gbps per telescope.
- LOFAR: about 19 GB/s for the entire array) are sent via a high-speed fibre network to the central processing facility, where they are pre-processed on-line.
- SKA: For the first phase, the summed data rate is estimated to be about 3 TB/s and the expected archive data volume is 100 PB per year
- LIGO & VIRGO: Present data rates exceed 5 Mbytes per second per interferometer continuous. Enhanced experiments will produce a raw-data stream of about 2 TB/day.

Conclusion

- The demand for real-time or nearly real-time data streaming is increasing rapidly in many scientific areas for different reasons. This is particularly true in the area of astronomy, astrophysics and astroparticle physics.
- Similar architectural solutions are usually shared by experiments that produce the same kind of data.

Synergies & Commonalities

- Highly parallelized calibration and analysis based on GPUs and other hardware accelerators could be adopted by both classes of experiments. **(HW solutions)**
- A very attractive perspective is about the usage of deep learning techniques in the reconstruction process. **(SW solutions)**

Acknowledgement

- H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).