Handling Cosmological Simulations @ PIC

Theory GWS Science Platform Workshop
IVOA May 2021 Interoperability Meeting

J.Carretero on behalf of PIC team

Outline

- PIC and Cosmology
- Challenges
- PIC hadoop platform
  - CosmoHub
- Interoperability

Port d'Informació Científica

PIC
port d'informació
científica

Port d'Informació Científica

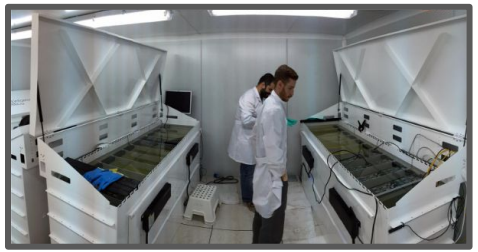
PIC
port d'informació
científica

- Founded in 2003: collaboration between IFAE and CIEMAT
- Team of 19 people (50% scientists - 50% engineers)
  - Agile teams that embed in scientific groups
    - understand the experiment,
    - follow the evolution of data management requirements,
    - work with engineers to develop and prototype solutions.
- Run production data services for large international collaborations
  - Collaborative environments, distributed infrastructures
- Flexibility to adapt to evolving needs
  - Integration of data processing tools/methods - trans-disciplinar cross-pollination
  - Infrastructure - experimental can-do attitude
    - Adapt off-the-shelf servers for immersion (fans, thermal paste, HDs, mechanical support…)
    - DIY servers to explore high density configurations

PIC data center

PIC
port d'informació
científica

- Facilities, ~150 kW IT
  - ~120 kW in 150 m^2 air-cooled room
    - high efficiency, PUE 1.44
  - ~30 kW in 25 m^2 liquid immersion cooling system
    - PUE 1.1



- Data processing services
  - Disk - dCache
    - 50 nodes, 10 PiB, support NFSv4, WebDAV, xrootd, gridftp...
  - Tape - Enstore
    - ~7000 slots, T10K and LTO8, 33 PiB
  - Computing - HTCondor
    - 321 nodes, 8100 cores, 10 GPUs
  - Computing - Hadoop
    - 16 nodes, 144 cores, 2 TiB RAM, 432 TiB SATA, 32 TiB NVMe, 2x10 Gbe
    - Spark pipelines and Cosmology data processing cluster and analysis web portal (https://cosmohub.pic.es)

- Connectivity
  - 2x10 Gbps to Academic Network, now upgrading to 2x100Gbps
  - Largest data mover in Spanish academic network: 70PB in+out per year

Supported projects

PIC
port d'informació
científica

- Particle Physics (**Spanish Tier-1 Large Hadron Collider** (LHC - WLCG):
  - LHC (Atlas, CMS, LHCb), neutrinos (T2K Japan, DUNE)
- Astrophysics:
  - MAGIC
  - Cherenkov Telescope Array (CTA)
  - Magnesia
- Cosmology
  - PAU
  - MICE (simulations)
  - DES
  - Euclid (**SDC-ES**)
  - DESI
  - LSST?
- Gravitational Waves
  - VIRGO/LIGO

Supported (Cosmology) projects

- Particle Physics (**Spanish Tier-1 Large Hadron Collider** (LHC - WLCG):
  - LHC (Atlas, CMS, LHCb), neutrinos (T2K Japan, DUNE)
- Astrophysics:
  - MAGIC
  - Cherenkov Telescope Array (CTA)
  - Magnesia
- Cosmology
  - PAU
  - MICE (simulations)
  - DES
  - Euclid (**SDC-ES**)
  - DESI
  - LSST
- Gravitational Waves
  - VIRGO/LIGO

Challenge 1: How to handle massive data

PIC
port d'informació
científica

- Data volume evolution: from a laptop, to a data center, to grid
- Goals
  - Scalable storage
  - Short iteration cycles
    - Fast access
    - Efficient analysis
  - Reproducibility
  - Traceability
- Our approach
  - Hadoop, Hive, Spark (code to the data!)
  - Multidisciplinary teams
    - so that scientists and software engineers understand each other
    - stand behind developers / Tutoring / Teaching (Bootcamps, MOOCs)
  - Jupyter+Spark notebooks, jupytext, git+LFS
  - Optimize algorithms
    - Custom algorithms (spark + treecorr)

Challenge 2: Gaining (more) independent users

PIC
port d'informació
científica

- Goal:
  - Reach a broader community
  - User should not notice the transition from laptop, to cluster, to grid
  - Simple, usable interfaces
    - Guided processes
    - Avoid configuration files and terminals
    - While still providing access to advanced features for expert users
- Our approach
  - Interfaces
    - CosmoHub
    - JupyterHub (+ extensions: i.e. VNC)
    - HTTP/Webdav
    - VO protocols
  - Training:
    - Documentation / Quick start guides (e.g. HTCondor / Spark)
    - Bootcamps / MOOC

Motivation: Galaxy catalogs

PIC
port d'informació
científica

Project	Date	volume / night	Total volume	Number of objects (catalog)
SDSS	**2000 - now**	variable	116 TiB	2×10^6
MICE GC (sims)	**2013**	NA	42 TiB	5×10^8
DES	**2013 - 2018**	2.5 TiB	2 PiB	4×10^8
GAIA*	**2014 - 2019**	40 GiB	1 PiB	1.1×10^9
Euclid	**2020 - 2025**	100 GiB	580 TiB	1.5×10^9
LSST	**2022 - 2032**	15 TiB	50 PiB	1×10^{10}

* DR1 full sky star catalog

PIC Hadoop platform

PIC
port d'informació
científica

- Based on Hadoop
  - Open source Big Data Platform
  - Distributed storage and processing
    - Runs on commodity computer clusters
    - Scalable from dozens up to thousands of nodes
      - *Performance scales with HW*
  - Fault tolerant
    - Simple machines working together - no single point of failure
- Recent update:
  - Custom DIY nodes
    - 12 nodes AMD Threadripper 1920X
    - 128 GB RAM, 12 x 3 TB SATA HDD hot-swap
    - 2x1 TB NVMe SSD i 2x10-GBASE-T LAN
  - Hortonworks HDP 3.1.4
    - Hadoop 3.1.0
    - Hive 3.1.0
    - Spark 2.3.2

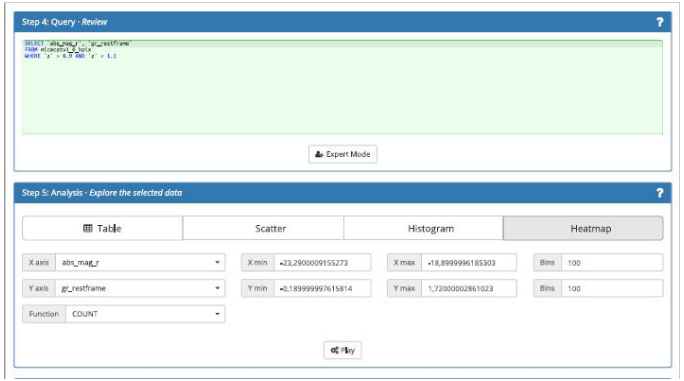

COSMO
HUB
Build your own Universe
Interactive data analysis of massive cosmological data without any SQL knowledge
Billions of observed and simulated galaxies
Superfast queries means superfast results
Features to make you work faster and easier
Online plotting preview and data download
https://cosmohub.pic.es/home

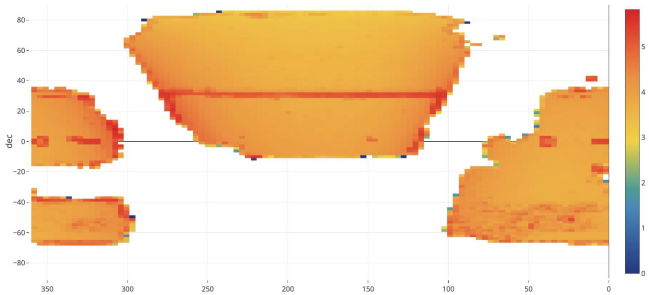
COSMO HUB

Tallada et al. 2020

PIC port d'informació científica

- Interactive exploration (visualization)
  - **Very fast** (85% < 30s)
  - Full dataset plots (over all rows)
    - May use sampling
    - Cone search tool
    - 1D histogram & 2D heatmap
  - Guided process (no SQL knowledge required)
    - Expert mode
    - UDF functions: e.g. healpix methods
- Distribution
  - Query time range: seconds to minutes
    - 75% in < 3 min
  - FITS format (custom SerDe)
  - Email with a link to download dataset
- Based on Python / Flask + Hadoop / Hive





DESI Legacy Survey z_phot < 0.1

Recently CosmoHub has been accepted by re3data.org as a Research Data Repository

COSMO HUB

~ 150 active users

~ 9000 custom catalogs

~ 40 TiB hosted data

> 10^{11} objects

Public catalogs

- LSST DESC DC2
- DES DR2
- Gaia EDR3
- GLADE (v2.3 & v2.4)
- PAUS+COSMOS photo-z catalog (v0.4)
- DESI Legacy Survey with Photoz (DR8)
- VIPERS photometry and spectroscopy (PDR2)
- CANDELS Bulge-Disk decomposition (2018)
- PAUS-COSMOS Early Data Release (v1.0)
- COSMOS2015 Laigle (v2.1)
- DES Y1A1 Morphological catalog (v1.0)
- DES Y1A1 Gold Data (v1.0)
- KiDS (DR4)
- ALHAMBRA S/G CLASSIFIED (v1.0)
- CFHTLenS (good fields) (v1.2)
- Gaia (DR2)
- Alhambra photometric redshifts (v1.0)
- PAU.MillGas Lightcone (2016-07-18)
- Gaia (DR1)
- MICECAT (v2.0)
- DEEP2 Redshift catalog (DR4)
- MICECAT (v1.0)

Future work / Interoperability

- *Implement VO protocols (ADQL, TAP, ...)*
  - Include CosmoHub in the IVOA registry service
- Multi-messenger (MM) - CosmoHub
- “Unified science platform”: from the reduced data to the paper in one login
  - Large volumes of data should not be transferred from data centers to scientists
  - Process/validation codes integrated in the data centers
- Social features
  - sharing plots/notebooks/datasets
  - collaborative edition
  - feed subscriptions

Conclusions

- Multidisciplinary team
  - combine multiples experiences and knowledge
  - key to bridge the gap between science and computing
- Careful selection of the technological solution
  - Hadoop / Hive / Spark
  - JupyterHub
- Simple interfaces
  - CosmoHub
  - Jupyter notebooks
  - remote VNC sessions
- PIC perspective (small data center):
  - Do not try to compete in the “megawatt-compute” arena, instead focus on
    - Data services: preservation, analysis, sharing
    - Use modern technologies to be effective and agile responding to scientists needs

Thanks for your attention
Questions?

Credits to: E. Acción, V. Acín, C. Acosta, A. Bruzzese, J. Carretero, J. Casals, R. Cruz, M. Delfino, J. Delgado, J. Flix, G. Merino, C. Neissner, A. Pacheco, C. Pérez, A. Pérez-Calero, E. Planas, M.C. Porto, B. Rodríguez, P. Tallada, F. Torradeflot

www.pic.es