Big Data Architectures for Logging and Monitoring Large Scale Telescope Arrays

Eva Sciacca, A. Costa, G. Tosti, J. Schwarz, P. Bruno, A. Calanducci,
A. Grillo, F. Vitello, U. Becciani, S. Riggi, V. Conforti, F. Gianotti

ISTITUTO NAZIONALE DI ASTROFISICA INAF

UNIVERSITÀ DEGLI STUDI DI PERUGIA

This work was conducted in the context of the ASTRI Project

- Background
- ASTRI and the ASTRI Mini-Array
- IoT and Big Data
- IoT Software Architecture
- Data Workflow
- Cloud Deployment
- Conclusions and Future Works

Background

Large volumes of **technical** and **logging data** result from the **operation** of **large scale astrophysical infrastructures**.

In the last few years several **"Big Data" technologies** have been developed to **deal** with a **huge amount of data**, e.g. in the **Internet of Things** (IoT) framework.

We are testing different stacks of **Big Data/IoT architectures** with the objective to have a **system** that can be updated, maintained and customized with a **minimal programming effort**.

We consider the end-to-end **mini-array** of **ASTRI telescopes** that are under **development by INAF** proposed to constitute a **pathfinder array** for the **Cherenkov Astronomy** for high energy ranges.

ASTRI (**A**strofisica con **S**pecchi a **T**ecnologia **R**eplicante **I**taliana) started in 2010 to support the development of **technologies** proposed for the Cherenkov Telescope Array Small-Sized Telescopes (SSTs).

The first result of the ASTRI project was the construction of **a prototype telescope** in 2014 in Italy (Sicily).

The ASTRI is an end-to-end project including the **full data archiving** and **processing chain, from raw data** up **to final scientific products**.

ASTRI-Horn prototype telescope installed at the astronomical INAF site on the slopes of the Etna volcano in Sicily, Italy

The next phase of the project, currently underway, is the construction of a series of **9 units of ASTRI telescopes** (named **ASTRI Mini-Array**) at the Teide Observatory (Canary Islands).

ASTRI-Horn Telescopes

The logging and monitoring system takes into consideration all the **telescopes sub-systems** to control the telescope motion and the camera activities, such as:

- The telescope Camera
- The Mount Control System (MCS) for the motion of the mechanical structure
- The Active Mirror Control (AMC) for controlling the primary and secondary mirrors.

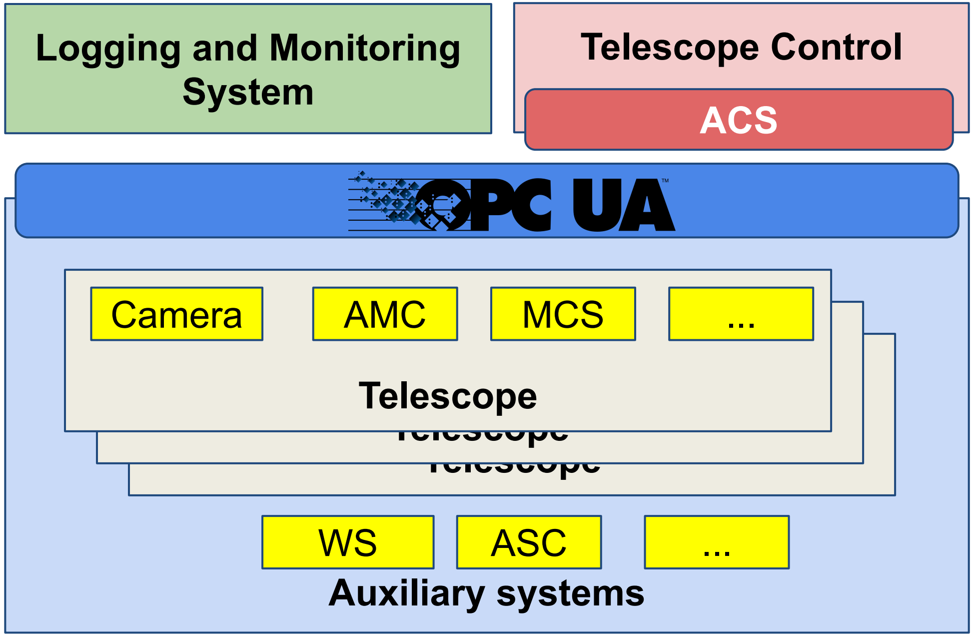
And **auxiliary systems** to detect the environmental conditions and to assess the quality of the observations, such as the Weather Station (WS) and the All Sky Camera (ASC)

Estimate a **technical/operational data load**:

- ~20.000 monitoring points
- ~ 14 GB / day -> ~5 TB / year

OPC-UA Interface protocol between the high-level control software and all the hardware devices.

The control components are implemented upon the **ALMA Common Software (ACS)** middleware.

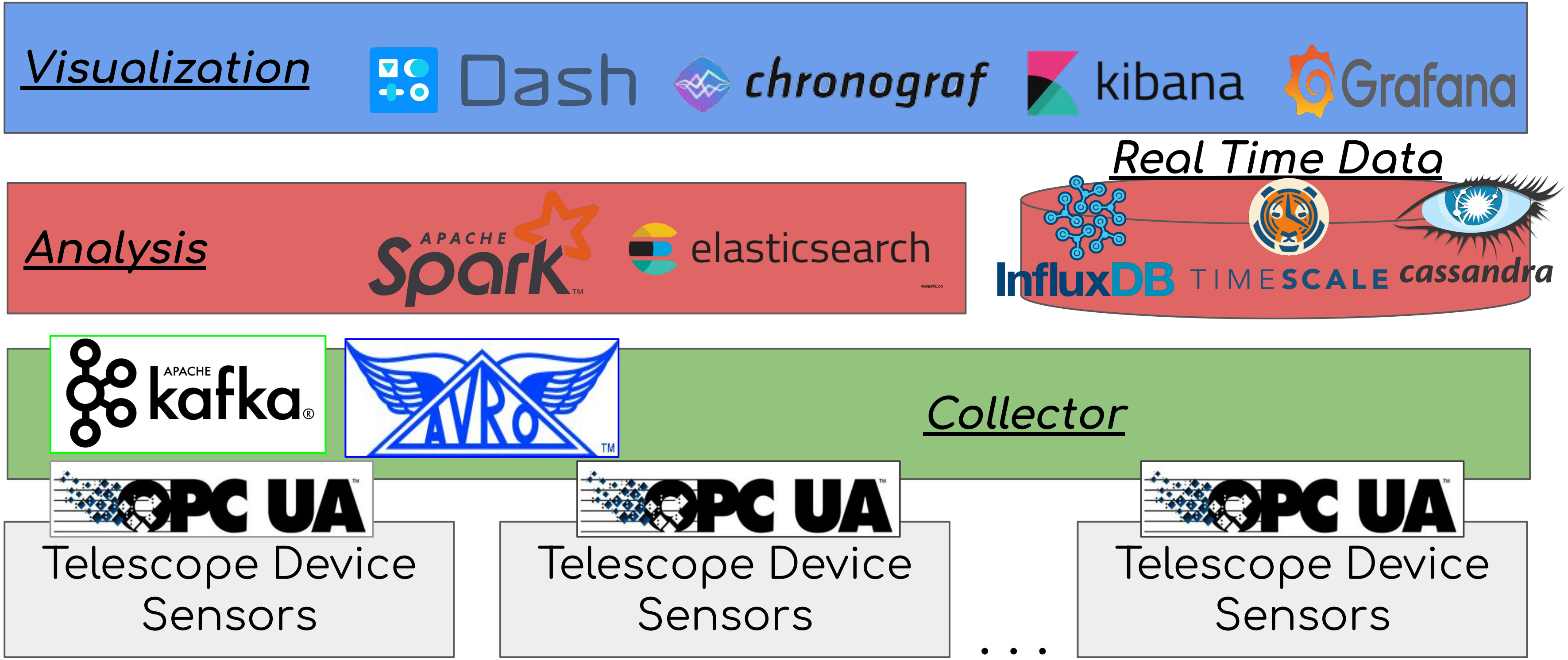


Internet of Things (IoT) is an emerging technology that is becoming an increasing topic of interest among **technology giants** and **business communities**.

IoT components are **interconnected devices** embedded with sensors, software and smart apps to **collect and exchange data** with each other or with **cloud/data centres**.

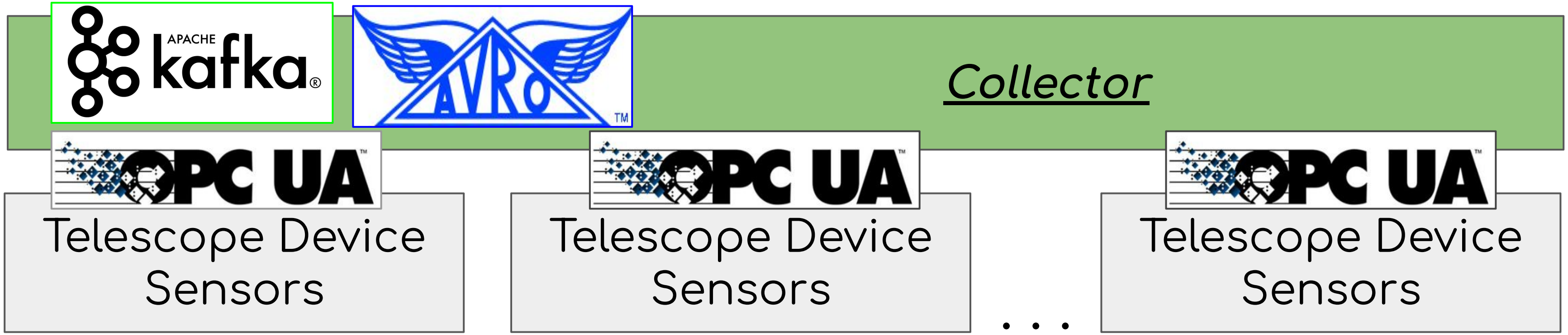
Data generated by IoT devices needs to be processed and analyzed using **Big Data analytics engine** in order to **extract the critical information** or to **understand behavioural patterns**.

IoT Software Architecture

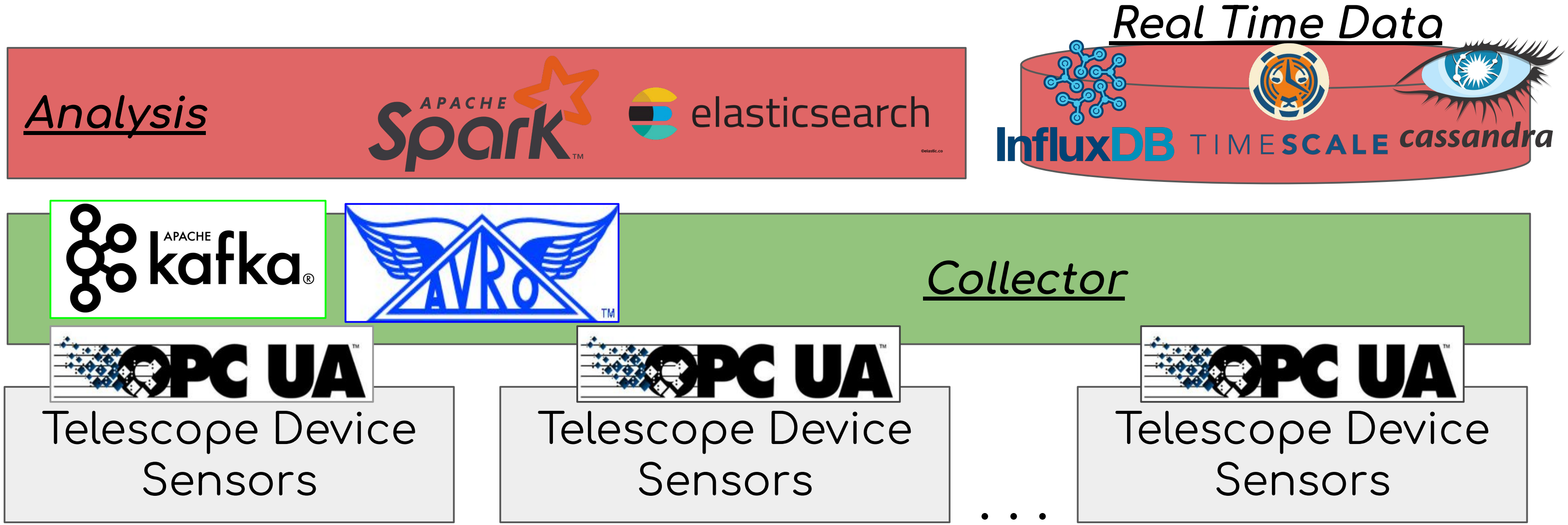


IoT Software Architecture: Collector

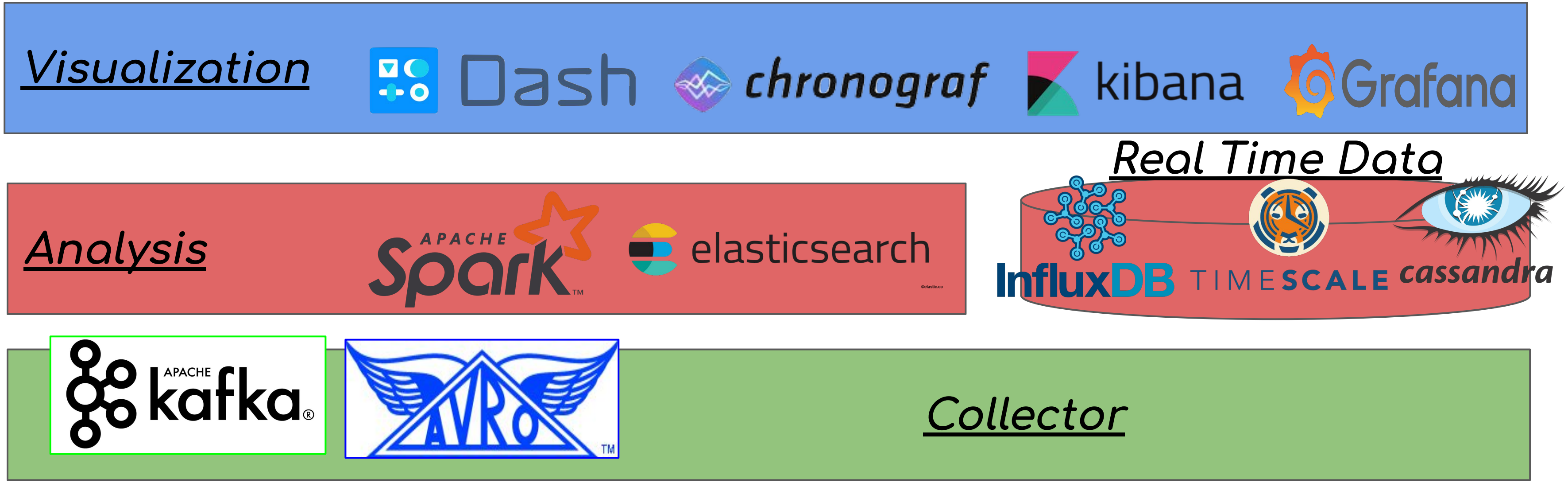
The **Collector** gathers together the **IoT messages** from the connected telescope devices which are captured by a **message broker** and are sent to the streaming application for processing.



The **Analysis** layer comprises a **streaming application** which **consumes IoT data streams** and processes them for data analysis. The IoT data processor stores the processed data in the **real time databases**.



The **Visualization** component retrieves data from databases and send them to **web pages** in fixed intervals so data are refreshed automatically. Dashboard **displays data** in charts and tables using a responsive web design and it is **accessible** on desktop as well as mobile devices.



Data Workflow: IoT data production and collection

Devices parameters are sent through **OPC-UA protocol** and described by an **Interface Control Document** (ICD).

The ICD comprises, in form of tables, for each control or **monitoring point**, a **complete description** of the information required, e.g. data type, OPC-UA node and connected alarms.

GET_AUX_WS_EXTTMP			
GET_AUX_WS_DEWPOINT			
GET_AUX_WS_WINDDIR			
GET_AUX_WS_WINDIR10M			
GET_AUX_WS_WINDSPD			
GET_AUX_WS_WINDGUST10M	yes	yes	ALL
GET_AUX_WS_WND10AVG	yes	yes	ALL
GET_AUX_WS_SOLARRAD	yes	yes	ALL
GET_AUX_WS_EXTUMDY	yes	yes	ALL
GET_AUX_WS_RAINRATE	yes	yes	ALL
GET_AUX_WS_RAINDAILY	yes	yes	ALL
GET_AUX_WS_RAIN1H	yes	yes	ALL
GET_AUX_WS_RAIN15M	yes	yes	ALL

In our application, the **IoT data producer** is an **OPC-UA simulator application** for connected devices.

Data are collected using an **OPC-UA client** and sent through an **Apache Kafka Producer** over a given **topic** serialized using an **Apache Avro** schema

Data Workflow: IoT data serialization


```
{
  "namespace": "avro.example",
  "type": "record",
  "name": "MonitoringPoint",
  "fields": [
    {"name": "id", "type": "string"},
    {"name": "env_id", "type": "string"},
    {"name": "serial_no", "type": "string"},
    {"name": "date", "type": "long"},
    {"name": "data", "type": { "type": "array", "items": "double"} }
  ]
}
```



```
{
"id": "SST01/WeatherStation01/ws_windspd",
"env_id": "SLN",
"serial_no": "WS-01",
"date": 13203093272171,
"data": [20.921472]
}
```

Data Workflow: IoT data analysis and storing

A **Kafka Consumer** get the **deserialized** data from the subscribed topic.

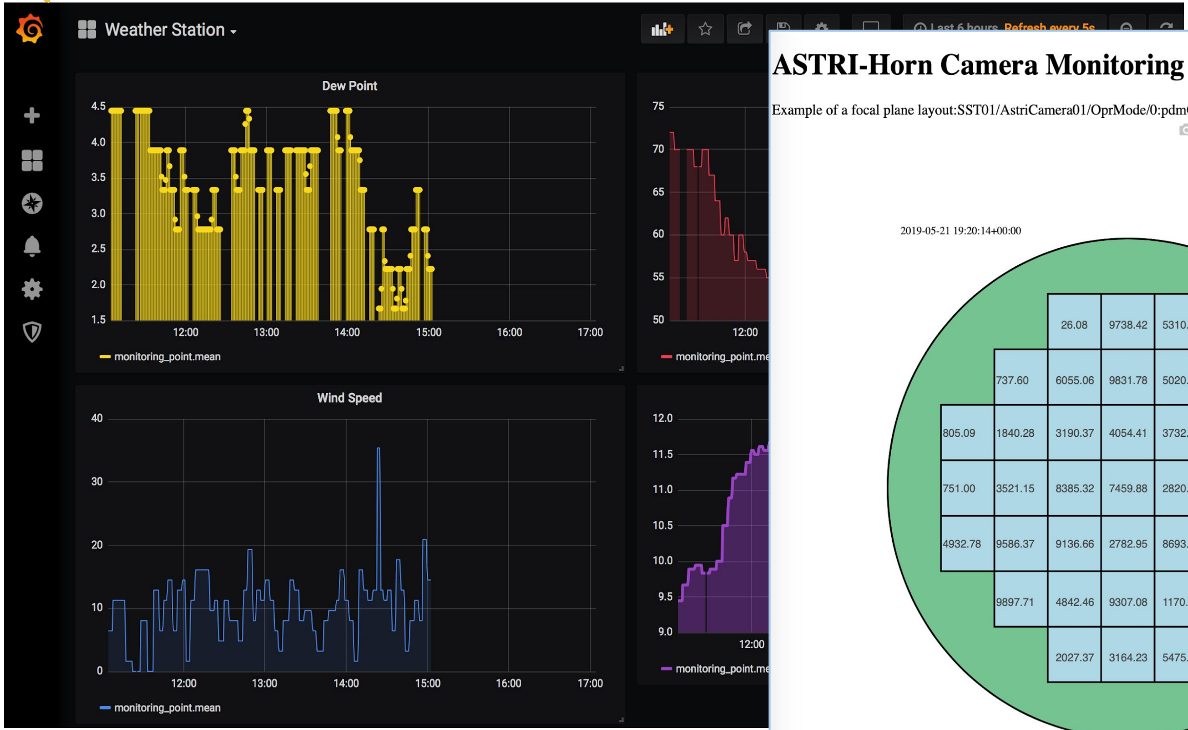
The **Spark** application analyzes the data stream and insert the analysis results into a **non-relational database** (optimized for real-time and Big Data applications, such as Cassandra or InfluxDB).

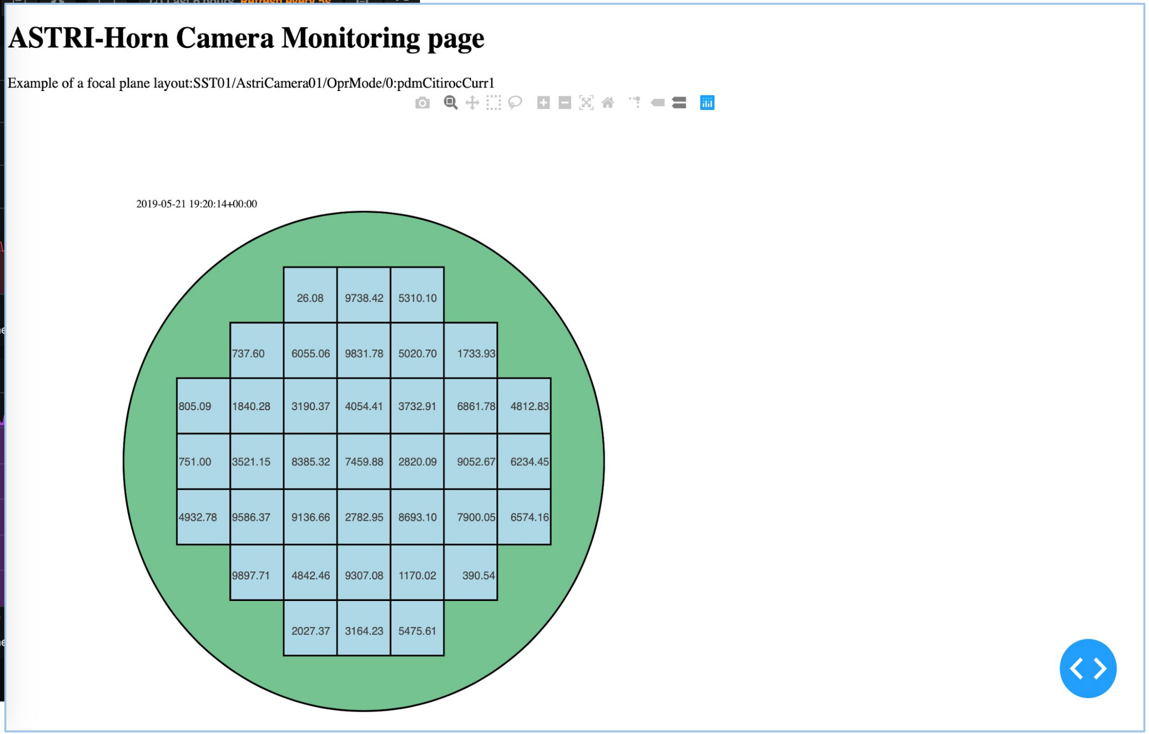
Finally, monitoring data are **visualized** through **interactive dashboards**.

Data Workflow: sample visualization

Grafana

Dash

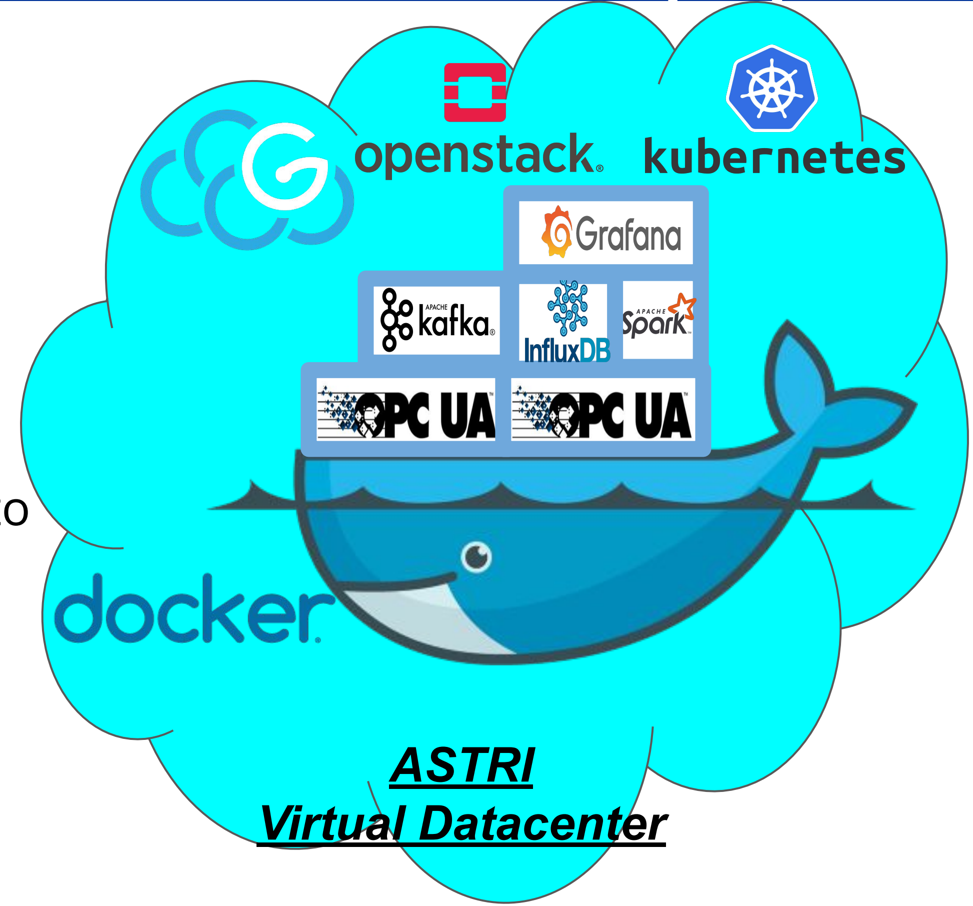




SW Stack deployed on the **GARR Cloud Platform** within a specific **virtual data center** using a **container based virtualization**.

The Cloud infrastructure provides the ability to **simulate** a **large number of sources** producing IoT data and **scale** easily the **various modules** according to the needed workload.

Testing sw architecture for an **entire array of telescope**, using any number of modules, each of them running as a separate **container**.



- ➔ We presented a **system** aimed at **monitoring and logging** the data needed to improve the **operational activities** of a **large scale telescope array**.
- ➔ The prototype system architecture was designed and built considering the latest **software tools** and **concepts** coming from **Big Data** and **Internet of Things**.
- ➔ The software stack is based on **open source software minimizing** the needs for **software development**.

Future Developments

- ➔ Planning **scalability tests** on the **GARR Cloud** infrastructure to **evaluate** the **system performances** simulating different virtual telescopes and scaling the various architectural modules according to the workload.
- ➔ Working on a **Web dashboard** that will allow the team to automatically increase or decrease live the number of simulated devices.
- ➔ Future works are planned to **integrate Machine Learning** algorithms to perform **anomaly detection** and **failure prediction.**
- ➔ Finally, the system will be **deployed** on the forthcoming **ASTRI Mini-Array**.

References and useful links

- ASTRI project: http://www.brera.inaf.it/astri/
- Pareschi, Giovanni. "***The ASTRI SST-2M prototype and mini-array for the Cherenkov Telescope Array (CTA).***" Ground-based and Airborne Telescopes VI. Vol. 9906. International Society for Optics and Photonics, 2016.
- Tanci, Claudio, et al. "***The ASTRI mini-array software system (MASS) implementation: a proposal for the Cherenkov Telescope Array.***" Software and Cyberinfrastructure for Astronomy IV. Vol. 9913. International Society for Optics and Photonics, 2016.
- Pena, Eduardo, et al. "***Framework to use modern Big Data Software Tools to improve operations at the Paranal Observatory.***" Observatory Operations: Strategies, Processes, and Systems VII. Vol. 10704. International Society for Optics and Photonics, 2018.
- Antolini, Elisa, et al. "***Telescope Control System of the ASTRI SST-2M prototype for the Cherenkov Telescope Array***." 16th Int. Conf. on Accelerator and Large Experimental Control Systems (ICALEPCS'17). JACOW, Geneva, Switzerland, 2018.