Automated scientific analysis pipeline and data quality for ASTRI Mini-Array

Nicolò Parmiggiani - INAF/OAS Bologna

for the ASTRI Project

OAS VHE Meeting - 08/06/2022

Summary

ASTRI Mini-Array

- This talk presents two software system of the ASTRI Mini-Array project:

    a. The Online Observation Quality system will be deployed in the on-site data center to perform the data quality checks during the observations.

    b. The Automated Scientific Analysis Pipeline will be deployed in the off-site data center (Rome) to execute two types of analyses:

        i. short-term analysis for quick-look scientific results during the observations

        ii. long-term analysis to produce high-level scientific results

Online Observation Quality System

OOQS Overview

ASTRI Mini-Array

1. The OOQS is part of the SCADA software system deployed in the on-site data center that manages the startup, shutdown, configuration, and control of all site assemblies and sub-systems.

2. The OOQS aims to execute data quality checks on the data acquired in real-time by the Cherenkov camera and intensity interferometry instruments deployed in the nine ASTRI Mini-Array telescopes.

3. The OOQS is designed to manage the high data rate generated by the instruments (up to 4.5 GB/s for the intensity interferometry observations) and the Cherenkov event rate of 1000 Hz.

4. We defined the Use Cases and the Software Requirements considering the experience and the know-how acquired in other projects: the ASTRI Horn prototype, and the AGILE space mission.

5. During the definition of the use cases, we collaborated with domain experts to collect information on the quality checks that the OOQS shall perform to evaluate the data acquired by the telescopes.

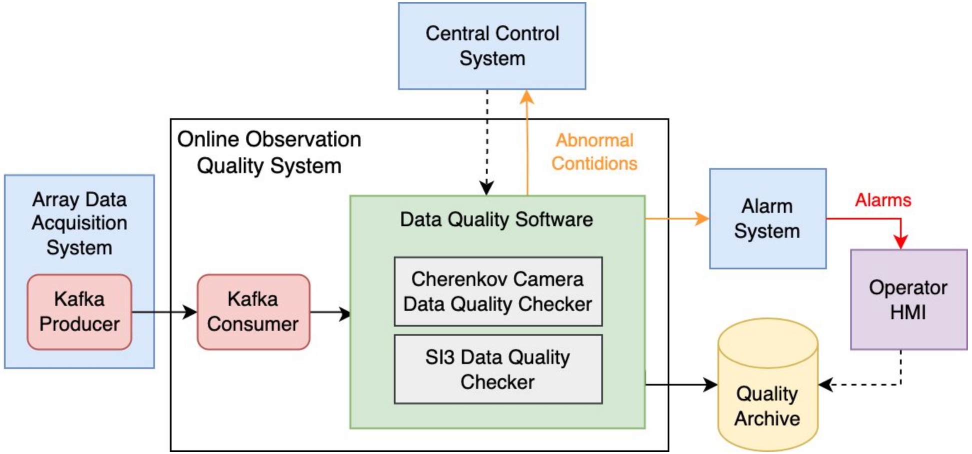
OOQS System Context

Mini-Array

1. The Central Control System manages the OOQS that receives input data from the Array Data Acquisition System through the Kafka service

2. The results from the data quality analyses are stored in the Quality Archive for further investigation. The operator can visualize them using the Human Machine Interface.

3. When the OOQS finds abnormal conditions, it sends a notification to the Central Control System and the Alarm System, triggering their reaction to promptly correct anomalies.



OOQS key technologies

1. Apache Kafka is an open-source distributed event streaming platform to implement high-performance and mission-critical applications.

2. Kafka can implement the publisher/subscriber pattern to send messages between sub-systems.

3. The OOQS software and dependencies (e.g., external packages and services) are installed in a Docker container.

4. Docker allows the developer to package and run an application in an isolated and reproducible environment.

5. Finally, we developed a continuous integration pipeline to support the development process that uses Docker containers to build, run and test the code.

OOQS key technologies - Slurm

ASTRI Mini-Array

1. We decided to manage the execution of the data quality checks with the Slurm workload manager.

2. Slurm can schedule and execute multiple processes in parallel to optimize the resources.

3. Slurm can run in a cluster composed of several workers and one controller to balance the job execution and manage the failures of the workers.

4. Slurm is fault-tolerant. It can automatically replace failed jobs with new ones.

5. Slurm can also manage the priority between processes. When a high-priority job is scheduled and the resources are not enough, Slurm suspends the low-priority job to execute the high-priority one. When the high-priority job ends, Slurm resumes the low-priority job.

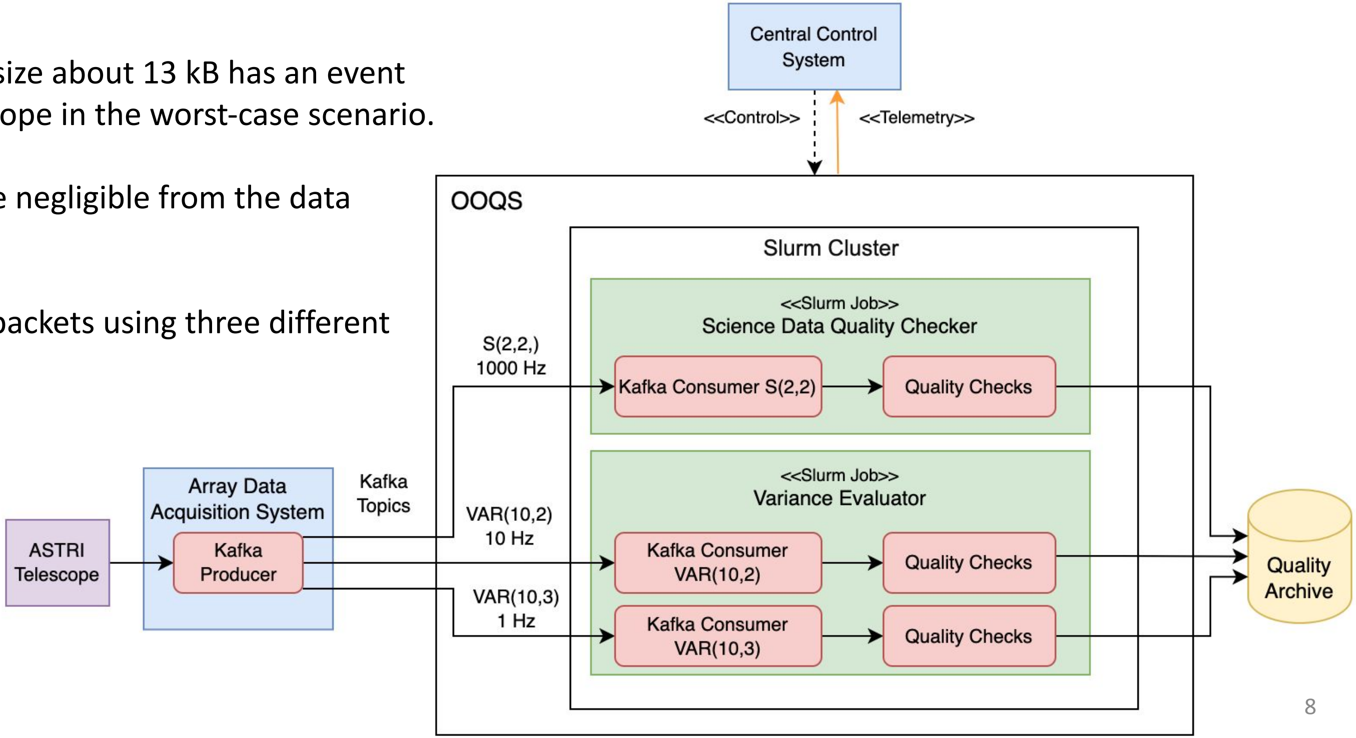
OOQS for Chrenkov Observations

1. The OOQS analyzes three packet types during Cherenkov Observations: S(2,2), VAR(10,2) and VAR(10,3)

2. The S(2,2) packet with a size about 13 kB has an event rate of 1000 Hz per telescope in the worst-case scenario.

3. The other two packets are negligible from the data throughput point of view.

4. The ADAS transfer these packets using three different Kafka topics



OOQS for Intensity Interferometry Observations

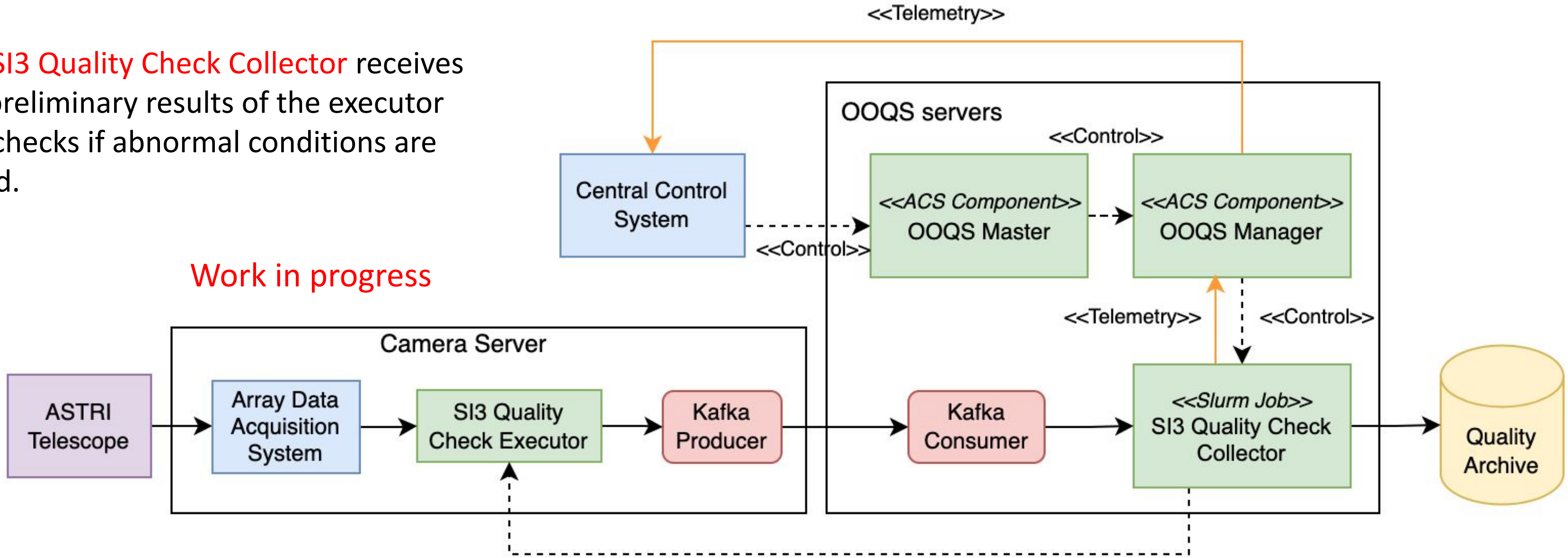
Mini-Array

1. During intensity interferometry observations, each telescope produces 100 Mevents/second with a data rate of 500 MB/s. The data rate for the full array of nine telescopes is about 4.5 GB/s

2. The SI3 Quality Check Executor performs the data quality checks on the SI3 data where they are acquired to avoid data transfer.

3. The SI3 Quality Check Collector receives the preliminary results of the executor and checks if abnormal conditions are found.

Work in progress



Conclusions

ASTRI Mini-Array

1. The OOQS is the software system that performs data quality checks in real-time on the data acquired by the nine telescopes of the ASTRI Mini-Array project during Cherenkov and intensity interferometry observations.

2. The ADAS sends the data acquired by the nine telescopes to the OOQS through Kafka.

3. The OOQS is a critical system because it detects abnormal conditions and sends notifications to the Central Control System and the Alarm System for a fast reaction.

4. In additions, the Operator visualizes the OOQS results stored in the Quality Archive through the Operator HMI during the supervision of the observations and takes corrective actions if needed.

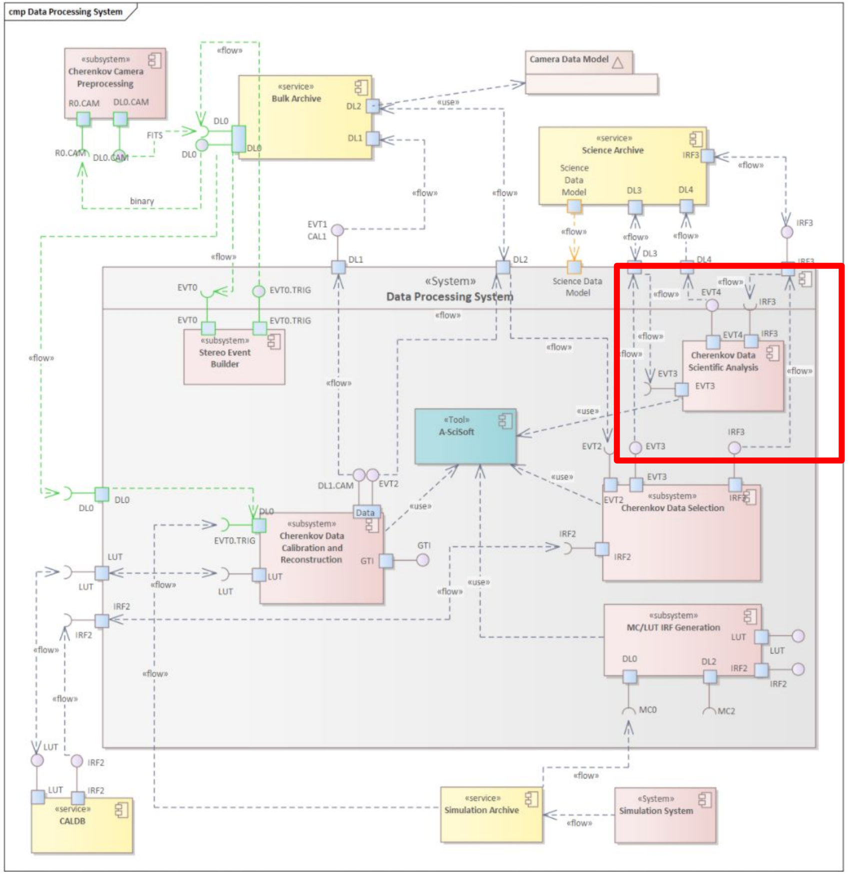
- Parmiggiani, N. et al., “The Online Observation Quality System Software Architecture for the ASTRI Mini-Array Project”, Proc. SPIE 2022
- Parmiggiani, N. et al., “The Online Observation Quality System for the ASTRI Mini Array.”, ICRC 2021, DOI: 10.22323/1.395.0692

Automated Scientific Analysis Pipeline

System Context Overview

ASTRI Mini-Array

- The Automated Scientific Analysis Pipeline of the ASTRI Mini-Array will be deployed in the off-site data center in Rome.
- It is part of the Data Processing System that manages the automated analyses starting from the raw data received from the Array Data Acquisition system on-site.
- This pipeline executes scientific analysis starting from the event list (EVT3) and the IRF3 generated by the automated reconstruction pipeline.
- The results of the scientific analyses are stored in the Science Archive



Pipeline Use Cases

ASTRI Mini-Array

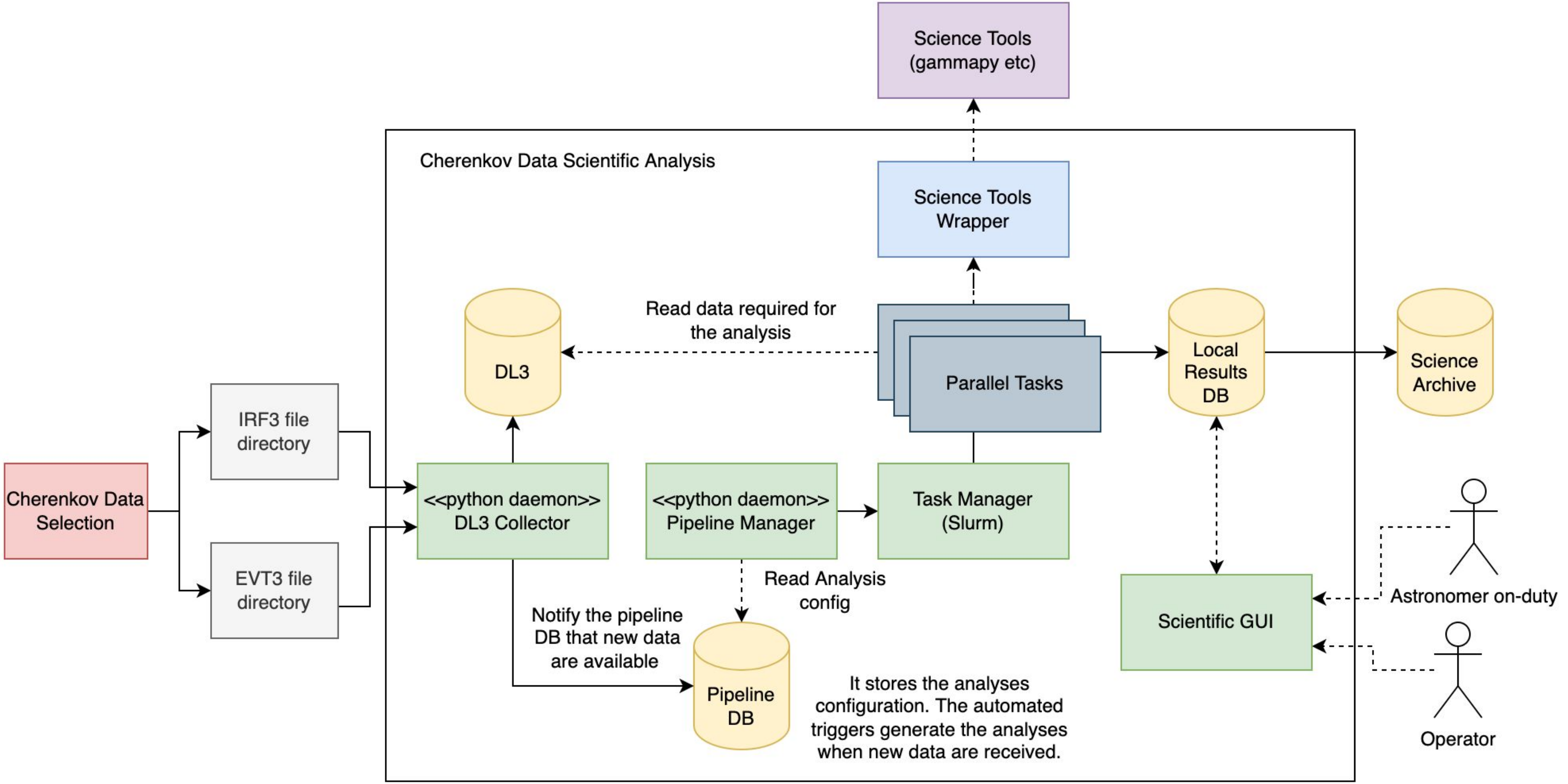
- The main goals of this system are to provide a quick-look scientific analysis results during the data acquisition and to produce high-level scientific results in an automated way.

- The pipeline must satisfy two main operating modes:

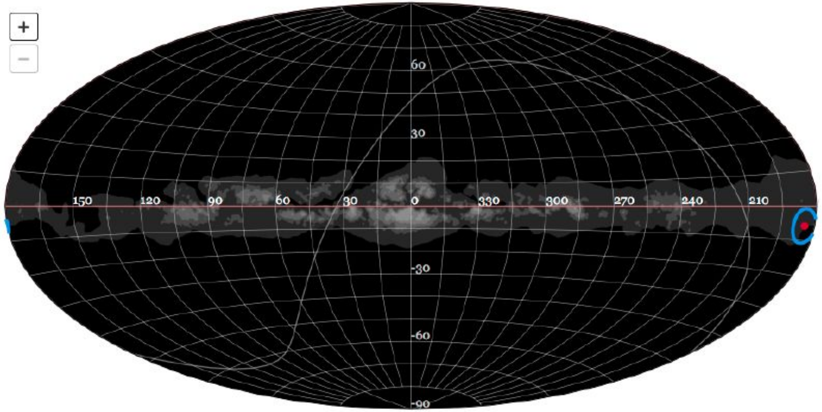
  - short-term analysis: executed automatically as soon as the event list is received (~ 20 min since data acquisition) and reduced at the data center (~10 min since data receiving) to generate scientific quick-look results that can be visualized using the web Scientific GUI by the Astronomer on-duty.

  - long-term analysis: executed without time constraints and based on the best available calibration factors. It generates more accurate scientific results that will be stored in the Science Archive and available to scientific users.

Pipeline software architecture



Graphical User Interface

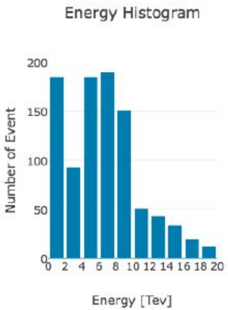


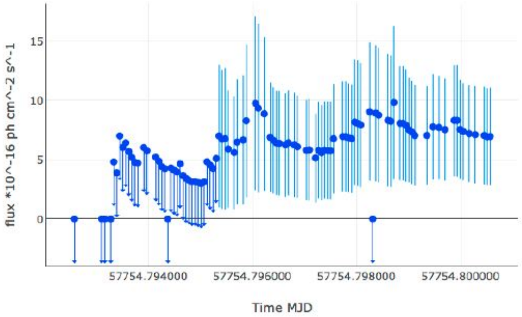
Show 10 entries Search:

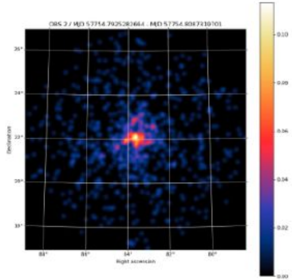
Name	Ra	Dec	Fov
Crab-astripipe	83.63	22.01	5

Showing 1 to 1 of 1 entries

Previous 1 Next







RTApipe framework

Mini-Array

- We developed the RTApipe framework to help the development of real-time analysis pipeline in the context of ground and space-based gamma-ray facilities
- This framework is designed for the multi-messenger and multi-wavelenght context
- This framework satisfies two main use cases:
    - execute periodical scientific analysis, searching for transient events and promptly generate science alerts to share the information with the community;
    - react to external science alerts and start scientific analysis on available data or wait until the required data are available.

Parmiggiani, N. et al., “The RTApipe framework for the gamma-ray real-time analysis software development”, Astronomy and Computing, Volume 39, April 2022. DOI: https://doi.org/10.1016/j.ascom.2022.100570

RTApipe framework main features

1. **Parallel analysis**: pipelines developed with the RTApipe framework can perform many parallel analyses using the Slurm workload manager (job scheduler).

2. **Flexibility**: pipelines can be configured to use different Science Tools that can be configured in several ways.

3. **Scalability**: Slurm can be configured in a single machine or in a cluster of several machines.

4. **Task Priority**: it is possible to configure the priority between different types of analysis.

5. **Pipeline Monitoring**: all the information about the analyses are stored into a database to obtain a real-time monitoring.

6. **Easy deployment**: all the services needed by the pipelines are installed and configured inside a Singularity or Docker container.

AGILE Know-How

Mini-Array

1. The RTApipe framework has been used since 2017 to implement several AGILE real-time scientific analysis pipelines.

2. These pipelines react to external science alerts executing automatically more than 100 analyses in parallel. In addition these pipelines analyze the data acquired by the AGILE instruments as soon as they are available.

3. The AGILE RTA is able to send automated notices to the GCN network when a GRB is detected inside the AGILE data. Since 2019, it has sent more than 70 notices to the GCN network without human intervention.

4. The AGILE researchers use a web Graphical User Interface during their daily work to monitor the pipeline results and follow-up external science alerts.

Other usage of the RTApipe

ASTRI Mini-Array

1. We are using the RTApipe framework to develop the automated scientific analysis pipeline for the CTA project (Andrea's talk). The RTApipe framework can satisfy the demanding requirements of the Science Alert Generation system of CTA, such as:

    a. perform analyses with a latency of 5 seconds;

    b. manage multiple observations in parallel;

    c. detect candidate science alerts (e.g GRBs).

2. We are using the framework to implement the RTA pipeline for the LST1 telescope (Ambra's talk)

3. The RTApipe was used to implement a prototype for the scientific analysis pipeline of the ASTRI Horn telescope. This implementation proves the capability of this framework to implement automated scientific pipelines for Cherenkov telescopes.

Conclusions

Mini-Array

1. The Automated Scientific Analysis Pipeline aims to provide quick-look scientific results during the data acquisition (e.g., counts maps and light curves) and to generate high-level accurate scientific results.

2. The pipeline executes two types of analyses: a short-term analysis as soon as the event list is received from the reduction pipeline and a long-term analysis executed without time constraints and using the best available calibration factors.

3. The scientific results generated by this pipeline are stored in the Science Archive. The results of the long-term analyses are available for scientific users.

4. The Astronomer on-duty can visualize the short-term results during the observations through a web Graphical User Interface.

5. We started to prepare the Use Case Document and to collect information from the ASTRI Science Tools team to complete the design of this software system.