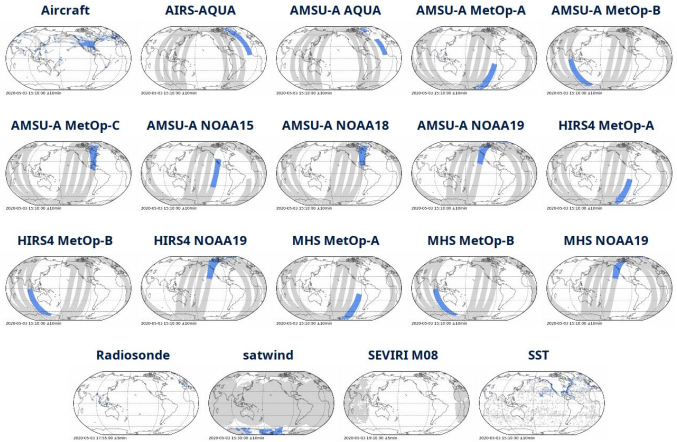
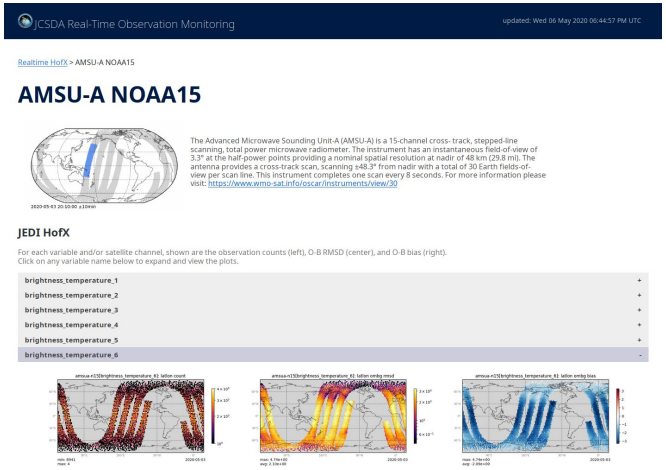
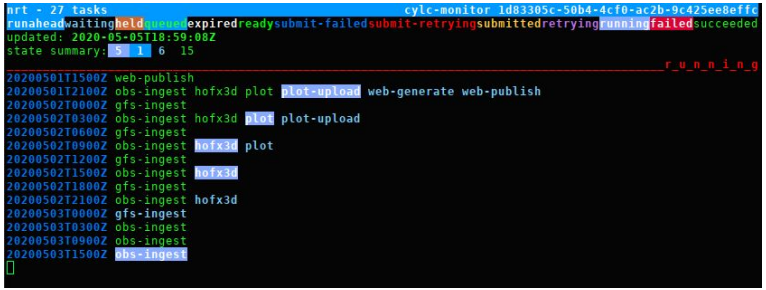
JEDI-Rapids:

Near-Realtime H(X)

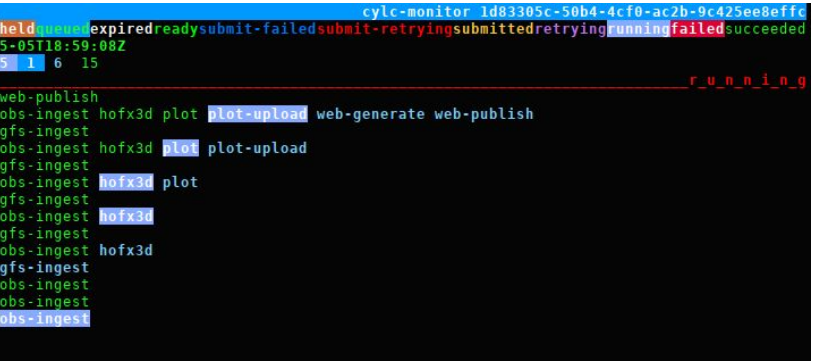


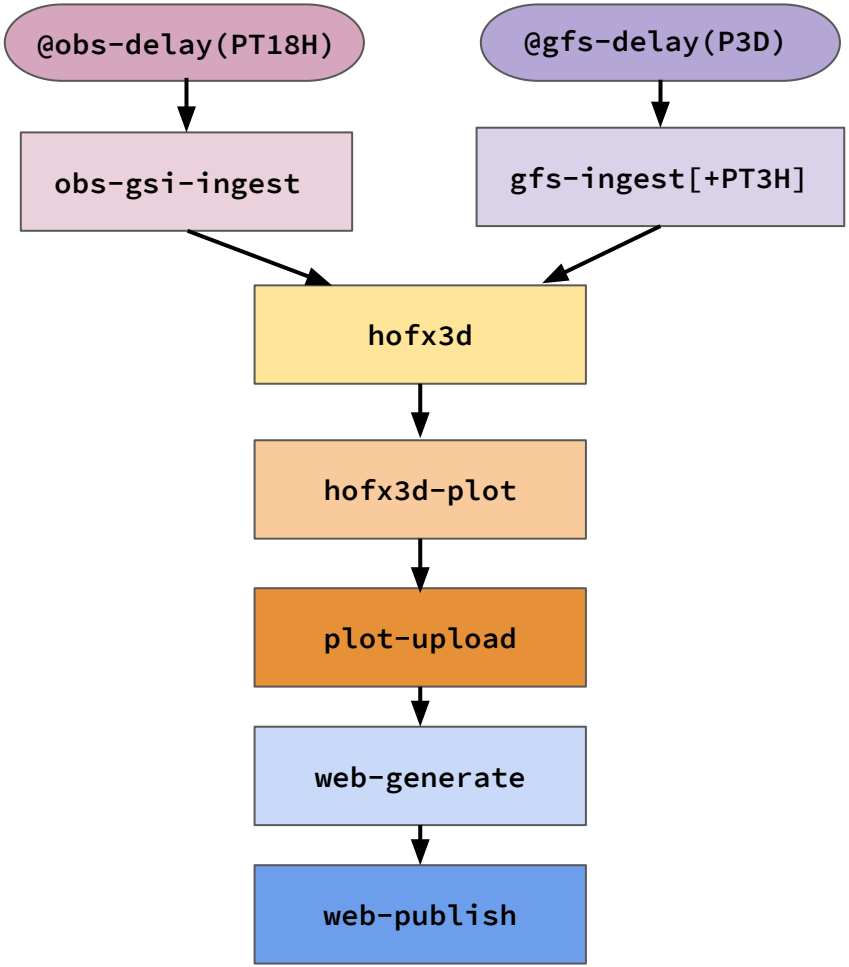




NRT Website Cylc-workflow

```
[scheduling]
    max active cycle points = 14
    [[xtriggers]]
        obs_delay = wall_clock(offset=PT12H)
        gfs_delay = wall_clock(offset=P3D)
    [[dependencies]]
        [[[ T00,T06,T12,T18 ]]]
            graph =  @gfs_delay => gfs-ingest
        [[[ T03,T09,T15,T21 ]]]
            graph = """
                    @obs_delay => obs-ingest
                    gfs-ingest[+PT3H] & obs-ingest => hofx3d
                    hofx3d => plot => plot-upload => web-generate => web-publish
                    web-publish[-PT6H] => web-generate
                    web-publish[+PT6H] => !WWWGEN
                    """
```

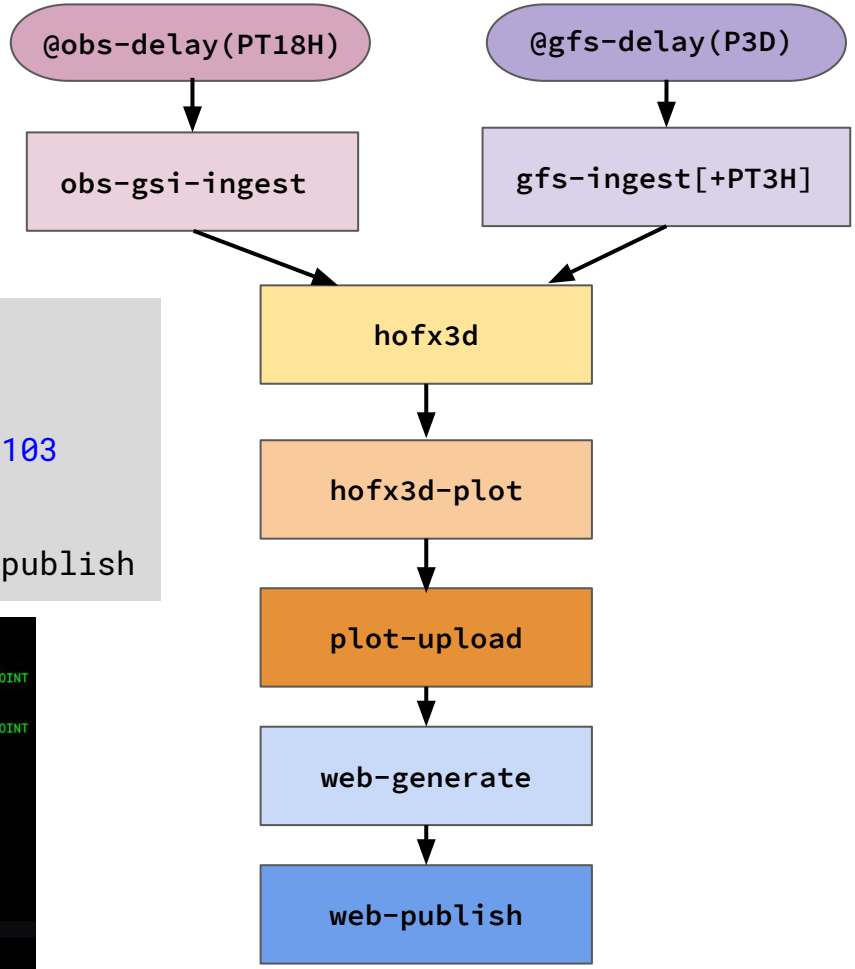




NRT Website Cylc-workflow

```
$ jedi-rapids run gfs-ingest-tar -t 050106
$ jedi-rapids run obs-gsi-ingest -t 050103
$ jedi-rapids run hofx3d -t 050103
$ jedi-rapids run obs-gsi-hofx3d-plot -t 050103
$ jedi-rapids run plot-upload-s3 -t 050103
$ jedi-rapids run web-nrt -t 050103
$ jedi-rapids run web-nrt -t 050103 --phase publish
```

```
#Tasks
[[obs-ingest]]
    inherit = INGEST
    script = $RAPIDS_SUBMIT -T 20 run obs-gsi-ingest --source $INGEST --target $DATA -t $CYLC_TASK_CYCLE_POINT
[[gfs-ingest]]
    inherit = INGEST
    script = $RAPIDS_SUBMIT -T 10 run gfs-ingest-tar --source $INGEST --target $DATA -t $CYLC_TASK_CYCLE_POINT
[[hofx3d]]
    inherit = PROCESS
    script = $RAPIDS_SUBMIT -T 30 -N $NODES run hofx3d --target $ANALYSIS -t $CYLC_TASK_CYCLE_POINT
[[plot]]
    inherit = PROCESS
    script = $RAPIDS_SUBMIT -T 20 run obs-gsi-hofx3d-plot --target $ANALYSIS -t $CYLC_TASK_CYCLE_POINT
[[plot-upload]]
    inherit = PROCESS
    script = $RAPIDS_SUBMIT -T 20 run plot-upload-s3 -t $CYLC_TASK_CYCLE_POINT
[[web-generate]]
    inherit = WWWGEN
    script = jedi-rapids run web-nrt --phase generate -t $CYLC_TASK_CYCLE_POINT
[[web-publish]]
    inherit = WWWGEN
    script = jedi-rapids run web-nrt --phase publish -t $CYLC_TASK_CYCLE_POINT
```



NRT Website Cylc-workflow

JCSDA Real-Time Monitoring

Model: GFS Resolution: c768
Obs Sources: [GSI] OOPS Products: [hofx3d]

Assimilation Window Width: PT6H Begin: 2020-05-03T15:00:00

NRT Website Cylc-workflow

JCSDA Real-Time Monitoring

Model: GFS Resolution: c768
Obs Sources: [GSI] OOPS Products: [hofx3d]

Assimilation Window Width: PT6H Begin: 2020-05-03T15:00:00

```
molah@s4-submit ~/.jedi-rapids/log/s4/slurm $ cd ~/.jedi-rapids/log/s4/gfs-ingest-tar/2020_0
2020_0429/ 2020_0430/ 2020_0501/ 2020_0504/ 2020_0505/ 2020_0506/ 2020_0507/
molah@s4-submit ~/.jedi-rapids/log/s4/slurm $ cd ~/.jedi-rapids/log/s4/gfs-ingest-tar/2020_0507/run.20200507_
run.20200507_000009/ run.20200507_060009/ run.20200507_120010/ run.20200507_122652/ run.20200507_122842/ run.20200507_1228
45/
molah@s4-submit ~/.jedi-rapids/log/s4/slurm $ ls ~/.jedi-rapids/log/s4/gfs-ingest-tar/2020_0507/run.20200507_000009/
gfs-ingest-tar.s4.20200504_0000Z.log
molah@s4-submit ~/.jedi-rapids/log/s4/slurm $ cat ~/.jedi-rapids/log/s4/gfs-ingest-tar/2020_0507/run.20200507_000009/gfs-i
ngest-tar.s4.20200504_0000Z.log
2020-05-07 00:00:09,537 [INFO]: Run logging started: Level:info Echo:True File:/home/molah/.jedi-rapids/log/s4/gfs-ingest-
tar/2020_0507/run.20200507_000009/gfs-ingest-tar.s4.20200504_0000Z.log
2020-05-07 00:00:09,546 [INFO]: [LMod] Load system environment: intel19-impi
2020-05-07 00:00:09,546 [INFO]: [LMod] Load modules: jedi/intel19-impi
2020-05-07 00:00:10,282 [INFO]: *** Run phase: ingest
2020-05-07 00:00:10,298 [INFO]: GFS Ingest -- Date:2020-05-04 00:00:00 Resolution: c768 Source:['/home/molah/eliu-nrt'] De
st:['/data/users/molah/jcsda/jedirapids/data']
2020-05-07 00:00:10,298 [INFO]: GFS Ingest -- Source Folder:/home/molah/eliu-nrt/2020050400_fcst.tar -> Target Folder:/dat
a/users/molah/jcsda/jedirapids/data/gfs/master/bg/c768/2020_05/20200504_0000Z
2020-05-07 00:00:10,328 [INFO]: Untar: '/home/molah/eliu-nrt/2020050400_fcst.tar' -> '/scratch/users/molah/jedirapids/scra
tch/gfs-ingest-tar/20200504_0000Z/master/2020050400_fcst'
2020-05-07 00:00:10,370 [ERROR]: [JEDIApp] Caught exception: Unable to run:
'tar -xivf /home/molah/eliu-nrt/2020050400_fcst.tar --wildcards --no-anchored 20200504.000000* --xform='s#^.+/##x''
2020-05-07 00:00:10,370 [ERROR]: ***RapidsExecutionError***: Unable to run:
'tar -xivf /home/molah/eliu-nrt/2020050400_fcst.tar --wildcards --no-anchored 20200504.000000* --xform='s#^.+/##x''
molah@s4-submit ~/.jedi-rapids/log/s4/slurm $ tar -tvf /home/molah/eliu-nrt/2020050400_fcst.tar
drwxr-sr-x Emily.Liu/stmp    0 2020-05-06 07:01 ./2020050400/
molah@s4-submit ~/.jedi-rapids/log/s4/slurm $
```

NRT Website Cylc-workflow

```
-rw-r--r-- 1 eliu domain users 40004270080 May  4 08:42 2020050200_fcst.tar
-rw-r--r-- 1 eliu domain users  1141729280 May  2 09:21 2020050200_obs.tar
-rw-r--r-- 1 eliu domain users 40004270080 May  4 08:44 2020050206_fcst.tar
-rw-r--r-- 1 eliu domain users  7117209600 May  2 15:24 2020050206_obs.tar
-rw-r--r-- 1 eliu domain users 40004270080 May  4 07:18 2020050212_fcst.tar
-rw-r--r-- 1 eliu domain users  8020183040 May  2 21:15 2020050212_obs.tar
-rw-r--r-- 1 eliu domain users 40004270080 May  4 07:25 2020050218_fcst.tar
-rw-r--r-- 1 eliu domain users 10780590080 May  3 03:18 2020050218_obs.tar
-rw-r--r-- 1 eliu domain users 40004270080 May  5 07:17 2020050300_fcst.tar
-rw-r--r-- 1 eliu domain users 10247546880 May  3 10:09 2020050300_obs.tar
-rw-r--r-- 1 eliu domain users 40004270080 May  5 07:18 2020050306_fcst.tar
-rw-r--r-- 1 eliu domain users  7153756160 May  3 15:13 2020050306_obs.tar
-rw-r--r-- 1 eliu domain users 40004270080 May  5 07:53 2020050312_fcst.tar
-rw-r--r-- 1 eliu domain users  7450859520 May  3 21:13 2020050312_obs.tar
-rw-r--r-- 1 eliu domain users 40004270080 May  5 07:35 2020050318_fcst.tar
-rw-r--r-- 1 eliu domain users 11408373760 May  4 03:14 2020050318_obs.tar
-rw-r--r-- 1 eliu domain users       10240 May  6 14:35 2020050400_fcst.tar
-rw-r--r-- 1 eliu domain users 10057461760 May  4 09:33 2020050400_obs.tar
-rw-r--r-- 1 eliu domain users       10240 May  6 14:35 2020050406_fcst.tar
-rw-r--r-- 1 eliu domain users  6951290880 May  4 15:18 2020050406_obs.tar
-rw-r--r-- 1 eliu domain users       10240 May  6 14:35 2020050412_fcst.tar
-rw-r--r-- 1 eliu domain users  7671306240 May  4 21:17 2020050412_obs.tar
-rw-r--r-- 1 eliu domain users       10240 May  6 14:35 2020050418_fcst.tar
-rw-r--r-- 1 eliu domain users 11659479040 May  5 03:50 2020050418_obs.tar
-rw-r--r-- 1 eliu domain users       10240 May  7 07:01 2020050500_fcst.tar
-rw-r--r-- 1 eliu domain users 10129162240 May  5 09:13 2020050500_obs.tar
-rw-r--r-- 1 eliu domain users       10240 May  7 07:01 2020050506_fcst.tar
-rw-r--r-- 1 eliu domain users  7046082560 May  6 15:30 2020050506_obs.tar
-rw-r--r-- 1 eliu domain users       10240 May  7 07:01 2020050512_fcst.tar
-rw-r--r-- 1 eliu domain users  8011079680 May  6 16:21 2020050512_obs.tar
-rw-r--r-- 1 eliu domain users       10240 May  7 07:01 2020050518_fcst.tar
-rw-r--r-- 1 eliu domain users 12116869120 May  6 16:10 2020050518_obs.tar
-rw-r--r-- 1 eliu domain users 10316277760 May  6 15:18 2020050600_obs.tar
-rw-r--r-- 1 eliu domain users  7148789760 May  6 15:39 2020050606_obs.tar
-rw-r--r-- 1 eliu domain users  7920629760 May  6 21:13 2020050612_obs.tar
-rw-r--r-- 1 eliu domain users 11211233280 May  7 03:13 2020050618_obs.tar
-rw-r--r-- 1 eliu domain users  9778636800 May  7 09:13 2020050700_obs.tar
drwxrwxr-x 2 eliu domain users       51712 May  7 08:30 gps_bufr/
molah@s4-submit ~/.jedi-rapids/log/s4/slurm $
```

JCSDA Real-Time Monitoring

Model: GFS Resolution: c768
Obs Sources: [GSI] OOPS Products: [hofx3d]

Assimilation Window Width: PT6H Begin: 2020-05-03T15:00:00

JEDI-Rapids [github repo]

Release v0.1 (beta)

README.rst

JEDI-Rapids: JEDI Data Assimilation Workflow Applications

JEDI-Rapids is the command line interface to the JEDI Data Assimilation project. The JEDI-Rapids system provides a collection of *JEDI-Apps* that drive key processes in the data assimilation workflows. Apps are composeable, configurable and can be run individually or as part of a managed workflow run by a workflow engine like Cylc or ecFlow.

The structure of the JEDI-Rapids system mirrors that of the OOPS in that the JEDI-Rapids apps are generic with respect to model and observation type.

There are also apps to manage, list, and transfer data products such as model backgrounds and observations. Together the with with the core JEDI DA applications, the JEDI-Rapids Apps make it easy to distribute data products from remote decentralized data repositories and build complex, cycling workflows of interacting DA applications.

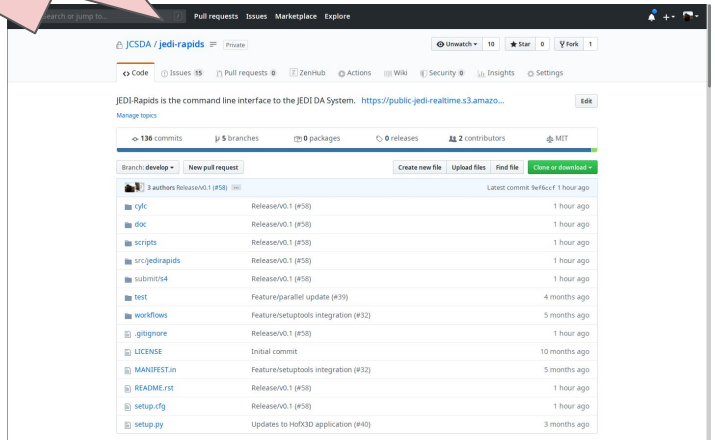
Getting Started

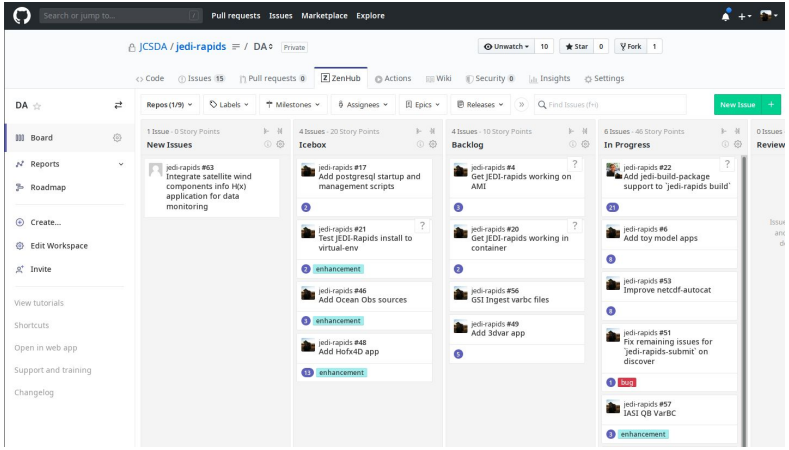
System Requirements

JEDI-Rapids itself is light wight and low-overhead and can run on any Linux or OSX system with Python 3.6 or newer. The JEDI-Rapids system is designed primarily for use in HPC, Cloud, Linux Cluster, or Workstation environments with sufficient resources for large multi-node MPI/OpenMP jobs. However, any system capable of building the `fv3-bundle` set of JEDI packages and running the included CTests, is capable of installing JEDI-Rapids and running apps. For details on JEDI environment setup and prerequisites see the JEDI Building Testing and Running documentation.

Supported systems:

- S4 - Univ. of Wisconsin-Madison (SSEC)
  - S4 JEDI Modules documentation
- Discover - NASA
  - Discover JEDI Modules documentation
- Cheyenne - NCAR
  - Cheyenne JEDI Modules documentation
- AWS EC2





JEDI-Rapids: Live Tutorials

Live Demo

76 lines (46 sloc) | 1.86 KB

Raw Blame History

Tutorial: HofX3D Low-res Quick Start Guide

This tutorial shows how to get a JEDI system up and running quickly using low-resolution GFS background data and smaller observation files. This tutorial can be run on smaller machine or a laptop with limited resources.

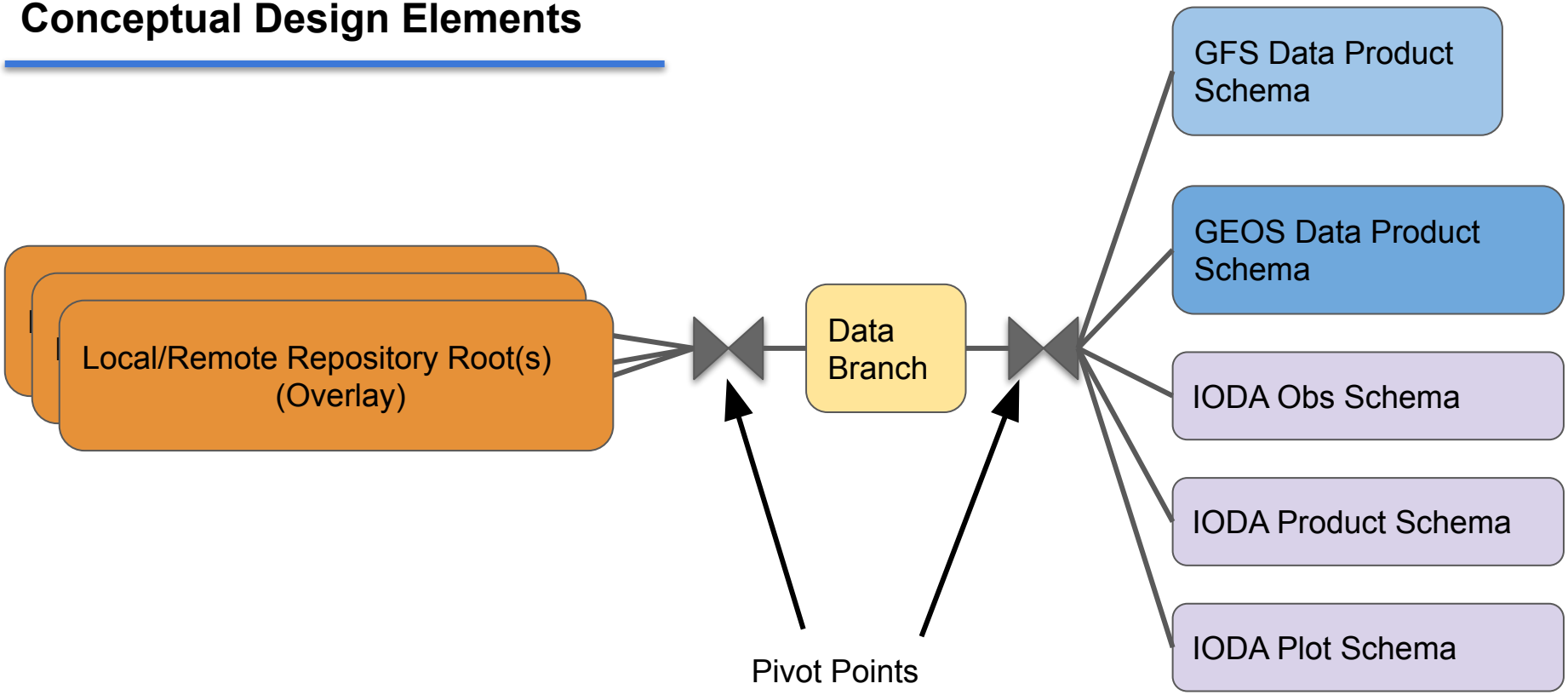
```
mjo@wwyzzerdd ~ $ jedi-rapids sync gfs -t 050206 -r 48 --source s4
2020-05-06 23:31:37,697 [INFO]: GFS Sync -- Source: s4:/data/users/molah/jedirapids/data
2020-05-06 23:31:37,697 [INFO]: GFS Sync -- Target: ['/nfs/olah/home/mjo/scratch2/jedirapids/data', '/nfs/olah/home/mjo/scratch0/
jedirapids/data', '/nfs/olah/home/mjo/glacier0/jedirapids/data']
2020-05-06 23:31:37,699 [INFO]: GFS Sync -- Time:20200502_0600Z Resolution:c48 SourceBranch:master TargetBranch:master
2020-05-06 23:31:37,699 [INFO]: GFS Sync -- Source Folder:gfs/master/bg/c48/2020_05/20200502_0600Z Target Folder:gfs/master/bg/c4
8/2020_05/20200502_0600Z
2020-05-06 23:31:38,206 [INFO]: GFS Sync -- s4:/data/users/molah/jcsda/jedirapids/data/gfs/master/bg/c48/2020_05/20200502_0600Z -
> ['/nfs/olah/home/mjo/scratch2/jedirapids/data/gfs/master/bg/c48/2020_05/20200502_0600Z']
receiving incremental file list
./
20200502.060000.coupler.res
20200502.060000.fv_core.res.tile1.nc
20200502.060000.fv_core.res.tile2.nc
```

Prerequisites

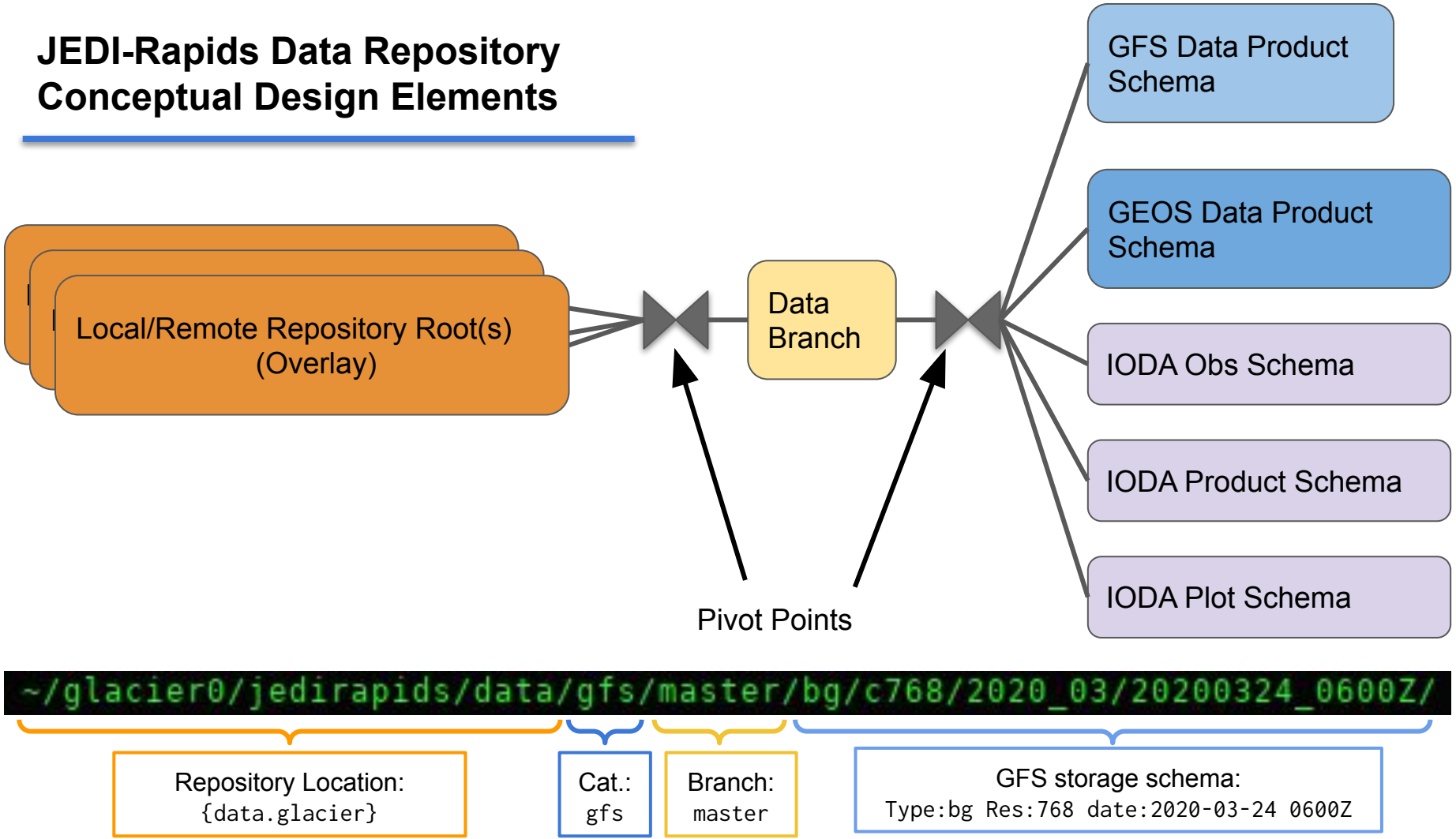
This tutorial assumes you have passed the Build Tutorial and have a working JEDI prefix install to run with.

Check your prefix install name and info with:

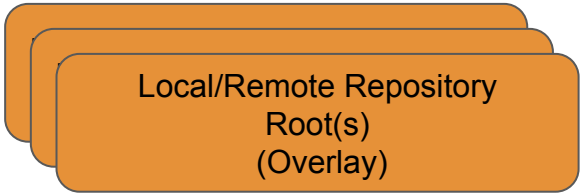
JEDI-Rapids Data Repository Conceptual Design Elements



JEDI-Rapids Data Repository Conceptual Design Elements



JEDI-Rapids Data Repository Repository Selection



Commands use --source and --target arguments that can take variables referring to system-specific repoitory paths as configured in user's system.yaml. Some operations can refer to remote system repository config variables also.

- Local repository -> Local repository transfer

```
$ jedi-rapids sync gfs -t 040106 --source {data.user} --target {data.glacier}
```

- Remote repository -> Local repository transfer

```
$ jedi-rapids sync gfs -t 040106 --source s4:{data}_--target {data.glacier}
```

- Repository can be a filesystem path or a structured list of named repositories

```
$ jedi-rapids sync gfs -t 040106 --source /my/special/datarepo --target {data}
```

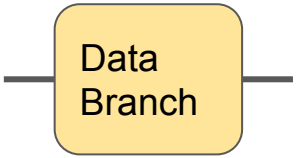
system.yaml - Defines default repository categories and priority-ordered list of repositories filesystem locations

```
path:
  data:
    user: /nfs/olah/home/mjo/scratch2/jedirapids/data
    scratch0: /nfs/olah/home/mjo/scratch0/jedirapids/data
    glacier: /nfs/olah/home/mjo/glacier0/jedirapids/data
  analysis:
    user: /nfs/olah/home/mjo/scratch2/jedirapids/analysis
    glacier: /nfs/olah/home/mjo/glacier0/jedirapids/analysis
```

- Default to system configured paths
  - Target defaults to first item in list
  - Source repositories form an overlay structure. First repo containing data will be used.

```
$ jedi-rapids sync gfs -t 040106 --source s4
```

JEDI-Rapids Data Repository
Data Branches



- Data branches can be any pathlike name:
  - e.g., `develop`, `feature/gfs-forecast`, `test/issue753`, `gfs/reanalysis/2020040100/PT6Z`
- Each data class (gfs, ioda, etc.) has its own branch namespace
- Each data class can define its own naming/usage schema
- All relevant JEDI-Rapids commands have a --branch argument sometimes several:

```
jedi-rapids run hofx3d -t 040601 --model gfs --model-branch develop --obs gsi --obs-branch test-converters --target-branch for_emily/try2
```

- Default branch name is normally master (configurable per data schema)
- Branches are at the top of each repo directory tree (easy to remove or copy or zip, etc.)
- Branches are conceptually combined over a repository list by overlay semantics.
- Concept is distributed and decentralized inspired by git branches

JEDI-Rapids Data Repository
Class-based storage schema

- ❖ “Data Classes” will define their own data naming schema
- ❖ IODA_DB defines window and obs naming schema
- ❖ IODA_DB also defines (obs-related) product naming schema
- ❖ Models each define their own list of data products and naming schema
- ❖ Folder and file schema are defined in YAML files.

Each Data Class provides methods to

- ➢ Format folders and filenames from metadata
- ➢ Enumerate a folder for data products return Pandas DataFrame DB
- ➢ Validate the contents of a data product
- ➢ Enumerate all data products in a repository (with constraints)
- ➢ Combine (fold) data product DataFrames across a multiple repositories

`rapids.model.gfs.GFSModel()` → GFS Data Product Schema

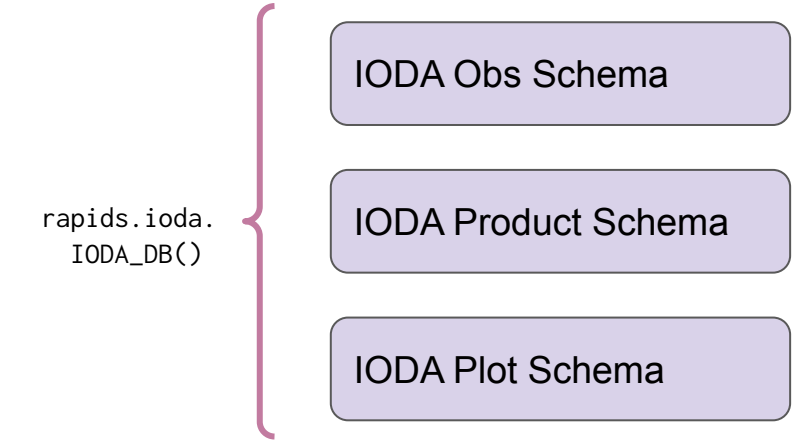
GFS Data Products
bg: warm restart
cs: cold start
fcsd: forecast data
geo: static geometry
ens: ensemble bg

`rapids.model.geos.GEOSModel()` → GEOS Data Product Schema

`rapids.ioda.IODA_DB()` →

- IODA Obs Schema
- IODA Product Schema
- IODA Plot Schema

JEDI-Rapids Data Repository
Class-based storage schema



jedirapids/conf/ioda.yaml

```yaml
ioda:
  obsdb:
    folder: "obs/{obs_source}/{branch}/ioda/{window_width}/{window_begin_date_month}/{window_begin_date_hour}"
    file: "ioda_{obs_source}_{instrument_platform}_{window_width}_{window_begin_date_hour}{extension}"
  prod: #OOPS IODA Data Product types (Model-independent data
    hofx3d:
      path: {{data.analysis}}
      folder: "ioda/{branch}/{model}/{model_resolution}/{obs_source}/{product}/{window_width}/{window_begin_date_month}/{window_begin_date_hour}/"
      file: "{product}_{model}_{model_resolution}_{obs_source}_{instrument_platform_operator}_{window_width}_{window_begin_date_hour}{extension}"
  plot:
    hofx3d:
      path: {{data.analysis}}
      folder: "plot/{branch}/{model}/{model_resolution}/{obs_source}/{product}/{window_width}/{window_begin_date_month}/{window_begin_date_hour}"
      file_prefix: "{product}_{model}_{model_resolution}_{obs_source}_{instrument_platform_operator}_{window_width}_{window_begin_date_hour}_"
```

JEDI-Rapids Data Repository
Class-based storage schema

jedirapids/model/gfs.yaml

```yaml
gfs:
  prod: #Model-dependent data input and/or output products
    bg: # Background aka warm-restarts
      folder: "{{model.name}}/{branch}/bg/{resolution}/{date_month}/{date_hour}"
      files: &bg_files
        coupler:  "{datetime}.coupler.res"
        tiles:
          core:     "{datetime}.fv_core.res.tile{tile_num:d}.nc"
          sfc_wnd:  "{datetime}.fv_srf_wnd.res.tile{tile_num:d}.nc"
          tracer:   "{datetime}.fv_tracer.res.tile{tile_num:d}.nc"
          phy_data: "{datetime}.phy_data.tile{tile_num:d}.nc"
          sfc_data: "{datetime}.sfc_data.tile{tile_num:d}.nc"
        optional_tiles: #These tiles are not produced upon regrid
        - phy_data
    ens:
      folder: "{{model.name}}/{branch}/ens/{resolution}/{date_month}/{date_hour}/"
      mem_subfolder: "mem.{memid}"
      files:
        <<: *bg_files  #Use bg.files as file description for ens mems
    cs: #Forecast cold start
      folder: "{{model.name}}/{branch}/cs/{resolution}/{date_month}/{date_hour}"
      files:
        gfs_ctrl: "gfs_ctrl.nc"
        tiles:
          gfs_data: "gfs_data.tile{tile_num:d}.nc"
          sfc_data: "sfc_data.tile{tile_num:d}.nc"
    geo: #Fixed geometry files
      folder: "{{model.name}}/{branch}/geo/{resolution}"
      files:
        grid_spec: "grid_spec.nc"
        tiles:
          grid: "{res}_grid.tile{tile_num:d}.nc"
          oro_data: "oro_data.tile1.nc"
    fcstd: #Long-term/fixed forecast data
      folder: "{{model.name}}/{branch}/fcstd/{date}"
      contents: "{{model.name}}_fcstd_{date}.CONTENTS.yaml"
```

GFS Data Product Schema

GFS Data Products
bg: warm restart
cs: cold start
fcsd: forecast data
geo: static geometry
ens: ensemble bg

JEDI-Rapids Data Repository

`jedi-rapids list`

```
mjo@wwwyzzerdd ~ $ jedi-rapids list --help
Usage: jedi-rapids list [OPTIONS] COMMAND [ARGS]...

  List entities: apps, backgrounds, obs, etc.

Options:
  --help  Show this message and exit.

Commands:
  apps    List available Apps.
  gfs     List GFS Data products.
  obs     List IODA Observations
  prefix  List prefixes built and configured with `jedi-rapid build`...
  prod    List IODA Data Products
```

`jedi-rapids list`: A general purpose command for convenient command-line information display.

- Options for listing data products as well as other conceptual items (prefix, apps, etc.)

JEDI-Rapids Data Repository
`jedi-rapids list gfs`

```
mjo@wwwyzzerdd ~ $ jedi-rapids list gfs --help
Usage: jedi-rapids list gfs [OPTIONS] [[bg|ens|cs|geo|fcstd]]

  List GFS Data products.

Options:
  -b, --branch <branch>           Data branch name.  [default: master]
  -t, --time <daterange>          Year (e.g., 2020), Month (e.g., 08, 201908),
                                  date (e.g., 0901, 20190901), or date range
                                  <start>:<end> [Multiple=true]
  -r, --resolution <resolution>   List selected resolution. [Multiple=True]
  --source <path>                 Source location: '/path/to/src' OR templated:
                                  '{{path_var}}'  [default: {{data}}]
  --help                          Show this message and exit.
```

`jedi-rapids list gfs`: Provided by GFS to list gfs-specific data products with GFS-defined format, arguments and options.

- Uses a concept of “type” to allow for different categories of data products (with different naming schemas)

JEDI-Rapids Data Repository
`jedi-rapids list gfs`

```
mjo@wwwyzzerdd ~/work/github/jedi-rapids $ jedi-rapids list gfs -t 0324:0326
================ List Backgrounds Model: gfs  Branch: master ================

[[ Begin: 2020-03-24 00:00:00 --- End: 2020-03-26 00:00:00 ]]
2020-03-24:    12:[00,    ,    ,    ]
               24:[00, 06, 12, 18]
               48:[00, 06, 12, 18]
               96:[00, 06, 12, 18]
              192:[00, 06, 12, 18]
              384:[00, 06, 12, 18]
              768:[00, 06, 12, 18]
2020-03-25:    24:[00, 06, 12, 18]
               48:[00, 06, 12, 18]
               96:[00, 06, 12, 18]
              192:[00, 06, 12, 18]
              384:[00, 06, 12, 18]
              768:[00, 06, 12, 18]
2020-03-26:    24:[00, 06, 12, 18]
               48:[00, 06, 12, 18]
               96:[00, 06, 12, 18]
              192:[00, 06, 12,    ]
              384:[00, 06, 12, 18]
              768:[00, 06, 12, 18]

[[ Data Path index ]]
[0]: path.data.user='/nfs/olah/home/mjo/scratch2/jedirapids/data'
[1]: path.data.scratch0='/nfs/olah/home/mjo/scratch0/jedirapids/data'
[2]: path.data.glacier='/nfs/olah/home/mjo/glacier0/jedirapids/data'

================================================================================
```

`jedi-rapids list gfs`: Provided by GFS to list gfs-specific data products with GFS-defined format, arguments and options.

- Uses a concept of “type” to allow for different categories of data products (with different naming schemas)

JEDI-Rapids Data Repository
`jedi-rapids list obs`

```
mjo@wwwyzzerdd ~ $ jedi-rapids list obs --help
Usage: jedi-rapids list obs [OPTIONS] [[godae|gsi|cosmic]]

  List IODA Observations

Options:
  -b, --branch <branch>       Data branch name.  [default: master]
  -t, --time <daterange>      Year (e.g., 2020), Month (e.g., 08, 201908), date
                              (e.g., 0901, 20190901), or date range
                              <start>:<end> [Multiple=true]
  -d, --duration <duration>   Window width (duration)  [default: PT6H]
  --instrument <instrument>   Instrument name
  --platform <platform>       Platform name
  --filetype [netcdf|odb]     IODA file type.  [default: netcdf, odb]
  --source <path>             Source location: '/path/to/src' OR templated:
                              '{{path_var}}'  [default: {{data}}]
  --help                      Show this message and exit.
```

`jedi-rapids list obs`:
Provided by `IODA_DB` to list observations from a “obs source”

- Defines concepts:
  - obs_source
  - instrument
  - platform
- All data from same obs_source and time window is in same directory (easy to symlink for runs)
- Individual obs can be found independently in different repositories in the overlay

JEDI-Rapids Data Repository

`jedi-rapids list obs`

`jedi-rapids list obs`:
Provided by `IODA_DB` to list observations from a “obs source”

- Defines concepts:
  - obs_source
  - instrument
  - platform
- All data from same obs_source and time window is in same directory (easy to symlink for runs)
- Individual obs can be found independently in different repositories in the overlay

JEDI-Rapids Data Repository
`jedi-rapids list prod`

```
mjo@wwwyzzerdd ~ $ jedi-rapids list prod --help
Usage: jedi-rapids list prod [OPTIONS] [[hofx3d]]

  List IODA Data Products

Options:
  -m, --model [gfs|l95]           Model  [default: gfs]
  -r, --resolution <resolution>   List selected resolution. [Default=full-res]
  -o, --obs [godae|gsi|cosmic]    Observation: source or source.instrument or
                                  source.instrument.platform or
                                  source.instrument.operator  [default: godae,
                                  gsi, cosmic]
  -b, --branch <branch>           Data branch name.  [default: master]
  -t, --time <daterange>          Year (e.g., 2020), Month (e.g., 08, 201908),
                                  date (e.g., 0901, 20190901), or date range
                                  <start>:<end> [Multiple=true]
  -d, --duration <duration>       Window width (duration)  [default: PT6H]
  --instrument <insturment>       Filter only selected instrument(s) by name.
  --platform <platform>           Filter only selected platform(s) by name.
  --filetype [netcdf|odb]         IODA file type.  [default: netcdf, odb]
  --source <path>                 Source location: '/path/to/src' OR templated:
                                  '{{path_var}}'  [default: {{analysis}}]
  --help                          Show this message and exit.
```

`jedi-rapids list prod`:
Provided by `IODA_DB` to list observation products of JEDI runs

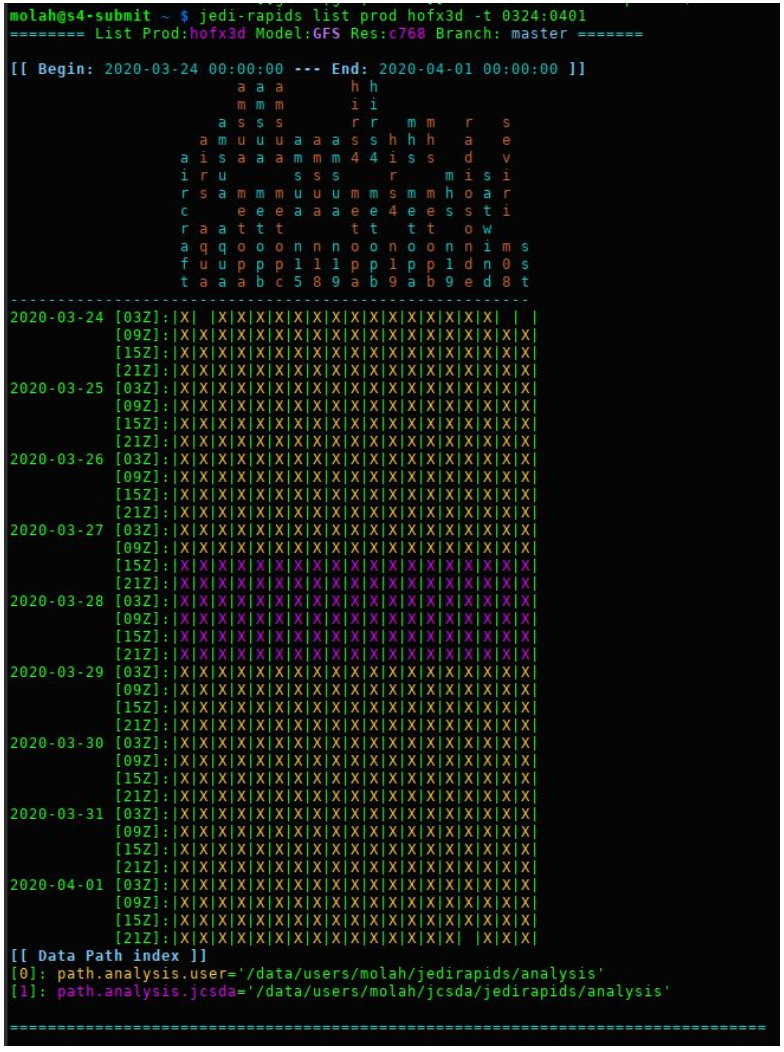
- Indexed on obs and model metadata
- Model-specific data products (e.g. initial conditions) are controlled by the model class. [These are obs-centric products.]

JEDI-Rapids Data Repository

`jedi-rapids list prod`

`jedi-rapids list prod`:
Provided by `IODA_DB` to list observation products of JEDI runs

- Indexed on obs and model metadata
- Model-specific data products (e.g. initial conditions) are controlled by the model class. [These are obs-centric products.]



JEDI-Rapids Data Repository

`jedi-rapids sync`

```
molah@s4-submit ~ $ jedi-rapids sync --help
Usage: jedi-rapids sync [OPTIONS] COMMAND [ARGS]...

  Sync data: backgrounds, obs, etc.

Options:
  --help  Show this message and exit.

Commands:
  gfs   Sync GFS data (backgrounds/coldstart/fcstdata)
  obs   Sync Observations.
  prod  Analysis IODA product sync.
```

`jedi-rapids sync`: A rsync-based tool to synchronize data products between local/remote repositories and branches.

JEDI-Rapids Data Repository

`jedi-rapids sync gfs`

`jedi-rapids sync gfs`:

Transfer GFS data products between local/remote repositories

- Each model provides its own sync command with appropriate arguments/metadata for each product
- Models have more than one data-product type to sync
- Checksum and provenance metadata is maintained
- Options to specify source and target repositories and branch names

```
molah@s4-submit ~ $ jedi-rapids sync gfs --help
Usage: jedi-rapids sync gfs [OPTIONS]

  Sync GFS data (backgrounds/coldstart/fcstdata)

Options:
  --type <datatype>               Data product type, bg, ens, cs, geo, fcstd
                                  [default: bg]
  -t, --time <daterange>          Time <datetime> or time sequence
                                  <start>:<end>:<step> [Multiple=true]
                                  [required]
  --source <url>                  Source URL. '/path/to/src' OR '<machine-
                                  name>[:/path/to/source]'  [default: {{data}}]
  --source-branch <branch-name>   Source data branch name.  [default: master]
  --target <url>                  Target URL. '/path/to/src' OR '<machine-
                                  name>[:/path/to/source]'  [default: {{data}}]
  --target-branch <branch-name>   Target data branch name.  [default: master]
  -r, --resolution <resolution>   List selected resolution. [Multiple=True]
  --overwrite                     Overwrite existing local targets.  [default:
                                  False]
  --cleanup-partial               Cleanup partial (incomplete) transfers.
                                  [default: False]
  --validate-source               Validate source before copy  [default: False]
  --validate-target               Validate target after copy  [default: False]
  --validate-existing-target      Validate existing target  [default: True]
  --stop-if-error                 Stop transfer if errors occur  [default:
                                  False]
  --stop-if-existing              Stop transfer if existing found  [default:
                                  False]
  --quiet                         Be more quiet  [default: False]
  --help                          Show this message and exit.
```

JEDI-Rapids Data Repository

`jedi-rapids sync obs`

`jedi-rapids sync obs`:
Transfer observation ioda files between local/remote repositories

- One sync app works for all observation sources
- Each source provides specification of instrument/platforms in a YAML document
- Handles NetCDF and ODB IODA files

```
molah@s4-submit ~ $ jedi-rapids sync obs --help
Usage: jedi-rapids sync obs [OPTIONS] [[godae|gsi|cosmic]]

  Sync Observations.

Options:
  -t, --time <daterange>          Time <datetime> or time sequence
                                  <start>:<end>:<step> [Multiple=true]
                                  [required]
  -d, --duration <duration>       Window width (duration)  [default: PT6H]
  --source <url>                  Source URL. '/path/to/src' OR '<machine-
                                  name>[:/path/to/source]'  [default: {{data}}]
  --source-branch <branch-name>   Source data branch name.  [default: master]
  --target <url>                  Target URL. '/path/to/src' OR '<machine-
                                  name>[:/path/to/source]'  [default: {{data}}]
  --target-branch <branch-name>   Target data branch name.  [default: master]
  --instrument <insturment>       Instrument name
  --platform <platform>           Platform name
  --filetype [netcdf|odb]         IODA file type.  [default: netcdf, odb]
  --overwrite                     Overwrite existing local targets.  [default:
                                  False]
  --cleanup-partial               Cleanup partial (incomplete) transfers.
                                  [default: False]
  --validate-source               Validate source before copy  [default: False]
  --validate-target               Validate target after copy  [default: False]
  --validate-existing-target      Validate existing target  [default: True]
  --stop-if-error                 Stop transfer if errors occur  [default:
                                  False]
  --stop-if-existing              Stop transfer if existing found  [default:
                                  False]
  --quiet                         Be more quiet  [default: False]
  --help                          Show this message and exit.
```

JEDI-Rapids Data Repository

`jedi-rapids sync prod`

`jedi-rapids sync prod`:
Sync observation-based data products between local/remote repositories.

- One sync app works with all observation sources
- Model and obs metadata is referenced in product folder/file naming
- Need to identify additional data product categories

```
molah@s4-submit ~ $ jedi-rapids sync prod --help
Usage: jedi-rapids sync prod [OPTIONS] [[hofx3d]]

  Analysis IODA product sync.

Options:
  -t, --time <daterange>          Time <datetime> or time sequence
                                  <start>:<end>:<step> [Multiple=true]
                                  [required]
  -d, --duration <duration>       Window width (duration)  [default: PT6H]
  --source <url>                  Source URL. '/path/to/src' OR '<machine-
                                  name>[:/path/to/source]'  [default:
                                  {{analysis}}]
  --source-branch <branch-name>   Source data branch name.  [default: master]
  --target <url>                  Target URL. '/path/to/src' OR '<machine-
                                  name>[:/path/to/source]'  [default:
                                  {{analysis}}]
  --target-branch <branch-name>   Target data branch name.  [default: master]
  -m, --model [l95|gfs]           Model  [default: gfs]
  -o, --obs [godae|gsi|cosmic]    Observation: source or source.instrument or
                                  source.instrument.platform or
                                  source.instrument.operator  [default: gsi]
  -r, --resolution <resolution>   List selected resolution. [Multiple=True]
                                  [Default=Full-Res]
  --instrument <insturment>       Filter only selected instrument(s) by name.
  --platform <platform>           Filter only selected platform(s) by name.
  --filetype [netcdf|odb]         IODA file type.  [default: netcdf, odb]
  --overwrite                     Overwrite existing local targets.  [default:
                                  False]
  --cleanup-partial               Cleanup partial (incomplete) transfers.
                                  [default: False]
  --validate-source               Validate source before copy  [default: False]
  --validate-target               Validate target after copy  [default: False]
  --validate-existing-target      Validate existing target  [default: True]
  --stop-if-error                 Stop transfer if errors occur  [default:
                                  False]
  --stop-if-existing              Stop transfer if existing found  [default:
                                  False]
  --quiet                         Be more quiet  [default: False]
  --help                          Show this message and exit.
```