iPlant Collaborative

Discovery Environment

Discovery Environment

- Scalable computing platform with Big Data management capability
- Web Accessible
- Extensible – Bring in your own tools

Data Management

- Access iPlant Data Store
- Share with collaborators (with iPlant and as well non-iPlant users)
- Search your own data plus Community data sets
- Automatically detect file and information type
- Data visualization
  - Tree view
  - Comparative Genomics (CoGe)
  - Genome Browser (Ensembl)
  - Tabular Data view
- Metadata management

Apps

- Over 450+ scientific apps with documentation
- No need of computer / command-line / Programming knowledge to use these apps
- Bring in your own command line tools and publish them
- Easy to use WYSIWYG drag-n-drop interface to build your own app.
- Ability to run your tools on HPC @ TACC
- Create pipeline by stringing your apps together into a linear workflow

Manage your analyses

- DE will stage your data and bring back results to your workspace
- All parameters and inputs are saved - Your work is reproducible
- Receive notifications when your analyses completes

Road Map

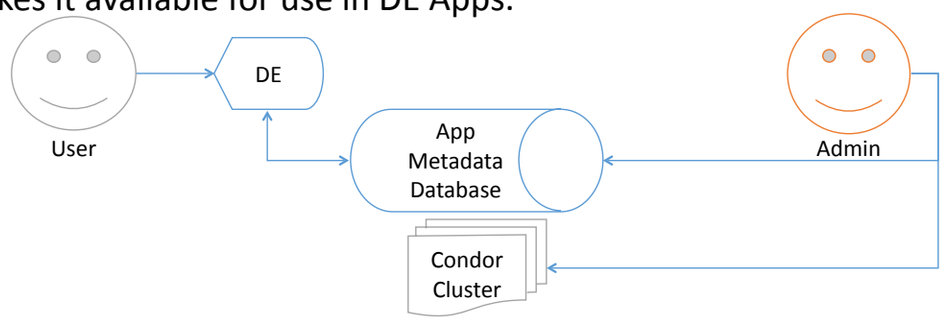
- Support for Batching – Submits 100’s of analyses with click of a button
- Extensive metadata management capabilities
- Projects View
- Support for more 3rd Party tools
  - Fig share
  - Google Drive
  - Drop box
- Public API – Build your own DE
- Tags, Comments
- Complex workflows – Conditional / iteration

Installing and Using Apps in the DE

- User makes a ‘tool request’ to have a new script or binary available for use in the DE.
- The tool is installed in the cluster and registered with the DE.
- The tool is now available for use in DE Apps.
- The user creates a new App, selecting the new tool and adding command line parameters, inputs, and outputs that will be required to run the tool on the cluster.
- The private App may now be run by the user that created it.
- Once the App creator is satisfied with it, the App can be make public and shared with other users.

Installing a tool in the DE for use in Apps

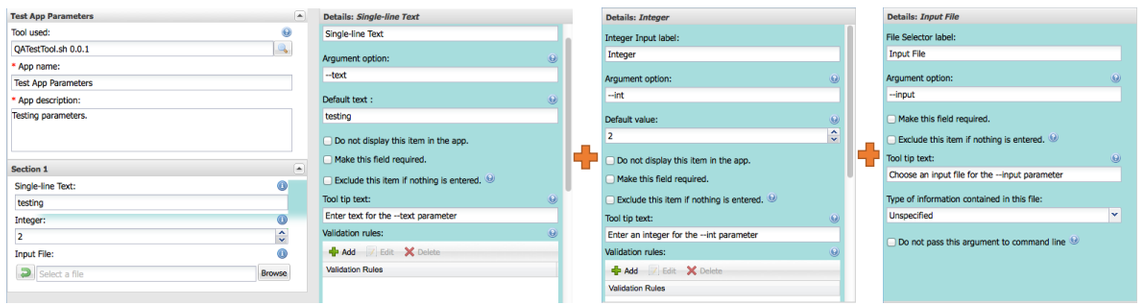
- A user may make a ‘tool request’ to have a new script or binary available for use in the DE.
- The tool is installed in the cluster and registered with the DE, which makes it available for use in DE Apps.



Creating a DE App from an available Tool

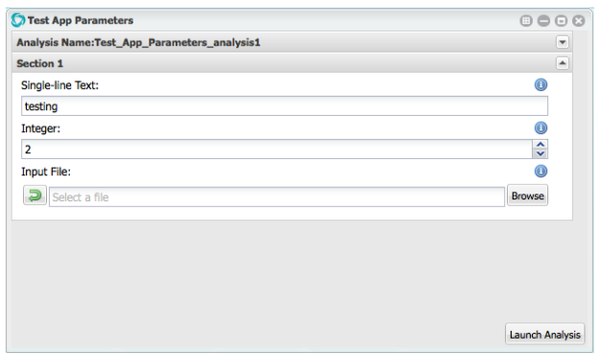
A user creates a new App, selecting the new tool and adding command line parameters, inputs, and outputs that will be required to run the tool on the cluster.

So the command
`QATestTool.sh --text 'testing' --int '2' --input 'test_input.txt'`
becomes an App like this:



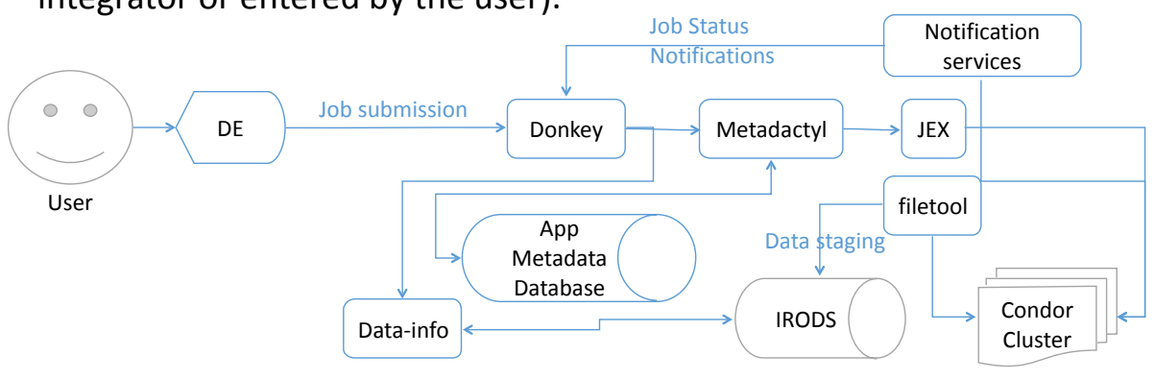
Running a DE App

- When the App is run, inputs are staged onto the node where the job will run and the tool is executed with the App's parameters (set by the integrator or entered by the user).



Running a DE App

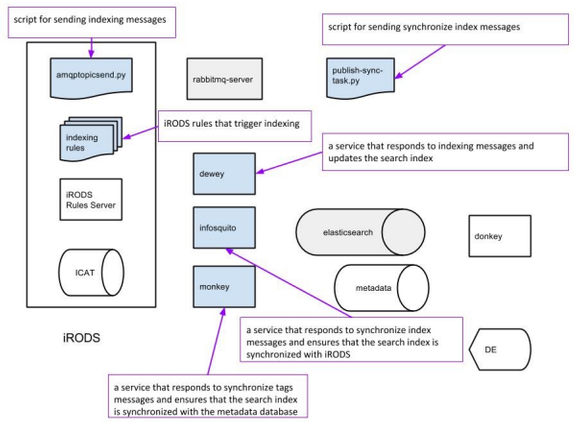
- When the App is run, inputs are staged onto the node where the job will run and the tool is executed with the App's parameters (set by the integrator or entered by the user).



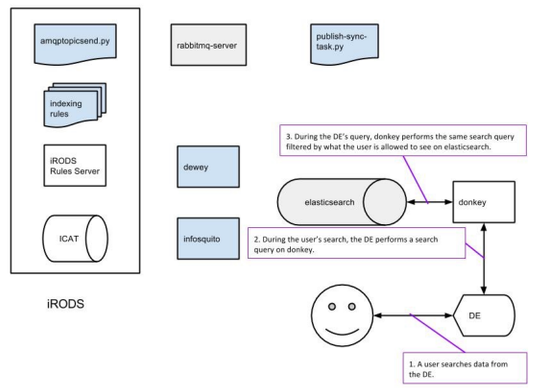
Search Work Flows

- User Search
- Tag Suggestion
- Create and Attach Tag
- Modify Tag
- Attach Existing Tag
- Detach Tag
- Synchronize Tags with Search Index
- Synchronize Data with Search Index
- Incremental Data Indexing

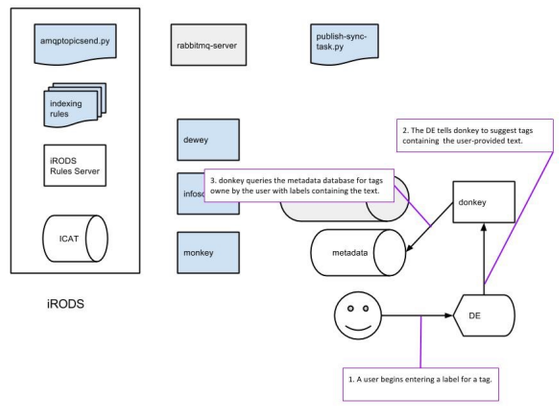
Search Components



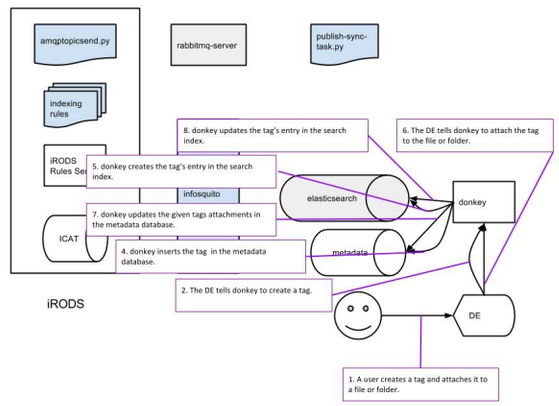
Search Flow


amqptopicsend.py
indexing rules
iRODS Rules Server
ICAT
iRODS
rabbitmq-server
dewey
infosquito
publish-sync-task.py
3. During the DE's query, donkey performs the same search query filtered by what the user is allowed to see on elasticsearch.
elasticsearch
donkey
2. During the user's search, the DE performs a search query on donkey.
DE
1. A user searches data from the DE.

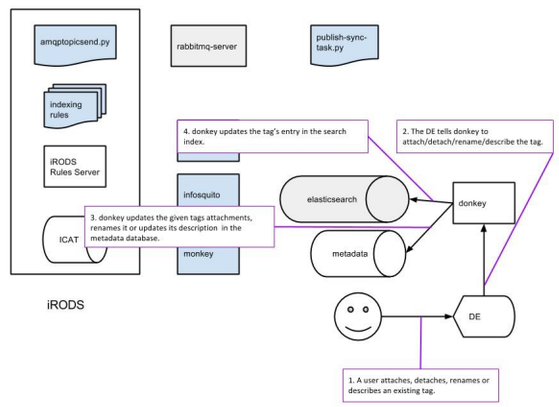
Tag Suggest Flow



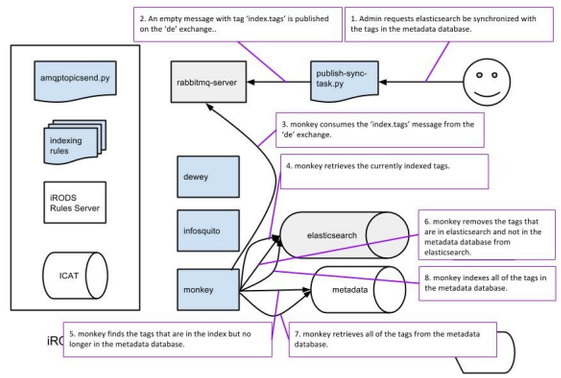
Create and Attach Tag Flow



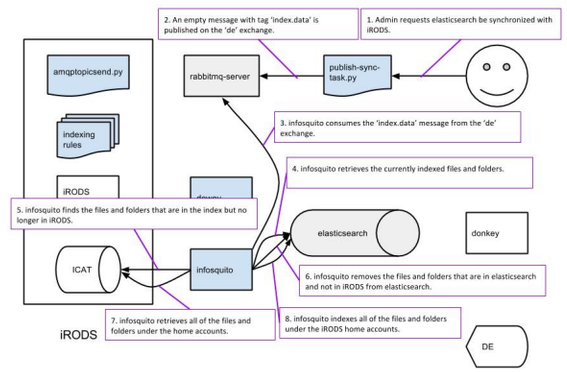
Attach/Detach/Modify Existing Tag Flow



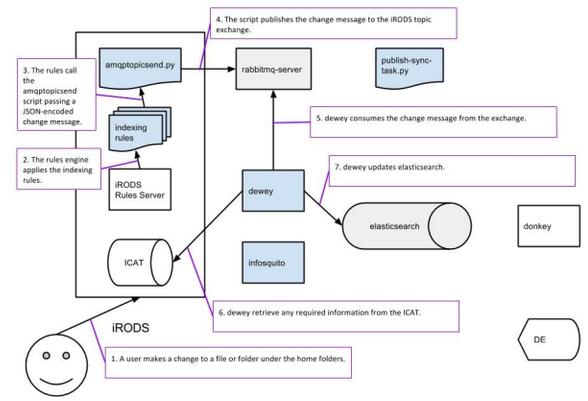
Tags Synchronization Flow



Data Synchronization Flow



Incremental Data Indexing Flow



Need more help?

ask.iplantcollaborative.org

The iPlant Collaborative is funded by a grant from the National Science Foundation Plant Cyberinfrastructure Program (#DBI-0735191).