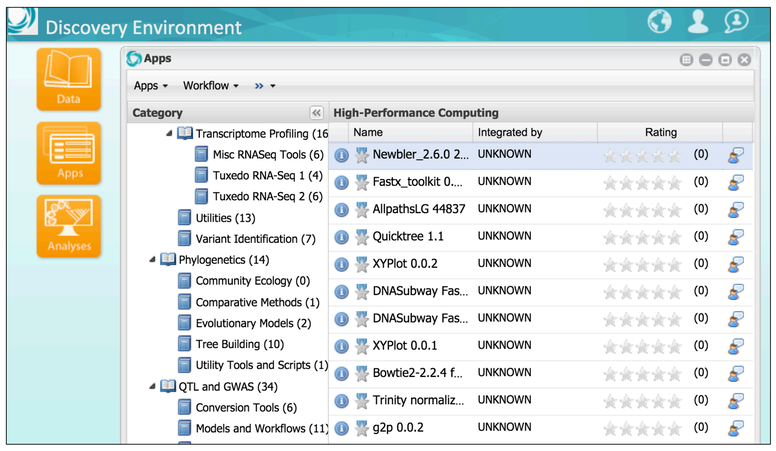
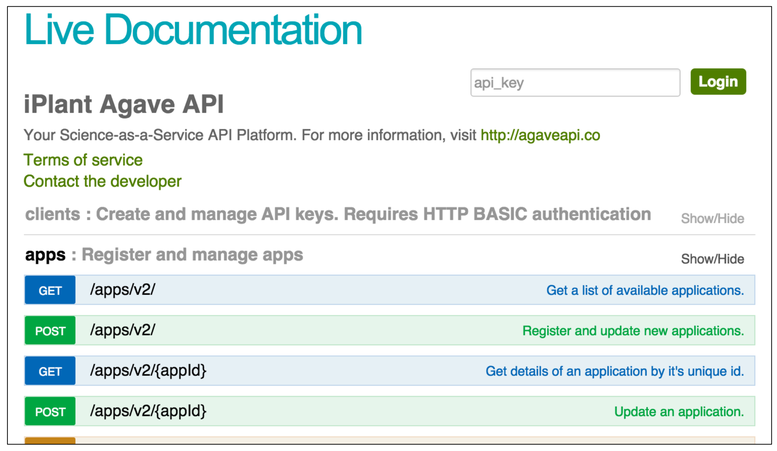
Reproducible computing using iPlant + Agave + Docker

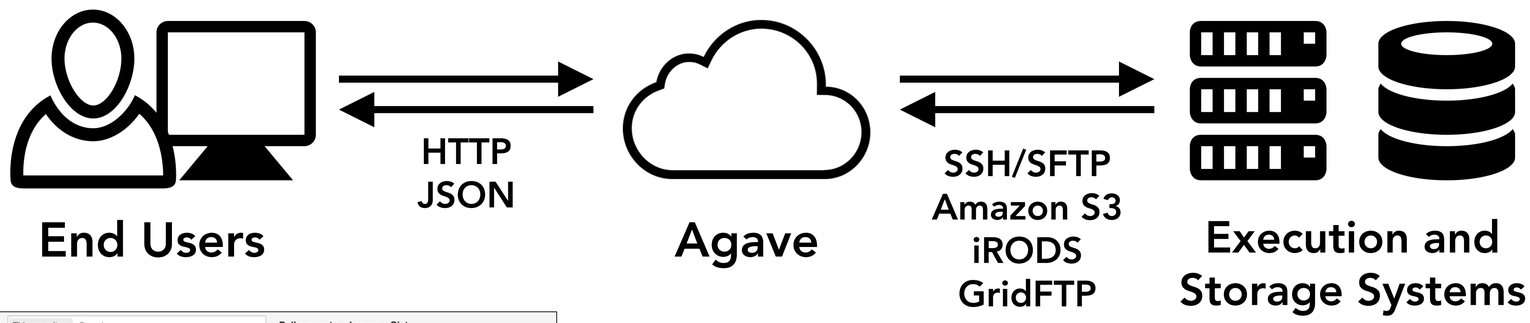
John M. Fonner, Rion Dooley, Matthew W. Vaughn

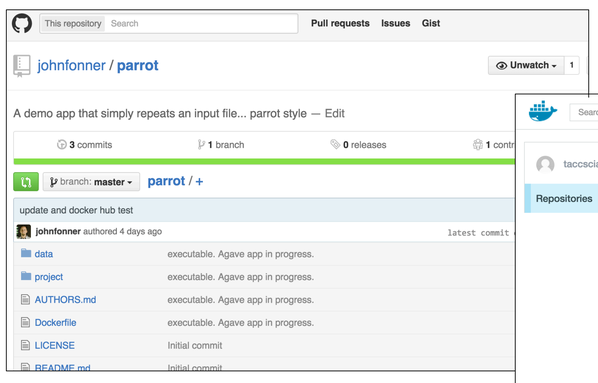
Introduction

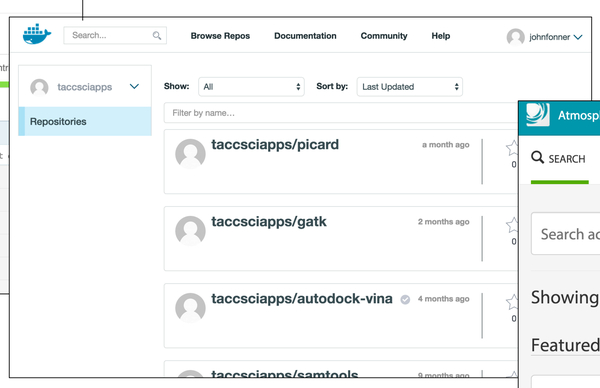
In genomics and other data intensive sciences, collaboration and reproducibility have been severely constrained by three pervasive problems: data availability, software portability, and the onerous computational skillset required to overcome those two things. The iPlant Collaborative and its partners have been developing open source cyberinfrastructure that handles authentication, data movement, job handling, software versioning, provenance, and metadata.

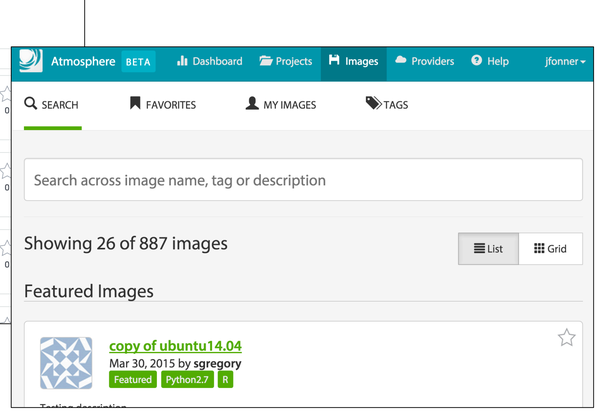












Try it…

- Boot up a system on Amazon, Azure, or Atmosphere: atmo.iplantc.org
- Make sure it has docker-machine: docker.com/docker-machine
- Tell Agave about the system: github.com/agaveapi/docker-provisioning-demo
- Add your own app, or try the simple “parrot” demo: github.com/johnfonner/parrot
- Submit a job through the Discovery Environment: de.iplantc.org

Congrats! You have app version control, automated builds, job histories, and data in a shareable location… and lots of ways to scale up, collaborate, and contribute.

poster available at:
github.com/johnfonner/presentations

iPlant

de.iplantc.org

The iPlant collaborative provides large-scale data and compute resources for the biology community and uses Agave for large-scale data management, job submission, and app management. These resources are provided through powerful user interfaces like the Discovery Environment web portal the Atmosphere cloud computing resource, all of which are free for the scientific community.

Agave API

agaveapi.co

Agave is a science-as-a-service, multi-tenant, RESTful API hosted in the cloud. Users can register their own execution and storage systems in addition to ones Agave provides. They can build up their own app store of scientific codes and workflows and share them with anyone. Agave manages all the back-end complexity of and has SDKs for half a dozen common programming languages.

Docker

docker.com

From Docker’s website, “Docker allows you to package an application with all of its dependencies into a standardized unit”. Among other benefits, Docker containers offer the consistency of virtual machines without the large image sizes. Docker Hub is a public repository for storing and fetching public Docker images, and it integrates with GitHub for automated builds and publication.