WAS: The Archive for the WEAVE Spectrograph

Jose Guerra, Adrian Martin, Emilio Molinari, Marcello Lodi, Fundacion Galileo Galilei. Telescopio Nazionale Galileo –TNG (Spain),

Gavin B. Dalton, Univ. of Oxford (United Kingdom); Scott C. Trager, Univ. of Groningen (Netherlands); Don Carlos Abrams, Isaac Newton Group of Telescopes (Spain); Piercarlo Bonifacio, Observatoire de Paris Meudon (France); Jose Alfonso Lopez Aguerri, Instituto de Astrofisica de Canarias (Spain); Antonella Vallenari, INAF - Osservatorio Astronomico di Padova (Italy); Esperanza E. Carrasco Licea, Instituto Nacional de Astrofisica, Optica y Electronica (Mexico); Kevin F. Middleton, STFC Rutherford Appleton Lab. (United Kingdom)

Introduction

Today for a foremost group of astronomers and astrophysics, WEAVE represents an intersection between science and engineering producing a new revolution in terms of observational data and its processing.

WEAVE Archive (WAS) is a compendium of advanced cluster technologies to enable high scalability, real-time analytics, high storage capacity and a new visualization tool-kit based on modern Web concepts. WAS also provides ways to exchange data to different formats (FITS files, CSV, JSON, VO Tables) and allows bulk copies for massive downloads.

Within the core of **WAS,** its database is currently based on **DataStax Enterprise** search tools; *Apache Cassandra* and *Apache Solr*. These packages enable the archive to work with multi-datacenter database technology capable to distribute and replicate data at global scale, allowing Cloud level search operations.
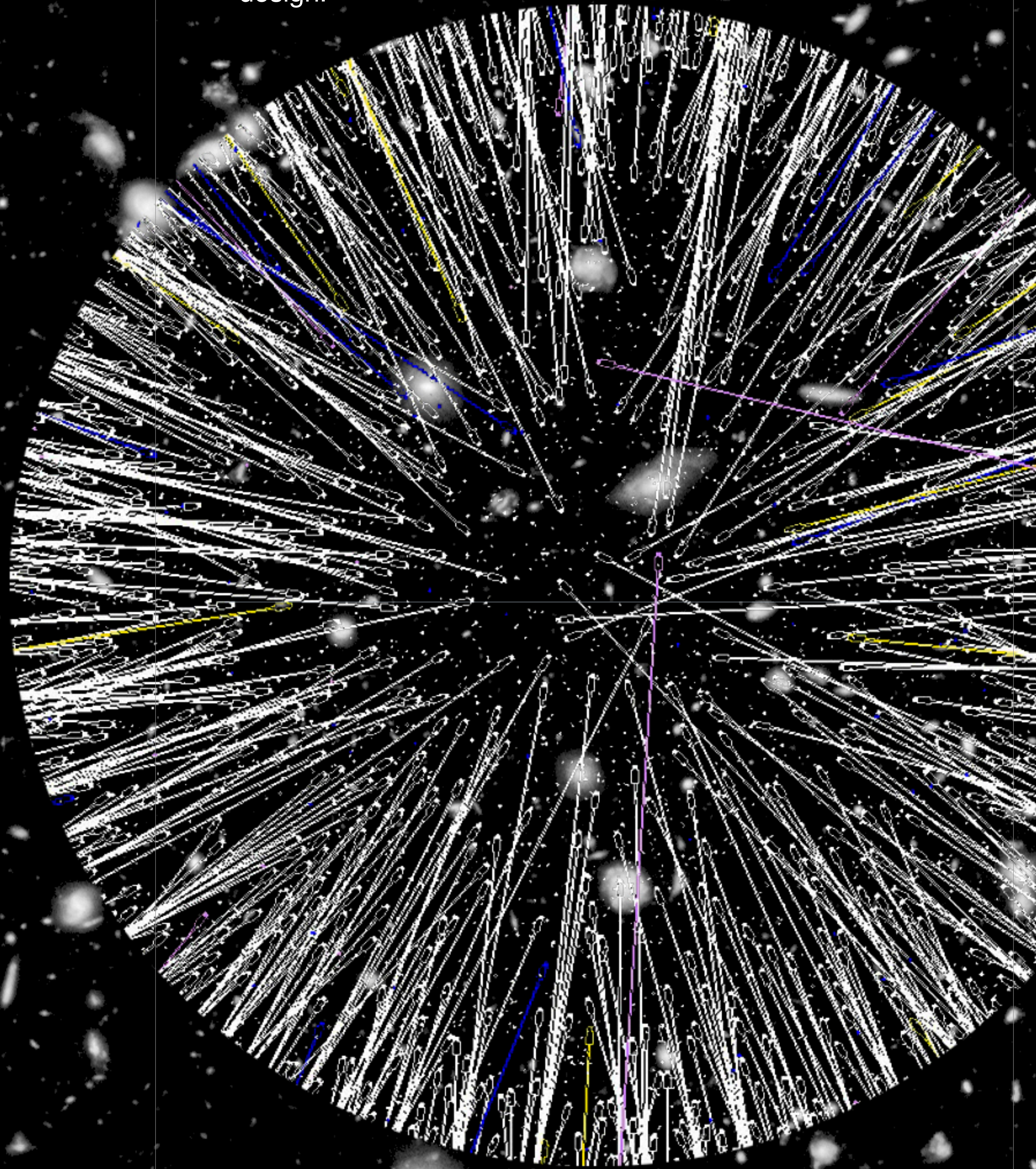
WAS database concept architecture.
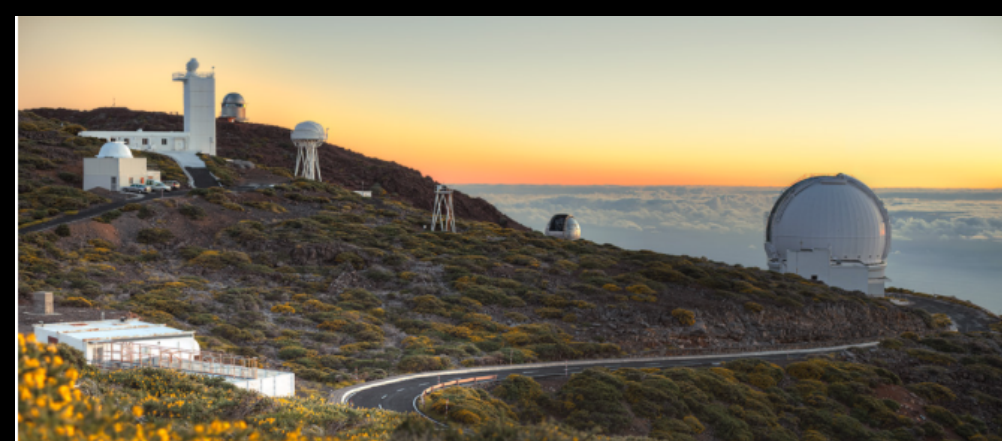
WHT at the ORM, La Palma. Source ING.
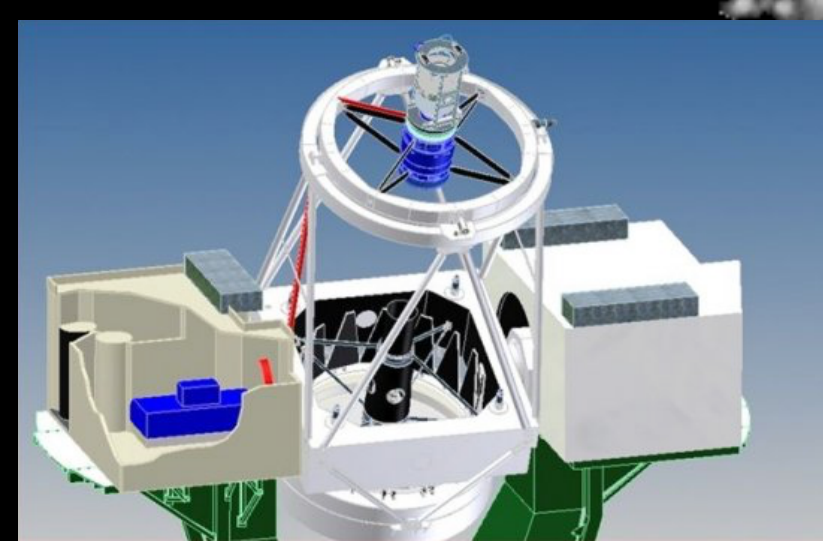
WEAVE mounted. Source IAC.

Technology

WAS is mostly using stream-line technologies from the database up to the web user interfaces. Standard open technologies and protocols are the key points of WAS design.

Fig. X.

WAS is based on modern programming languages and techniques expressed in Java, Python, HTML5 and Javascript, using the latest APIs in communication and visualization aids.
Other core technologies within WAS are HTTP RESTfull Web Services, Websockets and JSON encoding for the messaging services.
All these technologies are blended in with the latest hardware servers, virtualization and network equipment.
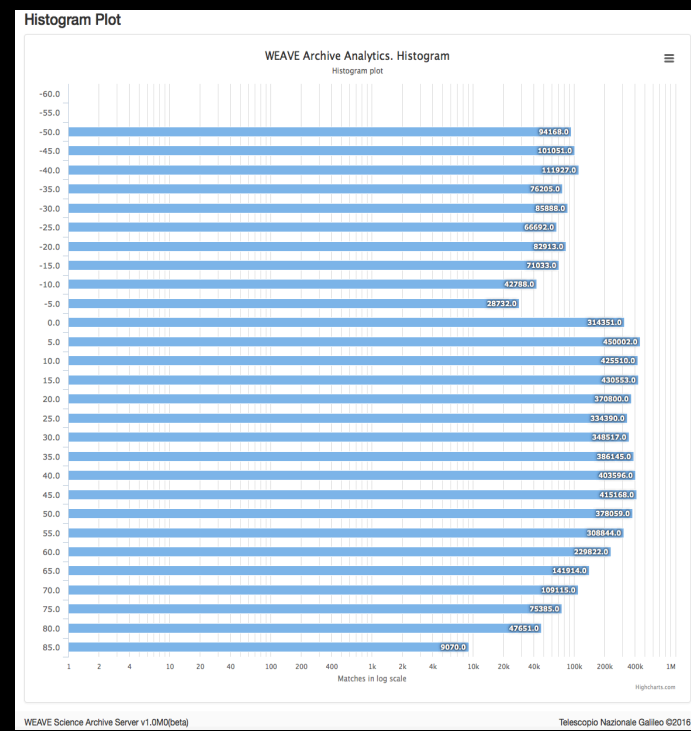
Results

One of the most powerful WAS outputs are real-time analytics and searches. Plot examples depicted below displays the very first WAS prototype outputs in tables, spectra, histograms, heatmaps and sky plots.
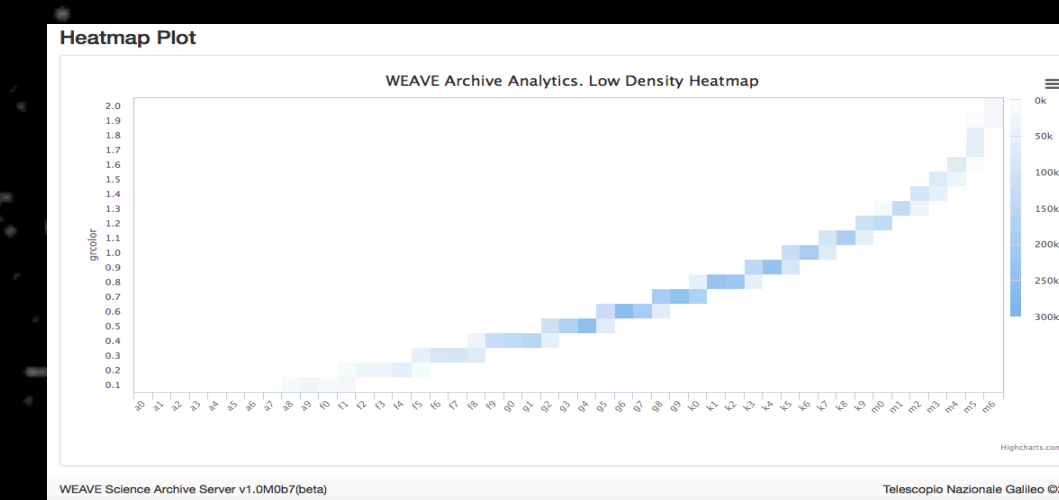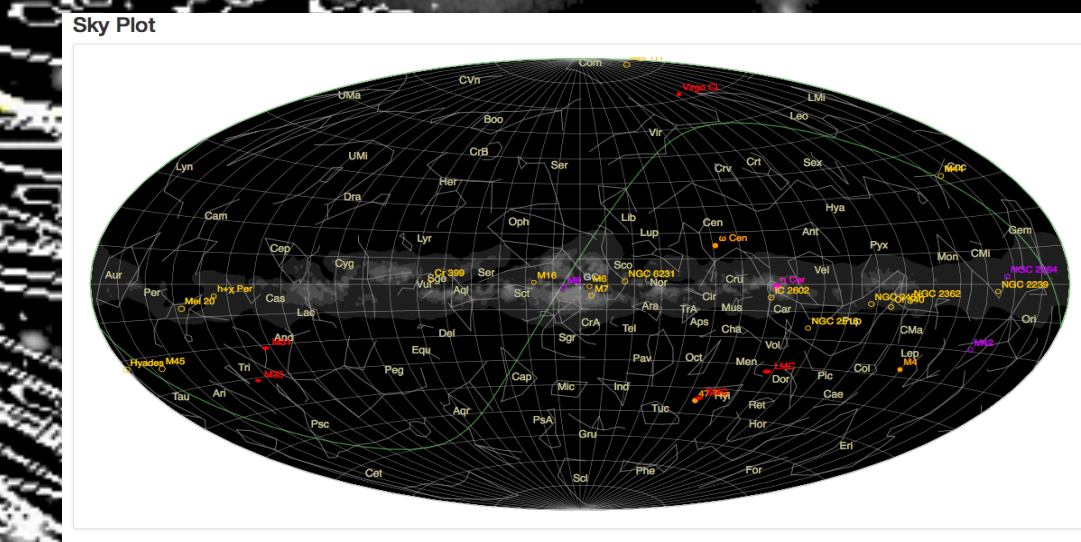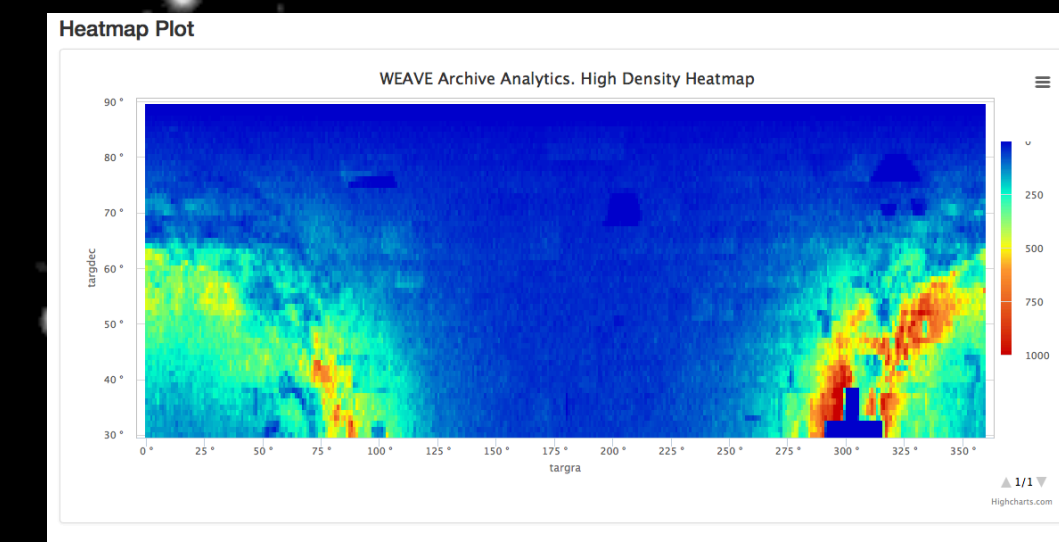
Spectra plots.

Histogram plot.

Tabular data display

Target color vs subclass heatmap plot

Mollweide galactic plot

RA vs DEC high density heatmap plot

Conclusions

Storage Capacity: Cassandra's capacity is far beyond WEAVE project could never produce for the foresee instrument's lifespan. Currently the database is dimensioned to only < 12TB in five years of not replicated meta data + spectra.
Fits files are planed to be kept by distributed files-system in GlusterFS.
Searching Features: Real-time searches and analytics using Apache Solr including +180 indexes only with the preliminary data-model.
Exporting Data: WAS will export in several ways, from single target FITS file until massive bulk downloads.
Visualization: A novel Web UI based on HTML5 and Javascript will support a custom visualization ecosystem**.**

Related Conference Readings:

[1] Final design and progress of WEAVE: the next generation wide-field spectroscopy facility for the William Herschel Telescope. SPIE 9908-53
[2] The HARPS-N archive through a Cassandra, NoSQL database suite ?. SPIE 9913-125