Renaissance Simulations Laboratory - a NDS Lab pilot

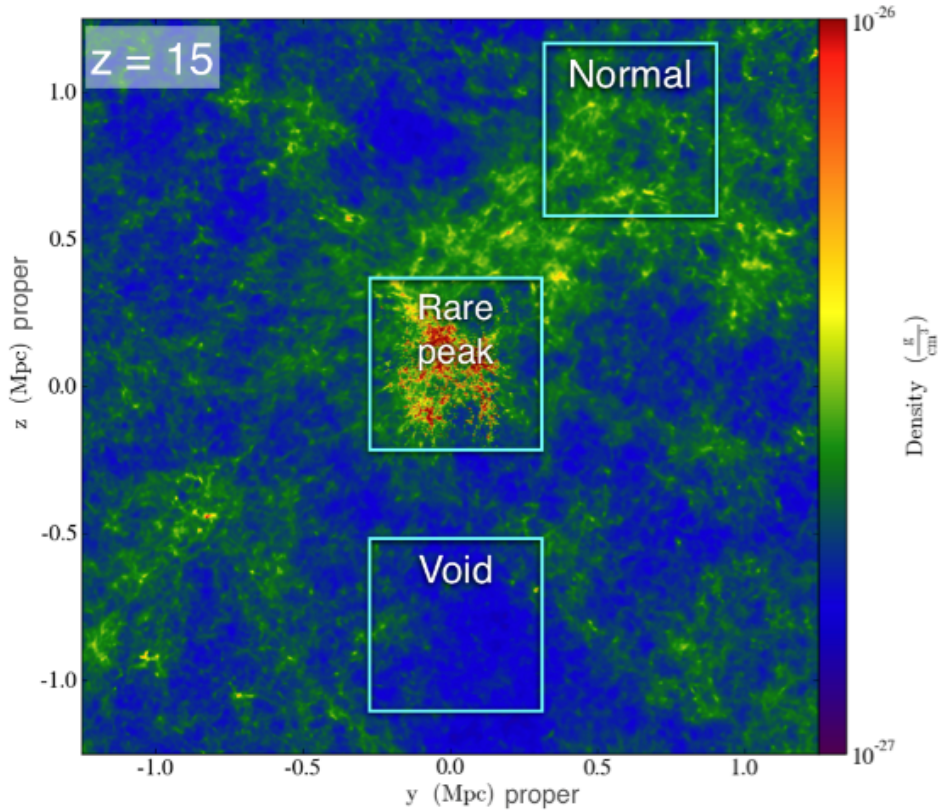
Hao Xu, Mike Norman (SDSC, UCSD)

Matt Turk, Kacper Kowalik (NCSA, UIUC)

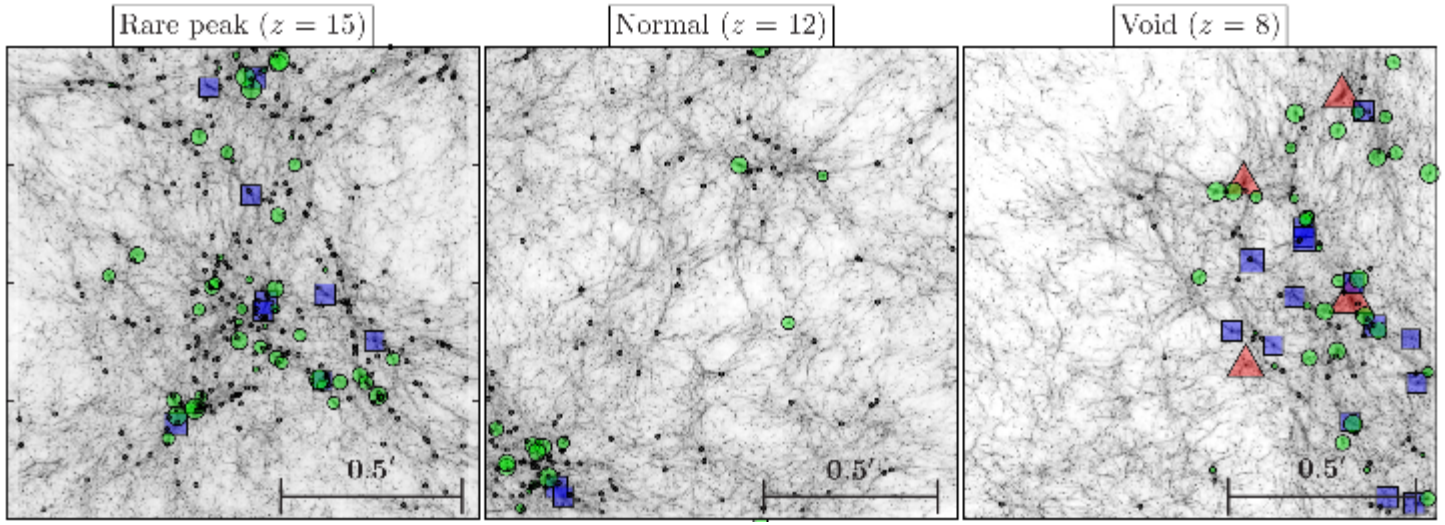
Purpose

- Data exploration lab for the *Renaissance Simulations*
- Provide rich set of data services (query, analyze, visualize, retrieve, etc.)
- Document our publications' analysis methods via curated *Jupyter* notebooks
- Encourage collaborative exploration of data via powerful execution environment and private or sharable *Jupyter* notebooks

Renaissance Simulations Laboratory



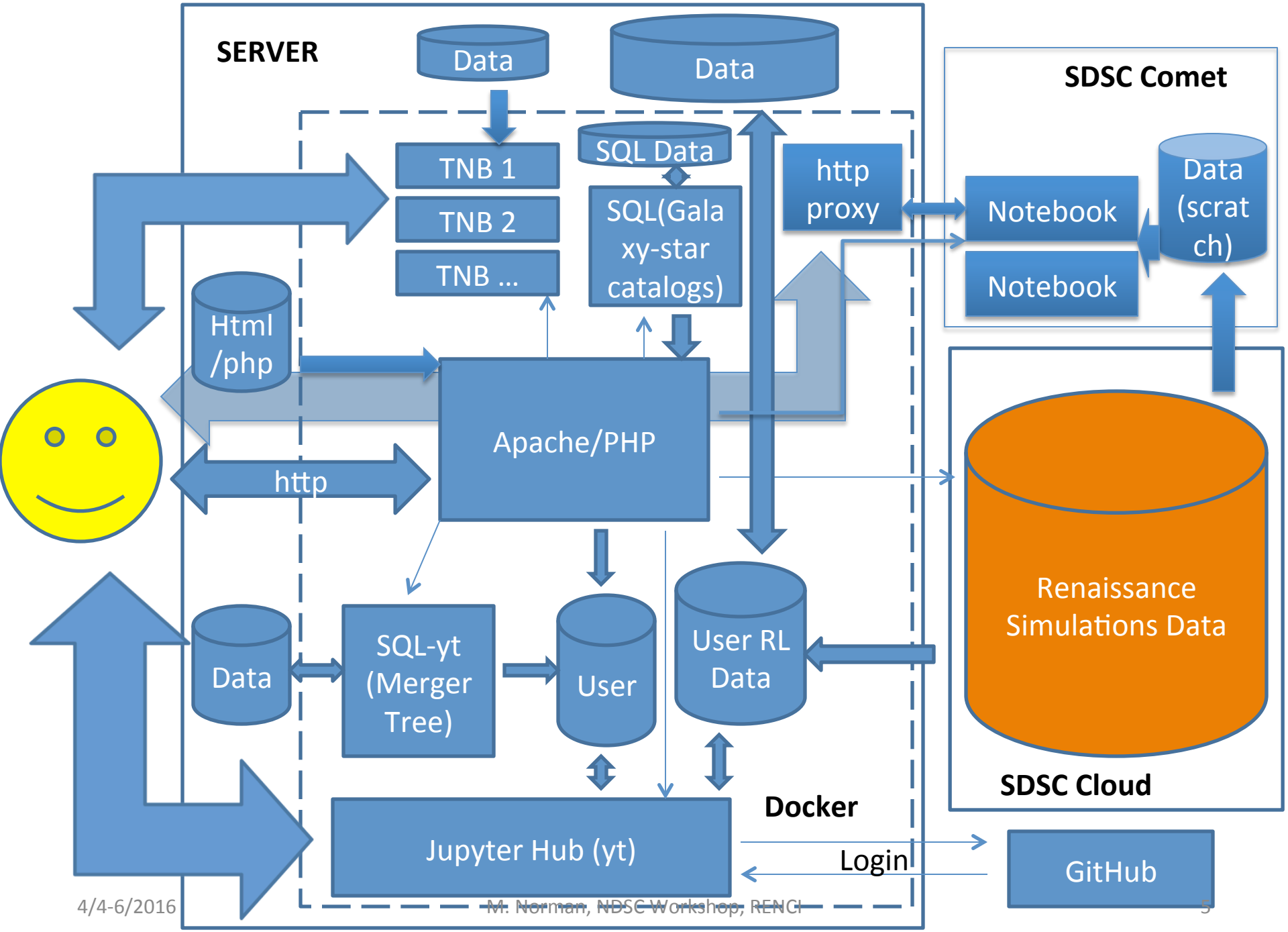
Renaissance Simulations



- 3 "first principles" simulations of the first generation of galaxies
  - Sample size: ~3000 galaxies, Data size: ~ 70 TB
- Required 100 million cpu-hrs on *Blue Waters* supercomputer and 3 years to complete
- Unique data sets will take years more to fully analyze

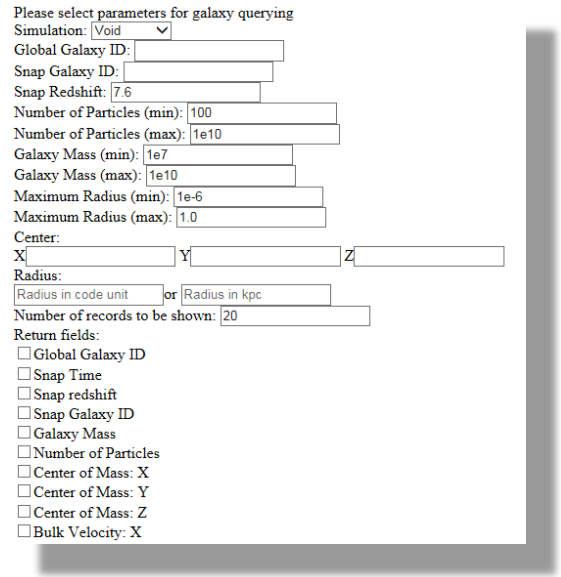
RSL technical approach

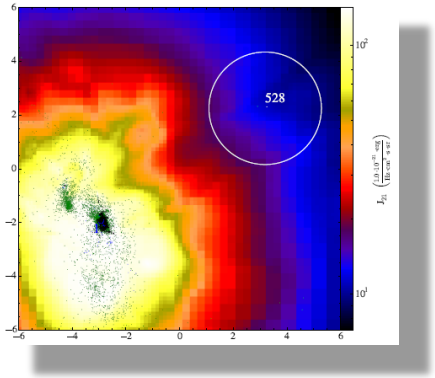
- Collection of containerized data services (Docker)
  - Apache/PHP server
  - Jupyter hub and notebooks
  - Jupyter II single user notebooks
  - Postgres database
  - yt analysis toolkit
- External data volumes (file, object stores)
- GitHub OAuth for Jupyter hub authentication
- JNBs executable on dedicated server or *Comet* supercomputer

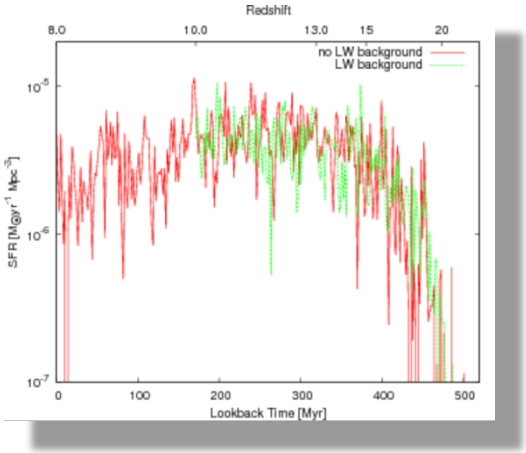

SERVER
Data
Data
SDSC Comet
TNB 1
TNB 2
TNB ...
SQL Data
SQL(Galaxy-star catalogs)
http proxy
Notebook
Notebook
Data (scratch)
Html /php
Apache/PHP
http
Renaissance Simulations Data
SDSC Cloud
Data
SQL-yt (Merger Tree)
User
User RL Data
Docker
Jupyter Hub (yt)
Login
GitHub

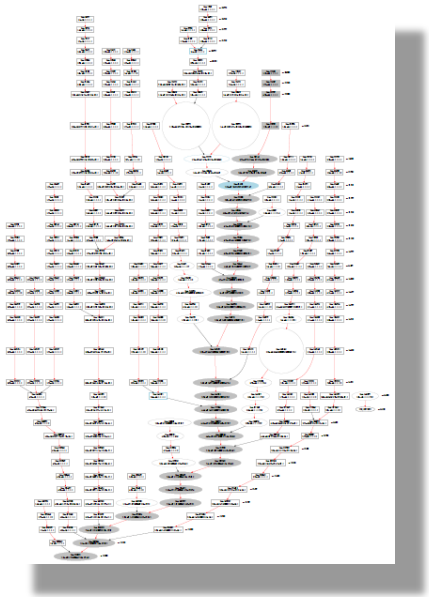
Current data services

- Query galaxy catalogs (3 x NZ)
  - Dark matter, baryonic, and stellar content
- Generate halo merger trees
- Spawn Jupyter NBs for running yt analysis scripts
  - Canned NBs for pubs
  - Copy/modify NBs
- Data download to local or remote directories









Future Directions
(NSF CDS&E proposal, Norman/Turk)

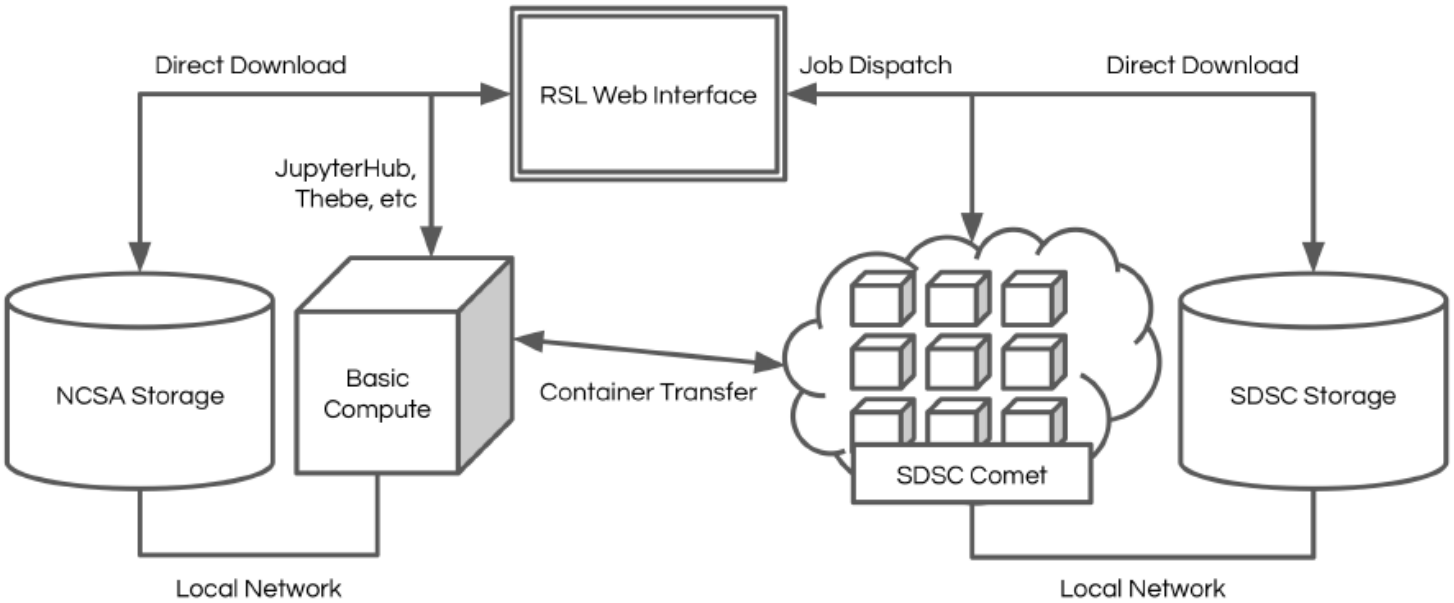


Figure 5: Architectural diagram of the RSL, including initial flow between cyberinfrastructure components.