



Contribution ID: 3

Type: **not specified**

CosmoHub on Hadoop: a web portal to analyze and distribute massive cosmological data

Wednesday, 26 April 2017 12:40 (25 minutes)

We present CosmoHub, a web platform to perform interactive analysis of massive cosmological data without any SQL knowledge. CosmoHub is built on top of Apache Hive, which is an Apache Hadoop ecosystem component, which facilitates reading, writing, and managing large datasets.

CosmoHub is hosted at the Port de Informació Científica (PIC) and currently provides support to several international cosmology projects such as the Euclid space ESA mission, the Dark Energy Survey (DES), the Physics of the Accelerated Universe (PAU) and the Marenstrum Institut de Ciències de l'Espai Simulations (MICE). More than two billion objects, from public and private data, as well as observed and simulated data, are available among all projects. In the last three and a half years more than 400 users have produced about 1500 custom catalogs occupying 2TB in compressed format.

CosmoHub allows users to access value-added data, to load and explore pre-built datasets and to create their own custom datasets through a guided process. All those datasets can be interactively explored using an integrated visualization tool which includes 1D histogram and 2D heatmap plots. In our current implementation, online analysis of datasets of a billion objects can be done in less than 25 seconds. Finally, all those datasets can be downloaded in three different formats: CSV.BZ2, FITS and ASDF.

The components, integration and performance of the system will be reviewed in this contribution.

Length of talk (minutes)

20

Scheduling constraints / preferences

Would like to have ethernet internet connection for a live demo.

Primary authors: Dr CARRETERO, Jorge; TALLADA CRESPI, Pau (Universitat Autònoma de Barcelona (ES)); CASALS HERNANDEZ, Jordi (University of Barcelona (ES)); CAUBET SERRABOU, Marc (Universitat Autònoma de Barcelona (ES))

Co-authors: DELFINO REZNICEK, Manuel (Universitat Autònoma de Barcelona (ES)); Mr TORRADEFLOT, Francesc; TONELLO, Nadia (PIC); Dr FOSALBA, Pablo; Dr DELGADO, Jordi; Dr SERRANO, Santiago; NEISSNER, Christian (PIC); ACOSTA SILVA, Carlos (Universitat Autònoma de Barcelona (ES))

Presenter: CASALS HERNANDEZ, Jordi (University of Barcelona (ES))

Session Classification: Computing and batch systems

Track Classification: Computing & Batch Services