CCIN2P3

cnrs

Centre de Calcul
de l'Institut National de Physique Nucléaire
et de Physique des Particules

Rubin-LSST data processing at CC-IN2P3

Quentin Le Boulc'h

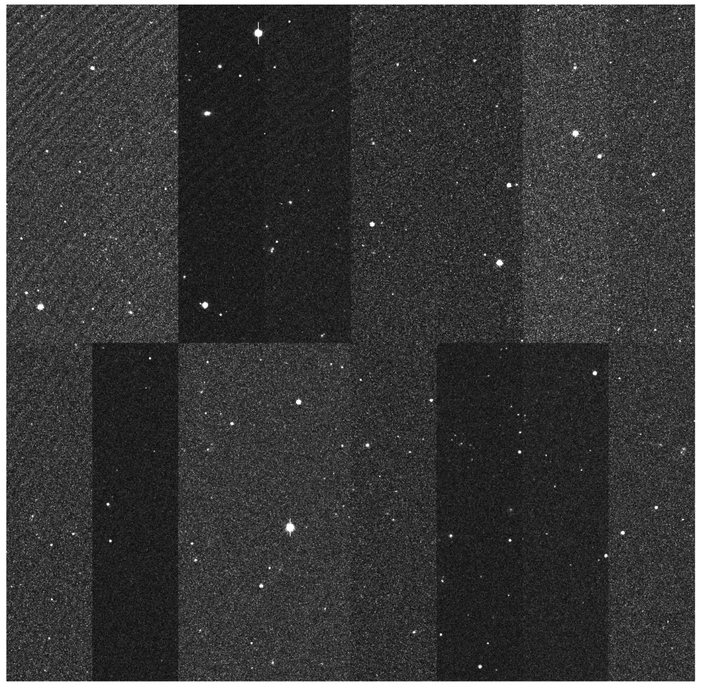
VERA C. RUBIN OBSERVATORY

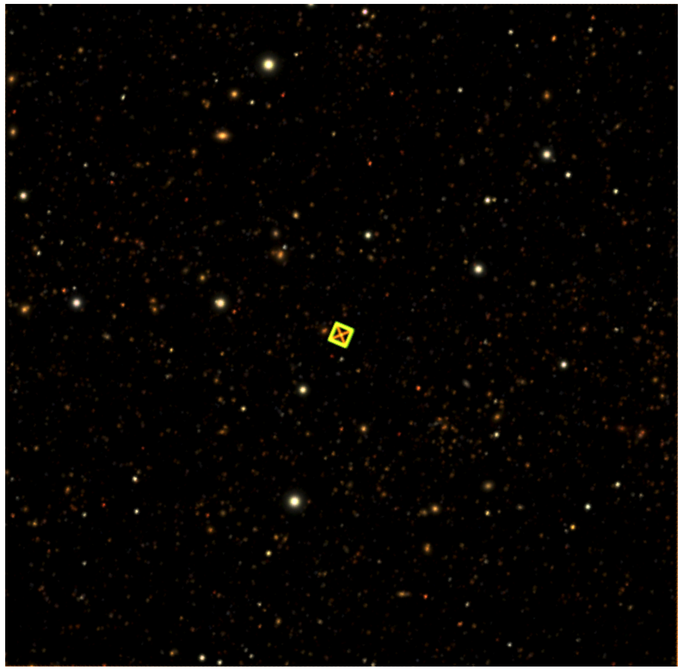
LSST

Data Release Production pipelines

DRP in a nutshell:

- **Process raw images** and calibrate them to remove any artifact from instrument
- Merge images together ("**coaddition**") to improve signal/noise
- **Measure differences** between images to detect changes
- Detect objects, measure their properties and populate **catalogs**





Data Release Production pipelines

DRP in a nutshell:

- **Process raw images** and calibrate them to remove any artifact from instrument
- Merge images together ("**coaddition**") to improve signal/noise
- **Measure differences** between images to detect changes
- Detect objects, measure their properties and populate **catalogs**


```
coord_ra,coord_dec,ccdVisitId,parentSourceId,x,y,xErr,yErr,ra,decl,calibFlux,calibFluxErr,ap03Flux,ap03FluxErr,ap03Flux_flag,ap06Flux,ap06FluxErr,ap06Flux_flag,ap09Flux,ap09FluxErr,ap09Flux_flag,ap12Flux,ap12FluxErr,ap12Flux_flag,ap17Flux,ap17FluxErr,ap17Flux_flag,ap25Flux,ap25FluxErr,
ap25Flux_flag,ap35Flux,ap35FluxErr,ap35Flux_flag,ap50Flux,ap50FluxErr,ap50Flux_flag,ap70Flux,ap70FluxErr,ap70Flux_flag,sky,skyErr,psfFlux,psfFluxErr,ixx,iyy,ixy,ixxPSF,iyyPSF,ixyPSF,gaussianFlux,gaussianFluxErr,extendedness,localPhotoCalib,localPhotoCalib_flag,localPhotoCalibErr,localW
cs_flag,localWcs_CDMatrix_2_1,localWcs_CDMatrix_1_1,localWcs_CDMatrix_1_2,localWcs_CDMatrix_2_2,blendedness_abs,blendedness_flag,blendedness_flag_noCentroid,blendedness_flag_noShape,apFlux_12_0_flag,apFlux_12_0_flag_apertureTruncated,apFlux_12_0_flag_sincCoeffsTruncated,apFlux_12_0_ins
tFlux,apFlux_12_0_instFluxErr,apFlux_17_0_flag,apFlux_17_0_instFlux,apFlux_17_0_instFluxErr,extendedness_flag,footprintArea_value,jacobian_flag,jacobian_value,localBackground_instFlux,localBackground_instFluxErr,localBackground_flag,localBackground_flag_noGoodPixels,localBackground_fla
g_noPsf,pixelFlags_bad,pixelFlags_cr,pixelFlags_crCenter,pixelFlags_edge,pixelFlags_interpolated,pixelFlags_interpolatedCenter,pixelFlags_offimage,pixelFlags_saturated,pixelFlags_saturatedCenter,pixelFlags_suspect,pixelFlags_suspectCenter,psfFlux_apCorr,psfFlux_apCorrErr,psfFlux_area,p
sfFlux_flag,psfFlux_flag_apCorr,psfFlux_flag_edge,psfFlux_flag_noGoodPixels,gaussianFlux_flag,centroid_flag,centroid_flag_almostNoSecondDerivative,centroid_flag_badError,centroid_flag_edge,centroid_flag_noSecondDerivative,centroid_flag_notAtMaximum,centroid_flag_resetToPeak,variance_fl
ag,variance_flag_emptyFootprint,variance_value,calib_astrometry_used,calib_detected,calib_photometry_reserved,calib_photometry_used,calib_psf_candidate,calib_psf_reserved,calib_psf_used,deblend_deblendedAsPsf,deblend_hasStrayFlux,deblend_masked,deblend_nChild,deblend_parentTooBig,deble
nd_patchedTemplate,deblend_rampedTemplate,deblend_skipped,deblend_tooManyPeaks,hsmPsfMoments_flag,hsmPsfMoments_flag_no_pixels,hsmPsfMoments_flag_not_contained,hsmPsfMoments_flag_parent_source,hsmShapeRegauss_flag,hsmShapeRegauss_flag_galsim,hsmShapeRegauss_flag_no_pixels,hsmShapeRegau
ss_flag_not_contained,hsmShapeRegauss_flag_parent_source,sky_source,detect_isPrimary,band,instrument,detector,physical_filter,visit_system,visit

54.289953664250866,-40.293267839666626,492028000,0,4044.0,19.0,,,54.289953664250866,-40.293267839666626,58514.17581773361,2927.8586378855653,27234.627307978462,760.8054085684626,True,50825.94174619848,1489.1644065091475,True,55210.77005531153,2205.2565913059484,True,58514.17581773361,2
927.8586378855653,True,54209.837612740215,4140.033553914013,True,,,True,,,True,,,True,,,True,-1.4600232384550038,136.57027351114698,56178.98154648083,1384.1157255531048,6.444413105039852,5.953385819426379,0.37244183717232615,6.792772073263132,6.604076569179897,0.16660893252961984,54641
.50071740754,1396.6248479067283,,2.034586657893343,False,0.0006318826663730968,False,3.9373687033725994e-07,-8.839573547989377e-07,3.948647274437616e-07,8.817207680087427e-07,0.0,True,True,True,True,False,True,28759.736328125,1439.0157470703125,True,26644.152709066868,2034.811004425342
,True,572,False,0.9982791522572211,-0.7176018936282345,67.12433357436944,True,False,False,False,False,False,True,False,False,False,False,False,False,False,False,0.9491527374465585,0.0,99.153404,True,False,True,False,False,True,False,False,False,True,False,False,False,True,True,,False,False,False,F
alse,False,False,False,False,False,False,0,False,False,False,False,False,True,False,False,False,True,True,False,False,False,False,False,True,y,LSSTCam-imSim,0,y_sim_1.4,1,492028

54.49983304318144,-40.3639773600175,492028000,0,888.4863930557776,24.209198837981198,0.05175733,0.051064942,54.49983304318144,-40.3639773600175,136507.95320678188,2973.916171931787,62369.20491397495,829.6969035947769,False,118911.6794915128,1557.6090883787397,False,131451.08343983756,2
261.0285768211343,False,136507.95320678188,2973.916171931787,False,139147.79069430396,4201.358672770262,False,,,True,,,True,,,True,,,True,2.5023856903109385,135.60757257602256,130928.03846184416,1507.8975564679422,7.2459476278218995,6.900482339774454,-0.2931864039555455,6.6977039565449
87,6.706558977593371,0.11463873596951761,134536.3620691466,1632.1172284821535,0.0,2.034586657893343,False,0.0006318826663730968,False,3.9136815637648663e-07,-8.84678115419472e-07,3.929615064435271e-07,8.824331076464853e-07,0.0,False,False,False,False,False,True,67093.703125,1461.532226
5625,False,68391.18410339952,2064.8599039284873,False,726,False,0.9981326068196241,1.2299233756413037,66.65116575285455,False,False,False,False,False,False,False,False,False,False,False,0.9870799510349546,0.0,93.84247,False,False,False,False,False,False,False,False,False,Fa
lse,False,False,False,True,True,,False,True,False,False,False,False,False,False,False,False,False,0,False,False,False,False,False,False,False,False,False,False,False,False,False,False,True,y,LSSTCam-imSim,0,y_sim_1.4,1,492028

54.379256129488596,-40.32228523879102,492028000,0,2709.0,40.0,,,54.379256129488596,-40.32228523879102,44790.555010869444,2918.376501887161,5499.503631493971,717.8918518678215,True,18718.58469832021,1455.4876483903606,True,31745.182740663953,2187.42615375103,True,44790.555010869444,2918
.376501887161,True,60327.18741636077,4139.915155478536,True,72983.0850524936,6094.671218002943,True,72170.01625376414,8532.389105549124,True,,,True,,,True,0.3396400412003608,136.48216863050106,16460.956411397176,1331.617384763296,42.43230174336507,67.44609844098791,10.562999314506914,6
.768722651926946,6.638581507680412,0.13803875894838405,65412.95736087909,3726.593018045784,,2.034586657893343,False,0.0006318826663730968,False,3.927323909035765e-07,-8.842682733053736e-07,3.9405655077013025e-07,8.820294386754979e-07,0.0,True,True,True,True,False,False,22014.572265625,
1434.36669921875,True,29650.832114875317,2034.7487993041134,True,786,False,0.9982222351243386,0.16693319003282553,67.08102999740352,True,False,False,False,False,False,False,False,False,False,False,False,False,False,False,0.9699220697895434,0.0,97.13245,True,False,False,False,False,True,False
,False,True,False,False,False,False,False,4556.2509765625,False,False,False,False,False,False,False,False,False,False,0,False,False,False,False,False,False,True,False,False,False,False,False,False,False,False,False,True,y,LSSTCam-imSim,0,y_sim_1.4,1,492028
```

Data Preview 0.2 (DP0.2)

- Data Preview 0:
  - Early integration test of pipelines and Rubin Science Platform
  - Provide data for preparation of science analysis

- DP0.2:
  - Reprocessing of simulated data using latest pipelines
  - Gen3 butler (data management)
  - Workflow automation

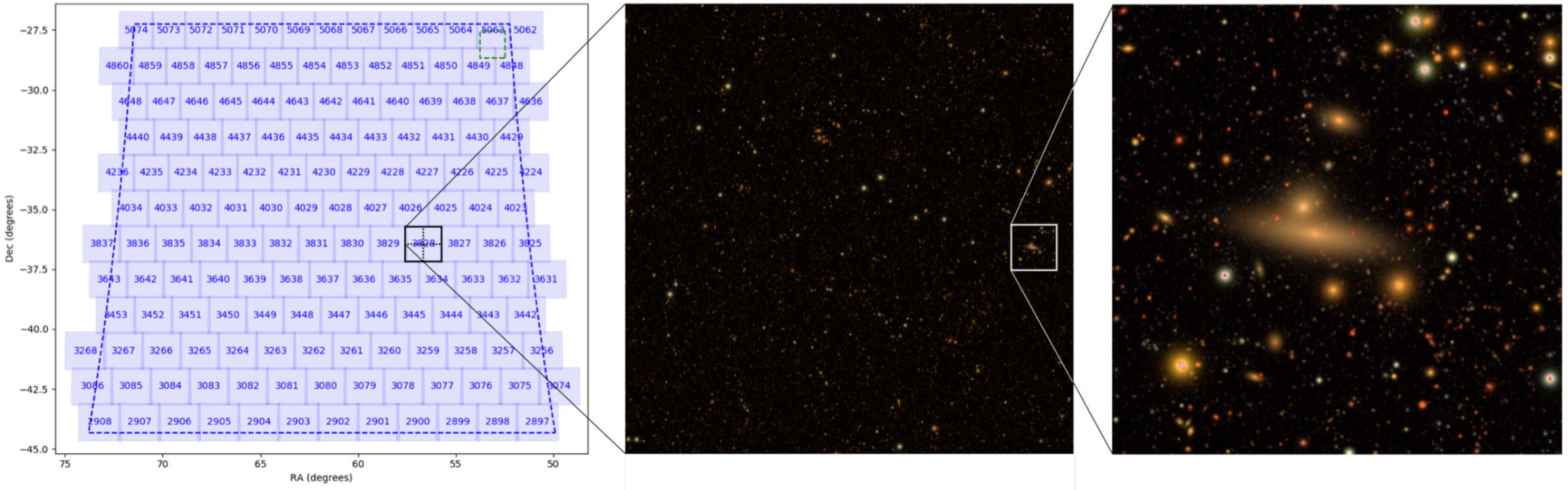
Data Product	DP0.1 DC2 Simulated Sky Survey	DP0.2 Reprocessed DC2 Survey	DP1 ComCam On-Sky Data	DP2 LSSTCam On-Sky Data	DR1 LSST First 6 Months Data
Raw images	☑	☑	☑	☑	☑
DRP Processed Visit Images and Visit Catalogs	☑	☑	☑	☑	☑
DRP Coadded Images	☑	☑	☑	☑	☑
DRP Object and ForcedSource Catalogs	☑	☑	☑	☑	☑
DRP Difference Images and DIASources	☐	☑	☑	☑	☑
DRP ForcedSource Catalogs including DIA outputs	☐	☑	☑	☑	☑
PP Processed Visit Images	☐	☐	☑	☑	☑
PP Difference Images	☐	☐	☑	☑	☑
PP Catalogs (DIASources, DIAObjects, DIAForcedSources)	☐	☐	☑	☑	☑
PP Alerts (Canned)	☐	☐	☑	☑	☑
PP Alerts (Live, Brokered)	☐	☐	☐	☑	☑
PP SSP Catalogs	☐	☐	☑	☑	☑
DRP SSP Catalogs	☐	☐	☐	☐	☑

Data Preview 0.2 (DP0.2)

Simulated Rubin images generated by DESC (Dark Energy Science Collaboration):

- 300 deg2 (full survey ~ 20 000 deg2)
- 5 years (full survey: 10 years)

  ➜ approximately **0.5% of the full survey**, or 10 nights of data gathering



Data Preview 0.2 (DP0.2)

Simulated Rubin images generated by DESC (Dark Energy Science Collaboration):

- 300 deg2 (full survey ~ 20 000 deg2)
- 5 years (full survey: 10 years)

  ➜ approximately **0.5% of the full survey**, or 10 nights of data gathering

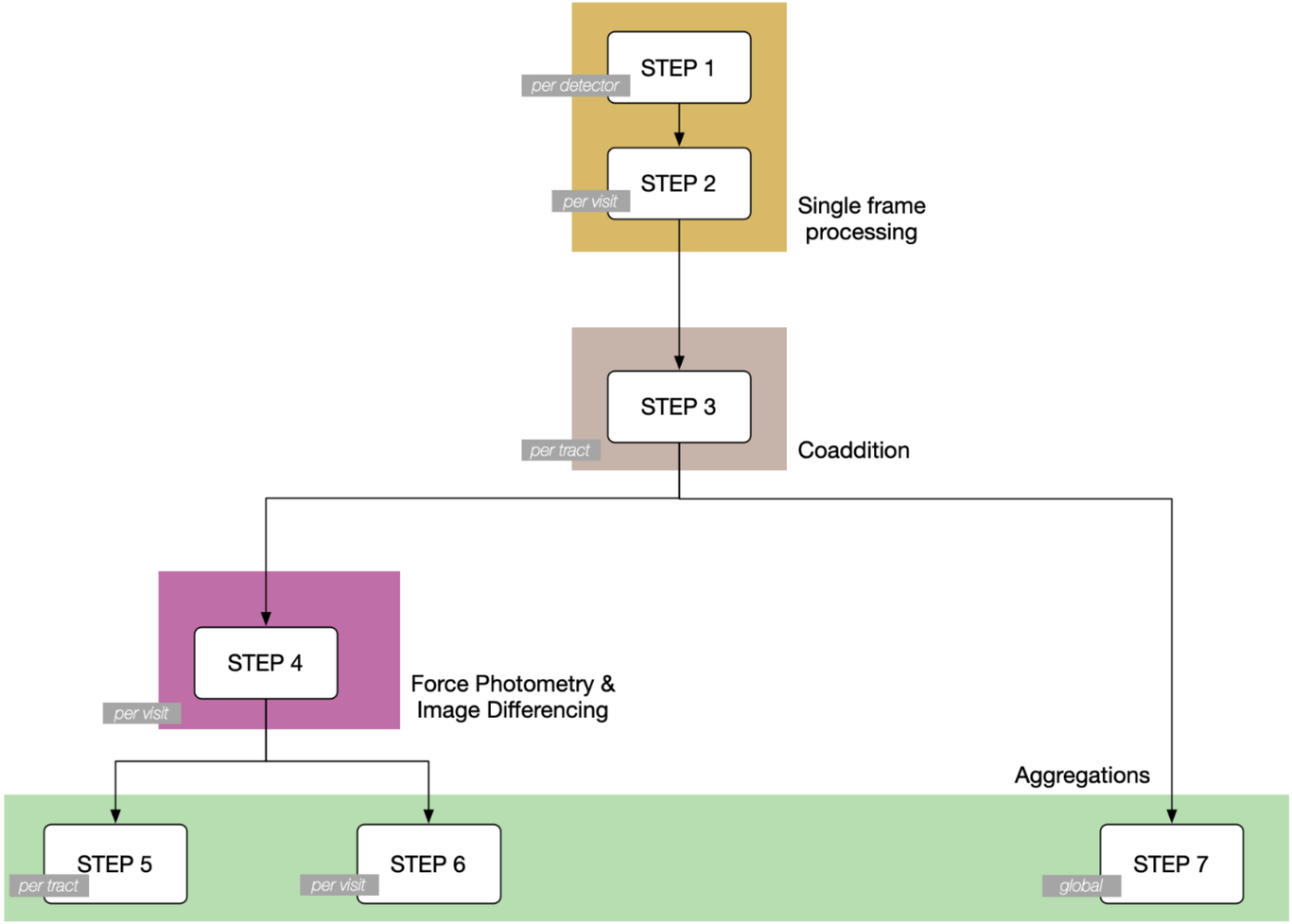
- 300 deg2 = about 30 pointings
- 5 year = hundreds of visits per pointing

➜ approximately **20,000 Rubin camera simulated exposures**, each with up to 189 detectors

➜ **3 M files**, 18 MB each, **50 TB** in total

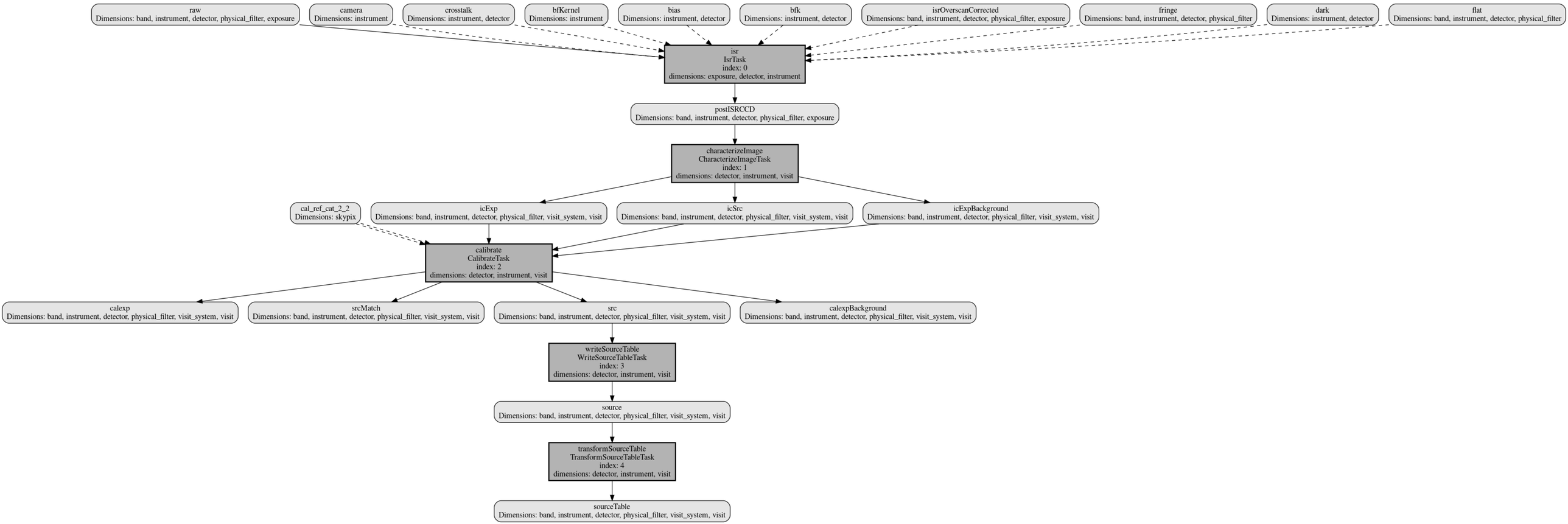
Workflows of pipeline tasks

- Set of ~ 80 pipeline tasks, grouped in 7 steps
- Each step process data at different levels: single CCD, full visit, coadded images, catalogs…



Workflows of pipeline tasks

- Pipeline tasks depend on each other through data production / consumption
- Some of the steps have rather straightforward workflows:



Workflows of pipeline tasks

- Pipeline tasks depend on each other through data production / consumption
- Others are more complex (step3):

Workflows of pipeline tasks

- Each task must be executed between 1 and 3 millions times
- Tasks execution time goes from few seconds to 24+ hours
- Maximum memory usage for each task varies from 1 GB to 200+ GB
- For a given task, memory usage depends on input data

➜ **Automation**!

Batch Processing Service (BPS): LSST framework for distributed pipeline execution

- Generate a workflow and submit it to an external **Workflow Management System (WMS)**
- A plugin is needed to interface BPS to the WMS
- Explore existing WMS that could be plugged into BPS to run the processing workflows on our computing platform

Workflows of pipeline tasks

- HTCondor: only available through grid interfaces at CC-IN2P3

- We needed a WMS that:
  - Can talk to **Grid Engine and Slurm**
  - Use **pilots jobs**
  - Is **scalable**
  - Does not require too much development to interface it with BPS

➜ **Parsl**: "extends parallelism in Python beyond a single computer"
  - Can submit jobs to Grid Engine and Slurm
  - Provides pilot jobs through its HighThroughputExecutor feature
  - Claimed to be scalable
  - Already existing plugin developed by J. Chiang (SLAC)

➜ All processing done using Parsl + plugin + configuration/customization

Execution of DP0.2 campaign

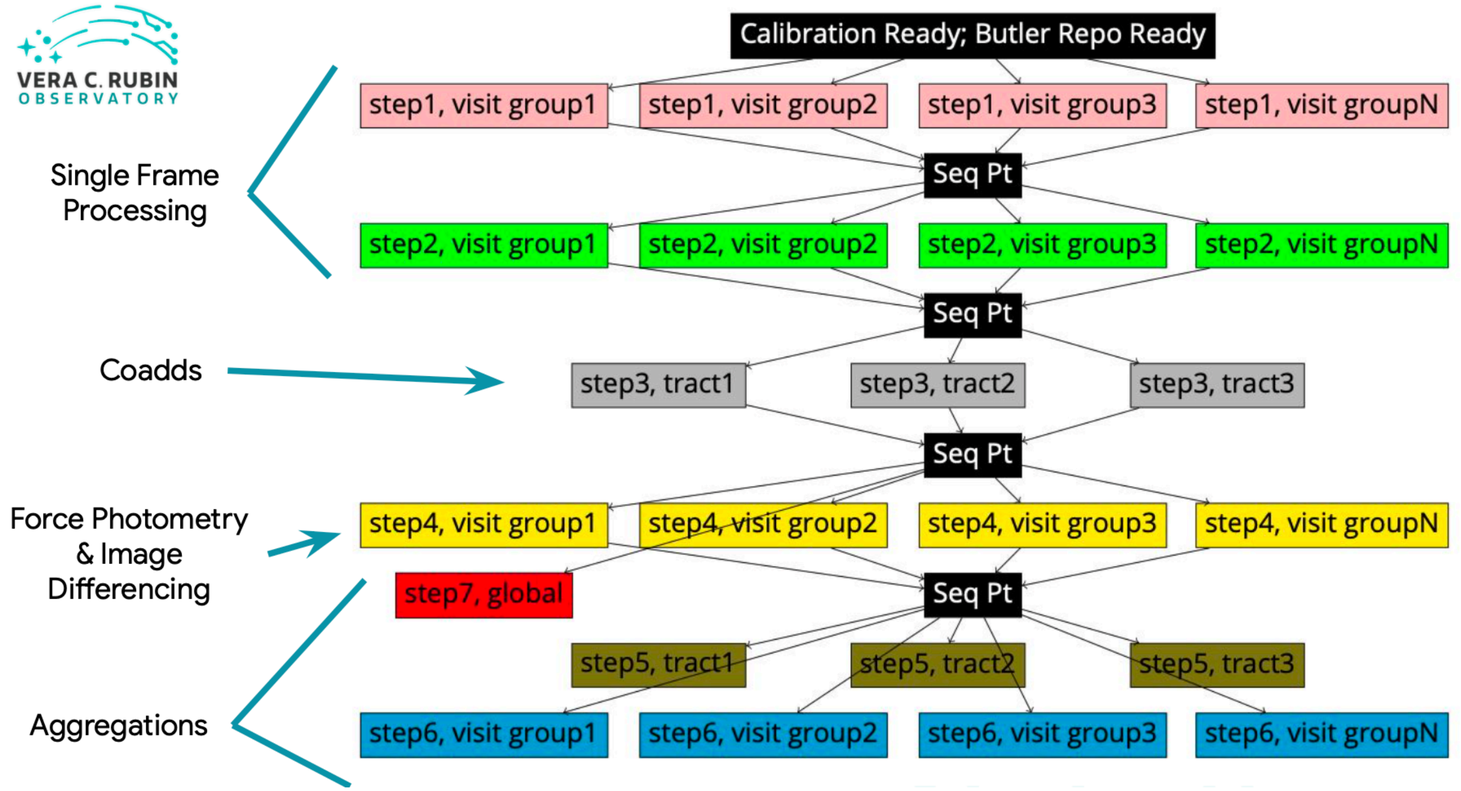
- Full production from step 1 to step 7
  - From April to October ’22
  - Using up to **3,000 simultaneous Slurm jobs**
  - Executed 57,903,740 pipetasks which consumed 2,347,306 elapsed hours

- Butler repository details:
  - Input dataset: 51 TiB, 2.9 M files
  - Products: **3 PiB**, 54 M directories, **201 M files** (including intermediates)
  - Registry database: 314 GiB

Execution of DP0.2 campaign

- Some of the steps can not be executed in a single workflow:
  - Workflow is too large and its generation used too much memory and time
  - Execution too long
  - Scalability issues

➜ Divided into sub-workflows of "reasonable" size (up to 32 workflows for step3)

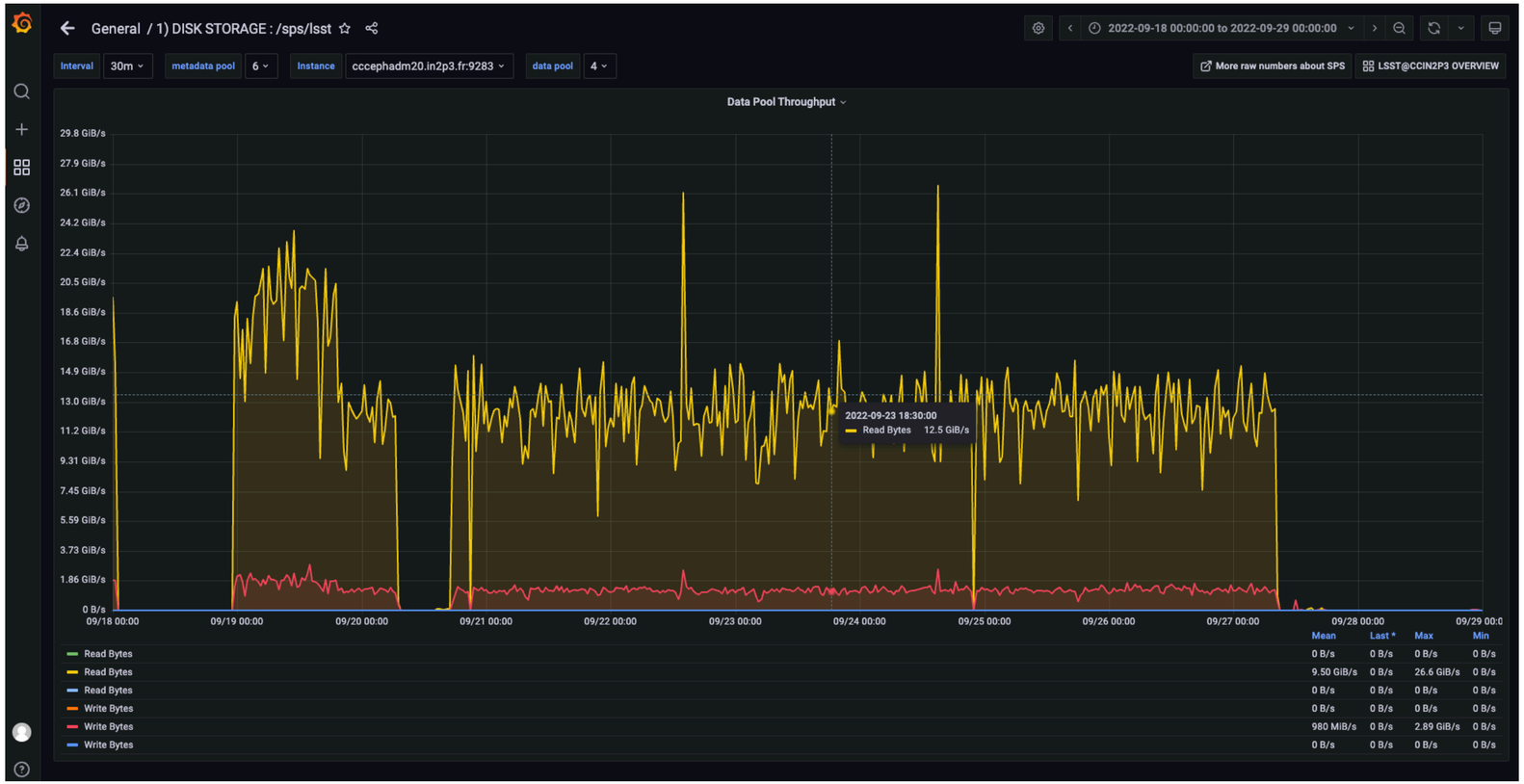


Execution of DP0.2 campaign

- Number of tasks in the workflows can be reduced thanks to the **clustering** feature of BPS:
  - Workflow graph split into subgraph where tasks are grouped together ("cluster")
  - A cluster is seen as a single task to be executed
  - Better scalability, but makes resources tuning and debugging more complex

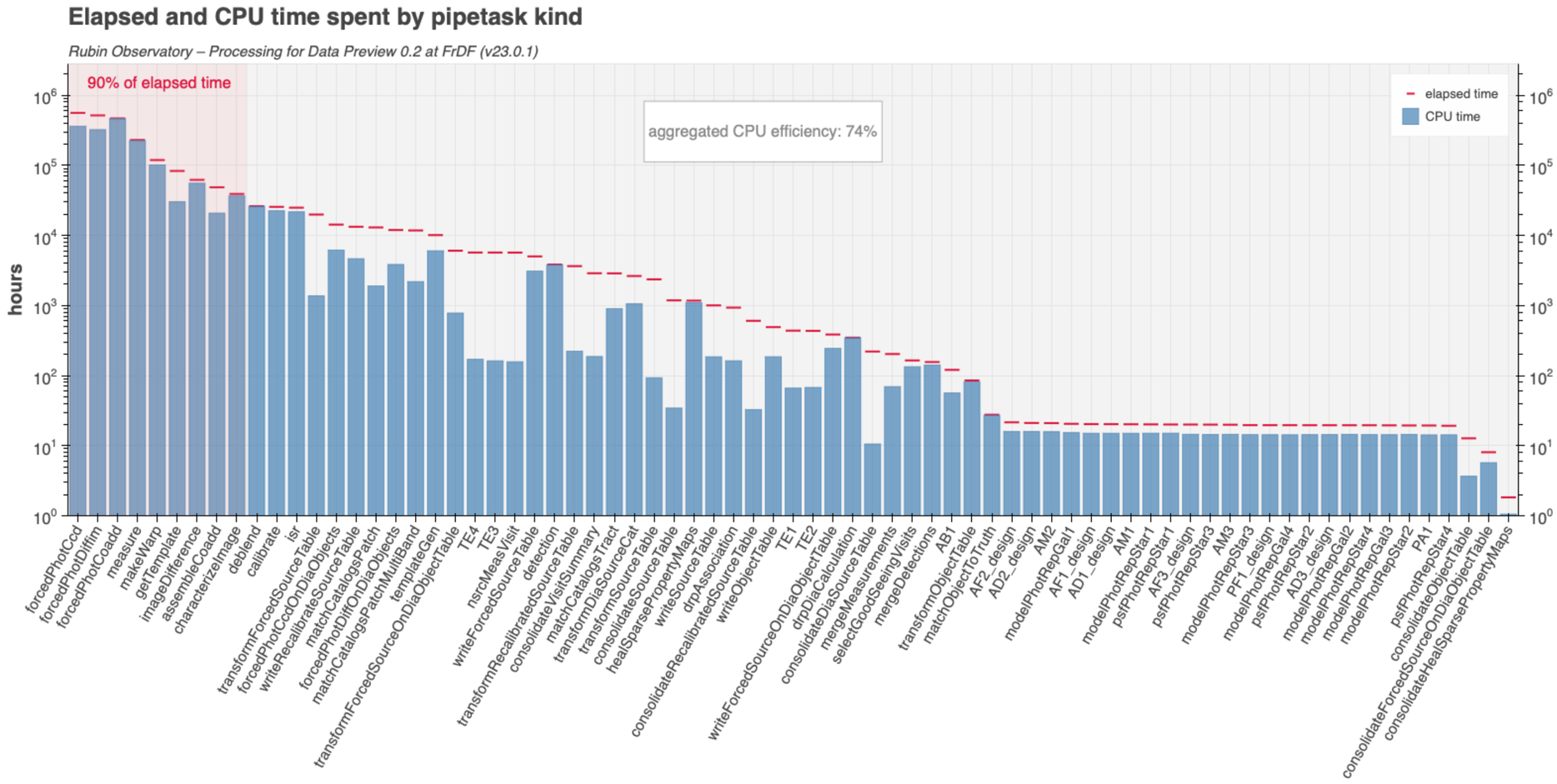
Execution of DP0.2 campaign

- Storage system: /sps/lsst (CephFS)
  - **Very good behaviour** overall
  - Performance issue for some tasks that required to copy the files to the local disk of the compute node: significant improvement (80% CPU efficiency, induced I/O throughput reduced to 12 GiB/s)



Execution of DP0.2 campaign

- 90% of total computing time used by 9 tasks: to be optimized in priority
- These tasks have very good CPU efficiency
- Overall CPU efficiency is good: 74%

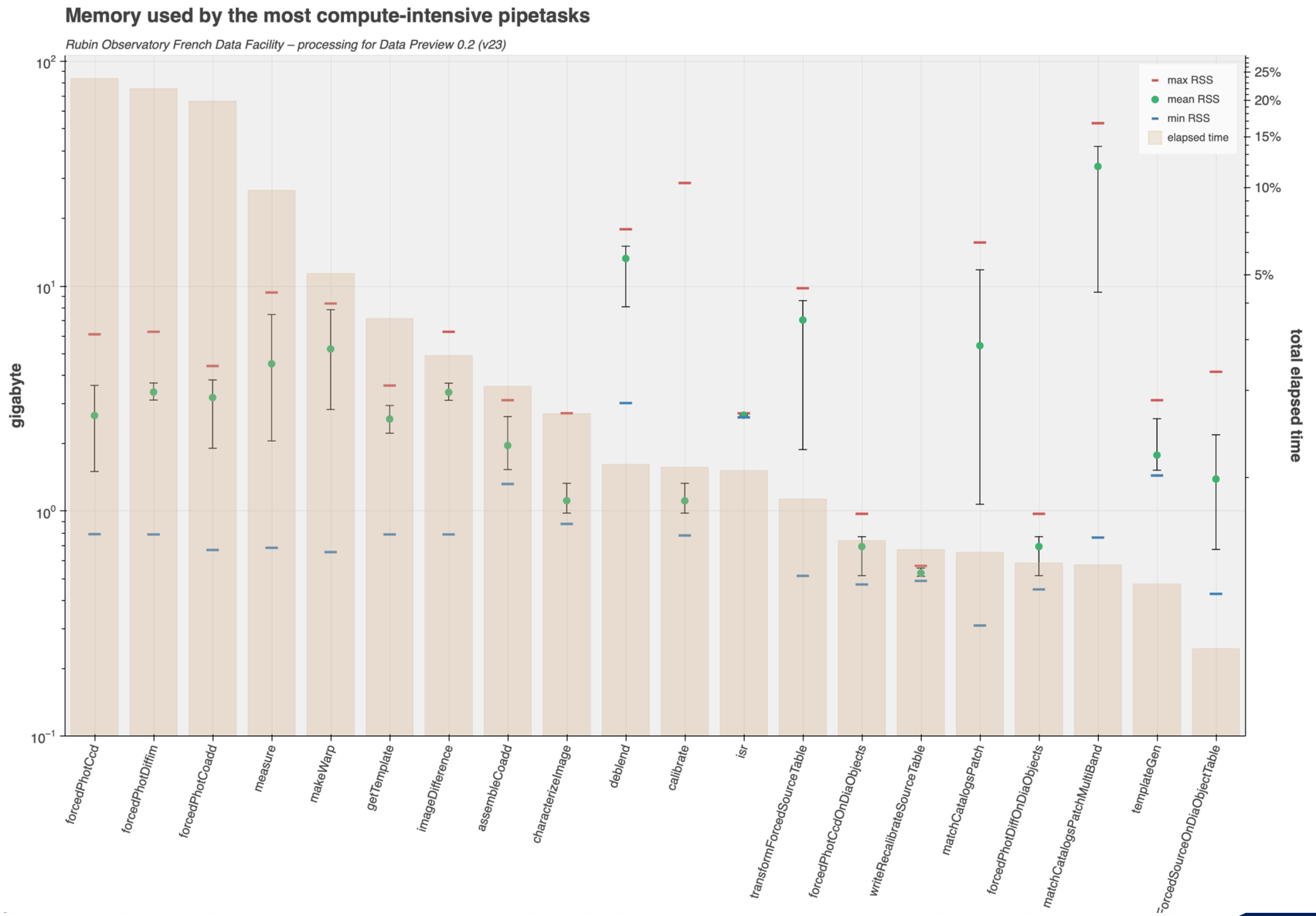


Execution of DP0.2 campaign

- Tasks with highest CPU usage have memory needs between 1 and 10 GB
- Tasks with moderate CPU usage have higher memory needs, up to 50 GB

Several options for improvement:

- Jobs tuning
- Pipeline optimisation
- Batch system configuration
- Hardware



Profiling

Current sources of informations:

- Metrics provided directly by the LSST processing framework
- Informations provided by the job scheduler (Slurm)
- Internal system metrics on nodes, storage system, network…
- Profiling?