Development of Big Data framework for Cherenkov Telescope Array

The Real-Time Analysis of CTA

N. Parmiggiani, A. Bulgarelli, L. Baroncelli, S. Tampieri, D. Beneventano, G. Zollino (INAF/OAS Bologna)

INAF OAS BOLOGNA

Why Big Data?

- 5 V of Big Data: Velocity, Volume, Variety, Veracity and Value
- New generation of observatories leads to Big Data
- Key technologies for feature observatories like SKA and CTA

SKA
SQUARE KILOMETRE ARRAY

Raw Data: 16 TB/s
Archive Data: 600 PB/y
Real-time: 5 TB/s

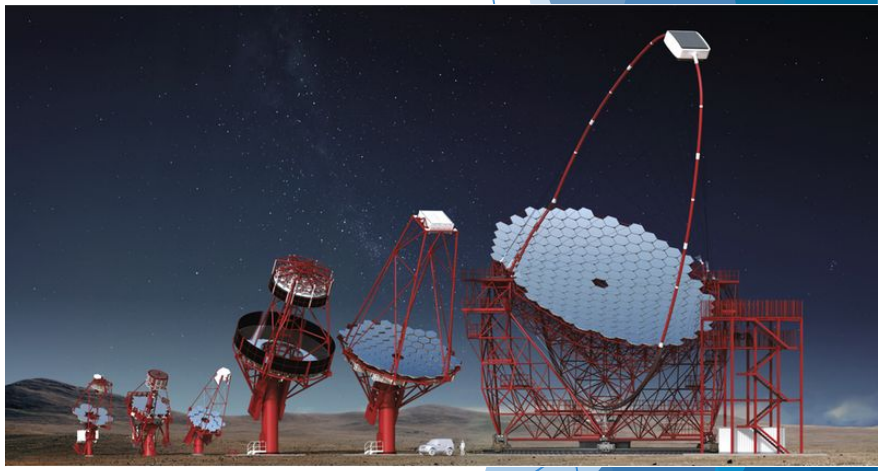
cta
cherenkov telescope array

Raw Data: 1000 PB/y
Archive Data: 3 PB/y
Real-time: 1-5 GB/s

INAF OAS
BOLOGNA

Cherenkov Telescope Array

- More than 100 telescopes located in the northern and southern hemispheres
- Three different telescopes: LST 23m, MST 11.5m and SST ~4m
- CTA consortium with 1420 members from >200 institutes in 31 countries
- CTA will be the world's largest and most sensitive high-energy gamma-ray observatory

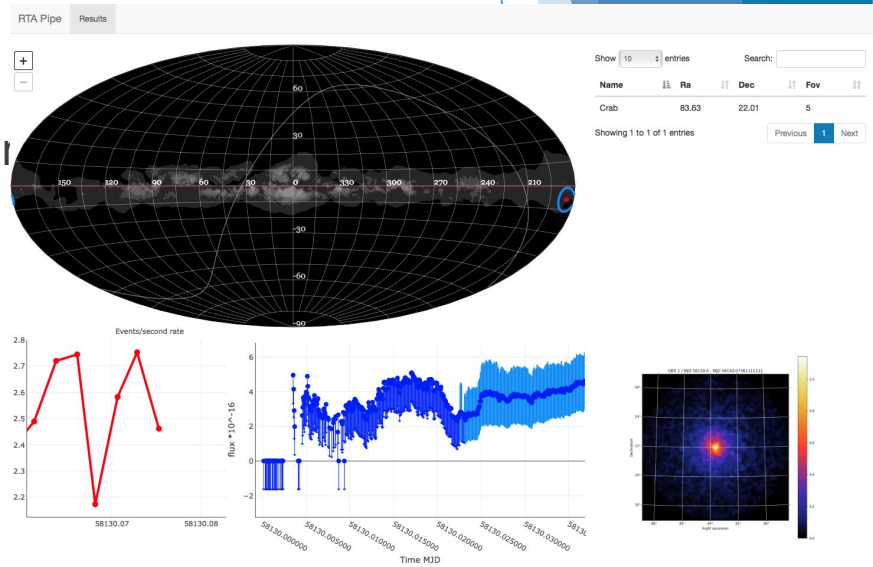


CTA Real-Time Analysis requirements

- At INAF/OAS Bologna we are developing several RTA prototypes of the Real-Time Analysis (ASTRI, LST1, CTA)
- Scientific analysis in real-time during the observation.
- Data rate of 2-5 GB/s
- On-site with telescopes (limited hardware)
- Generate Science Alerts with a latency of 30s
- Run multiple analysis in parallel to manage sub-array, different time scales or different analysis tools

RTA-SCI and RTA-GUI

- RTA-SCI is currently used and developed by the AGILE team for GW automatic response
- The RTA-SCI has different inputs and produces different scientific output: Light Curves, Counts Maps, Alerts, Detections ar TS Maps.
- The RTA-SCI runs multiple analysis in parallel: for this reason it requires a framework for scalability and flexibility
- RTA-GUI must provide fast updates for different simultaneous events



RTA-RECO and RTA-DQ

- During the reconstruction and the data quality analysis we have high data volume and data rate (kHz/s, GB/s).
- The data output from different reconstruction steps must be stored or buffered to run data quality analysis
- All the process should be done in-memory to reduce IO usage and improve the performance
- The data reduction streaming process can be done in parallel in order to use more computing power

Big Data for CTA @ OAS-Bologna

INAF OAS BOLOGNA

- CTA Real-Time Analysis requires Big Data technologies because the date have the 5 V of Big Data definition
- We are studying different technologies to develop a Big Data framework for data management and temporary storage
- We are testing an easy way to deploy these technologies in different contexts

Tools and Technologies

Relational Database

In-Memory non relational database

Cluster Computing Framework

Cluster workload manager

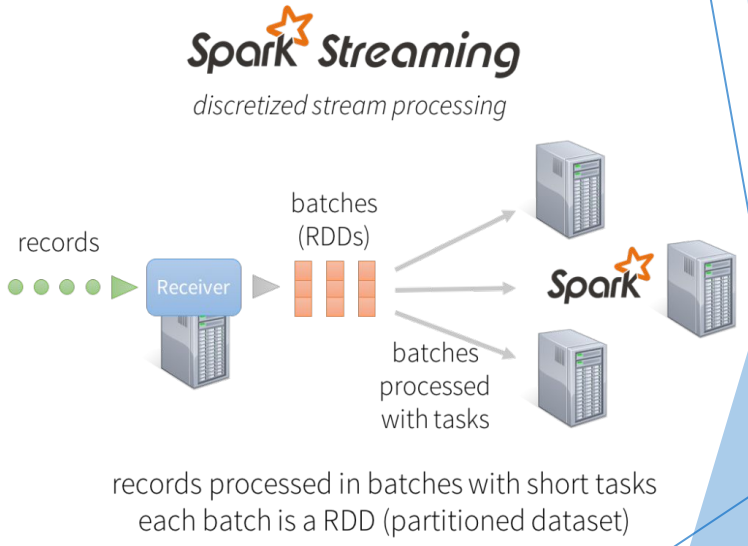
Container management

Handling real time data feeds

Spark - Cluster Computing Framework

Key features:

- Streaming application
- Parallel computing
- Scalability, Flexibility
- In-memory analysis
- Failure management
- Python development
- Web GUI monitoring



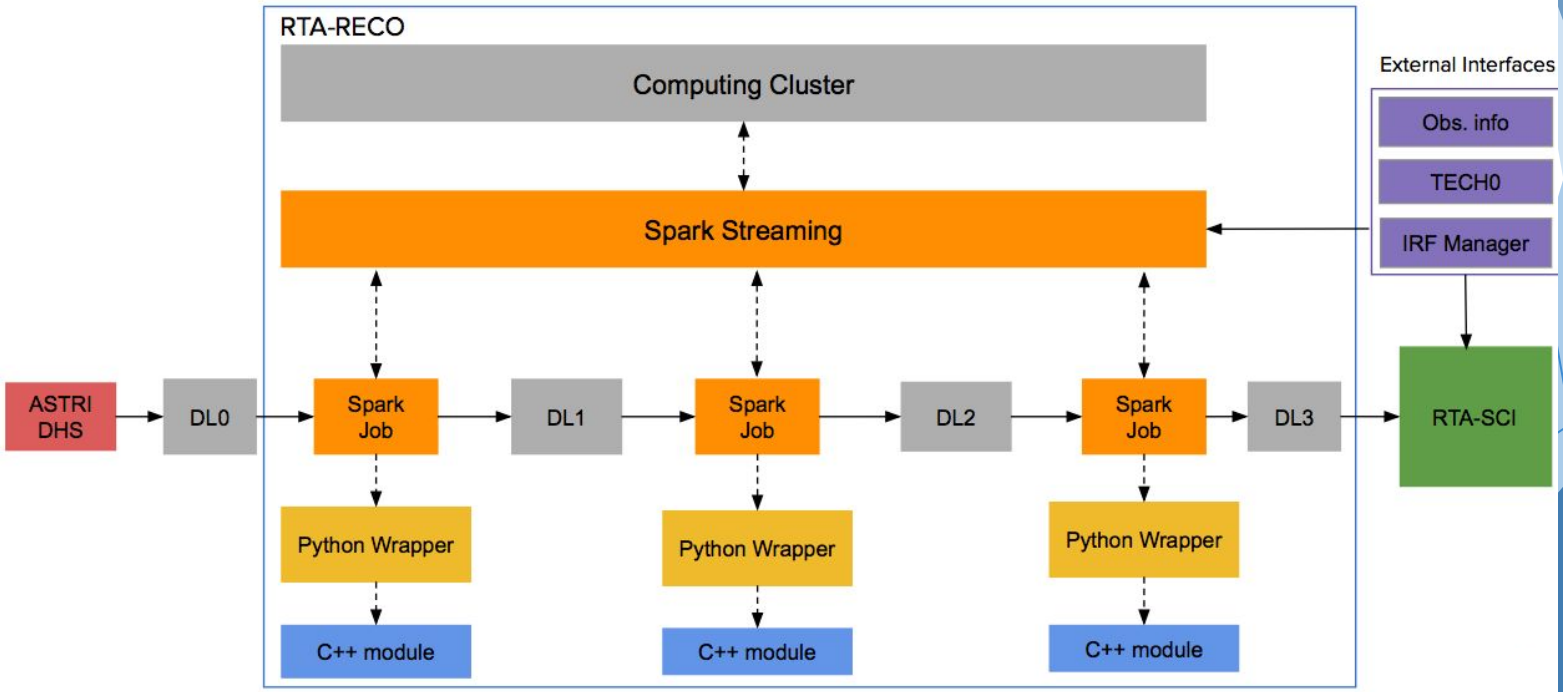
Singularity – containers manager

- All services (MySQL, slurm, Spark, Kafka and other) and all analysis software are installed inside a singularity container
- This container can be deployed on many different hardware machines and allow high scalability and flexibility
- Using singularity container we are able to use a process of continuous integration with Jenkins

INAF OAS BOLOGNA

Real-time Analysis for ASTRI

We are collaborating with INAF/OAR Roma ASTRI team to develop the ASTRI RTA



Real Time Analysis for LST1

- A collaboration with LST1 team is on going to implement the full LST1 RTA
- We tested our RTA singularity environment inside the La Palma IT cluster
- We tested and improved the cta-lstchain reconstruction performance using Spark and Kafka. More than 2000 Hz on a single machine.

Conclusions

- The need of Big Data technologies will increase with the future observatories
- We are developing a Big Data framework for the CTA RTA usin open source technologies.
- This framework can be easy deployed in different environment and context using singularity containers

INAF OAS BOLOGNA