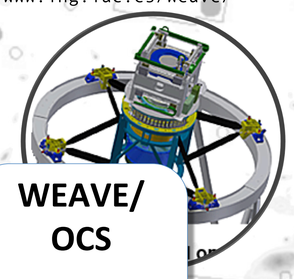
# The Multifaceted WEAVE Archive System

Emilio Molinari[1,2], Marcello Lodi[1], José Guerra[1], Chris Benn[3], Lilian Dominguez[3]
[1]TNG – INAF Fundación Galileo Galilei, Breña Baja, Spain - [2]INAF – IASF Milano, Milano, Italy - [3]Isaac Newton Group of Telescopes, Santa Cruz de la Palma, Spain
contact: emilio.molinari@tng.iac.es

The WEAVE MOS and IFU spectrograph will produce millions of spectra. This big-data product needs a robust high-performance tool to allow end users to access the data reliably and easily for their specific research projects. We are building a system which will take care of transport of data, their ingestion and subsequent population of a repository and a database. The back-end of the system will have access to the complete set of raw files and to the whole data product tree of the WEAVE pipeline and data flow. This will allow a different set of front-ends to be plugged into the WAS. A first, general GUI interface, which will cater for most of the users' needs, will be provided. The same system will give ultimately give broad access to the public releases. The underlying database structure and performance will be guaranteed with a redundant network of disks and CPUs, and robust open-source software for which troubleshooting support is available for a high fraction of the time. The WAS is part of the contribution of INAF (Italy) to the WEAVE project and will be implemented by TNG, Fundación Galileo Galilei in La Palma.
Follow the link http://www.ing.iac.es/weave/



**WEAVE/ OCS**

The spectrograph at WHT, La Palma, and its Observing Control Software (OCS)

FITS files

**CPS**

The Core Processing Software (CPS) at CASU, Cambridge, UK.

Catalogues

Spectra

**USERS**

WEAVE Science Team members as well as PI of open time proposals.

REQ

ANS

**APS**

The Advanced Processing Software (APS), developed by IAC, la Laguna, TF, Spain

astrophys

USERS allowed interface

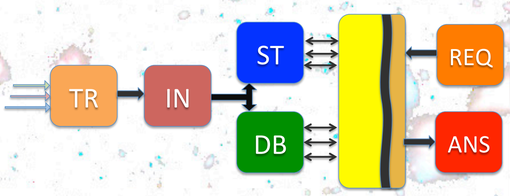
Main WEAVE data flow

Storage of intermediate products data flow.

Solr and Spark are the open source Apache modules for communication layer (the former) and for data mining (the latter). We will use Solr to allow requests and give answers. Spark will be a possible add on for massive data mining projects. Both come bundled with Datastax Cassandra.

Solr

Spark



Cassandra

We selected Apache Cassandra as the engine for our database. It is a challenging but also a natural choice because of its linear scalability and fault-tolerant database for mission-critical environment without compromising the performance. Cassandra's NoSQL datamodel offers a high column index performance with strong support for denormalization, materialized views and ultra-fast in-memory distributed options able to cope with the most challenging queries.

We partnered with Datastax company for the support and continuous upgrade of the Apache Cassandra open source software we are using. The collaboration is free of charge and will make of WEAVE a singular use case for the company.

DATASTAX

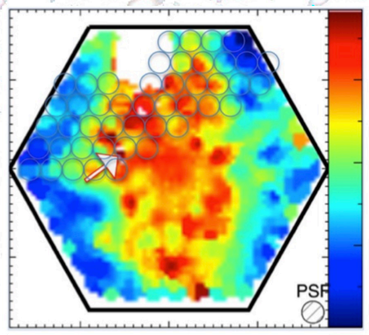



The basic structure of the WAS is shown here. The data are coming from the Users, OCS, CPS and APS via a **TR**ansport module, which synchronizes the data when they are produced. The **IN**gestion phase stores the raw and processed data in the **ST**orage file system (a distributed GlusterFS) and the (meta)data in the **DataBase**. A layer of interface will allow external procedure to access both, via a **REQ**uest and **ANS**wers protocol. The Astronomical User Interface is using this layer (based on SOLR) to interact with the real WAS.

# WAS

The WEAVE Archive System (WAS) will be developed, maintained and hosted by Fundación Galileo Galilei – INAF, in the premises of the its TNG (Telescopio Nazionale Galileo, to Italy) in the island of La Palma, Canary Islands, Spain.

www.tng.iac.es

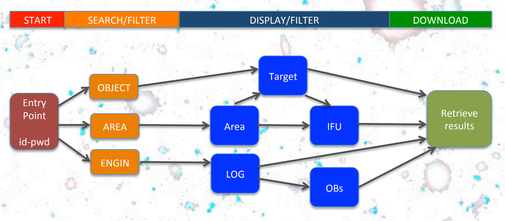



IFU (**integral field unit**) mode will be an important aspect of WEAVE, through its large IFU (78"x90") and 20 mini IFUs (11"x12"). The user interface of WAS will have limited tools for IFU visualization in the filter/search layer.

We have chosen OpenNebula as the environment for our virtualization scheme. Virtualization will ensure easier machine replacement and upgrade, while keeping low the hardware costs.

OpenNebula.org

The **Astronomical User Interface** will offer various way to query and filter the database content, prior to getting the results.
Results will be tables of metadata produced by the WEAVE reduction pipeline, as well as an interface to the real data: raw FITS as observed at the telescope, and a series of intermediate products such as calibrated spectra, extracted sky, spectral fits to libraries, …




IVOA

The level of compatibility with the Virtual Observatory is under consideration. The possibility of devising a TAP (Table Access Protocol) layer has to deal with the fact that Cassandra is a non relational database. The solution will undoubtedly be a great added-value to the WAS and VO itself.