yt-AGORA analysis pipeline on NERSC

snapshots->rockstar->sunrise->images+plots

Project AGORA : Assembling Galaxies Of Resolved Anatomy

Simple science objective

"To raise the realism and predictive power of galaxy formation simulations and the understanding of feedback processes"

Project AGORA : Assembling Galaxies Of Resolved Anatomy

Big data challenge

Simulation Design → Data Products

Halo Mass Range: M_{vir} (M_{sun}) = 10^{10}-10^{13}

Quiescent vs. Violent Assembly Histories

Common Initial Conditions

Evolving same physics with 7+ high-performance codes using different algorithms

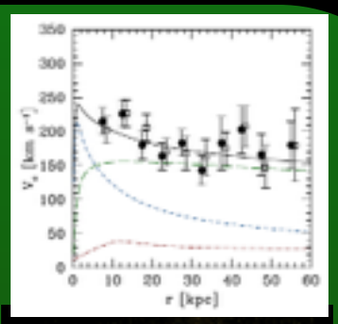
yt

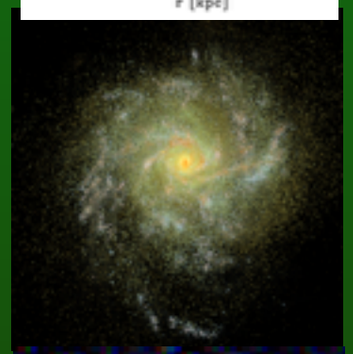
Common Analysis Platform

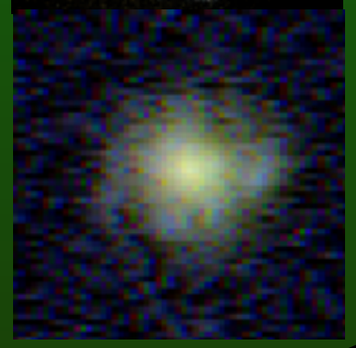
+

rockstar

sunrise







How is NERSC helping us with this challenge

Providing:

- A shared project space for storage (~100 Tb) and analysis /project/projectdirs/agora
- State-of-the art computational resources (5M cpu-hours), including GPU's (Xeon Phi's coming soon)
- Great infrastructure to develop a project Science Gateway
- Extensive support

How are we proposing to use this resources

- Having a common space where we can keep and share large data sets
- Having a common working installation of all the codes and scripts required for analysis
- Having a place where we can apply scripts and workflows to produce publication-ready data products
- Developing interactive data analysis and visualization tools that can be accessed via a Science Gateway

Common Space

We already have a common space that can be accessed from any computing resource at NERSC
/project/projectdirs/agora

You can start transferring data there as soon as you get a NERSC account

Each group should put their data under a directory named after their code, and control access to users with file ACL (e.g. setfacl)

This is how our space looks like now:

```
mrocha@edison06 agora$ ls
ART_NMSU  GIZMO  scripts  www
```

Common Installation

I have installed yt-3.0, Rockstar+consistent-trees, Sunrise with MPI+CUDA support, and a few other packages useful for analysis

```
mrocha@edison06 agora$ ls .AGORA_PIPE_INSTALL/
activate_yt-agora.sh  include  packages  share
bin                   lib      scripts   yt-agora_install
mrocha@edison06 agora$ ls .AGORA_PIPE_INSTALL/packages/
astropy                  ffmpeg-2.3.2  pycuda    sunrise_dependencies
consistent_trees-0.99.9.2  impression    rockstar  visnap
Downloads                mpi4py        sunrise   yt-agora
```

Sourcing the activate_yt-agora.sh will set up your environment so that all the installations work

```
mrocha@edison06 agora$ source scripts/activate_yt-agora.sh
(yt-agora)mrocha@edison06 agora$
```

This indicates that the yt-agora environment is active

Scripts and Workflows

The scripts directory contains scripts that we want to make available project-wide

```
(yt-agora)mrocha@edison06 agora$ ls scripts
activate_yt-agora.sh   run_rockstar.py   run_sunrise.py
dirac_streamRender.py  run_rockstar.py~
```

I am planning to setup tools such like fireWorks and/or qdo to create workflows that manage the order in which a set of scripts are applied to datasets, and how to take them through the queuing system.

More coming soon, stay tuned!

AGORA Science Gateway

A Science Gateway is a community-developed set of tools, applications, and data collections that are integrated through a portal or a suite of applications. Gateways provide access to a variety of capabilities including workflows, visualization, resource discovery, and job execution services.

Science Gateways allow researchers to focus on their scientific goals minimizing the administrative and technical challenges associated with the cyber-infrastructure they require.

We are building a Science Gateway for AGORA
www.agoragateway.org

How to get an account at NERSC

Submit this form online:

http://www.nersc.gov/users/accounts/user-accounts/get-a-nersc-account/

Choose a "standard" account, and find the repository name "agora"

If you already have a NERSC account e-mail me:migroch@gmail.com or Joel:joel@ucsc.edu to add you to the agora group