CODE-DE Processing Integration Workshop

25. & 26. June 2018
DLR Oberpfaffenhofen
Outline

- Purpose of the CODE-DE processing portal
- Access the processing portal
- Level-2 Processing
- Level-3 Processing
- Pixel extraction (Matchup analysis)
- Stored Requests
- Production monitoring
Purpose of the Processing Web GUI

- allow to interactively define and submit processing tasks
- allow to monitor processing
- offer access point for picking up the results
Access to the Processing Web GUI

Either from CODE-DE frontpage:

Or via direct link:
https://processing.code-de.org
Access to the Processing Web GUI

Log in using your CODE-DE credentials
Level-2 Processing

- Landing page
- Simplest processor type
- Level-2: typically anything to process a single input file
- Example applications: cloud screening, atmospheric correction, e.g. with Sen2Cor, water constituents retrieval, RGB image generation, …
Level-3 Processing

- Aggregation of input data
- Input: typically L2, but also L1 data
- Typical application: generate time series of cloud-free monthly global maps showing average value of some variable, e.g. amount of burned area
- Allows to fine-tune methods of aggregation
Pixel Extraction (Matchup Analysis)

- Analyses matchups between reference data and L1 or L2 data
- Typical application: validation of processor on basis of in-situ data
- Output: not raster data, but a CSV table of pixels and a report
Regions

- Define/edit/remove processing regions
- Used in processing steps to filter input data set
Requests

- Requests can be stored as templates
- Edit or remove processing requests
- Re-use previous processing requests
Monitoring

- Allows to monitor status of productions
- Check if running, completed, or erroneous
- Cancel ongoing production
- User may edit request and submit again
- Offers button for downloading processing results
Outline – Processing Command Line Interface

• Login to “front end” machine
• Access to data
• Submission of processing requests to the cluster (cpt tool)
• Workflow control scripts
Processing Web GUI

- **browser** (https)
- **CODE Processing Web GUI**
  - Request forms
    - Input collections
    - Processor selection
    - Parameters
  - Job submission
    - Progress monitoring
  - Result set access
    - Download
- **Storage**
  - EO data
    - Sentinel 1
    - Sentinel 2
    - Sentinel 3
    - ...
    - processing results
- **CODE Processing Cluster**
  - Processor tasks
  - Processor images
Project VMs and the command line interface

- **Gateway**
  - ssh
  - Command line client
  - Services
  - Workflows

- **Project VMs**
  - Project1 VM
  - Project2 VM
  - Project3 VM

- **CODE Processing Cluster**
  - Processor tasks
    - Docker containers
    - Linux executables
    - Sentinel Toolbox op.
    - Python w. Anaconda

- **Storage**
  - EO data
    - Sentinel 1
    - Sentinel 2
    - Sentinel 3
    - ...
  - Processing results
Project VMs - login

- Personal user accounts, membership in project group
- One project machine per project
- ssh login via gateway cd-hop.eoc.dlr.de (129.247.255.24)

```
ssh cvuser1@cd-hop.eoc.dlr.de
ssh cvuser1@cd-project1
```
Access to data

- **Sentinel data**
  
  `/gpfs/DATA2/www/Sentinel1/<year>/<month>/<day>/<zip-file>
  
  `/gpfs/DATA4/www/Sentinel2/...
  
  `/gpfs/DATA3/www/Sentinel3/OLCI/...
  
  `/gpfs/DATA3/www/Sentinel3/SLSTR/...
  
  `/gpfs/DATA3/www/Sentinel3/SRAL/...
  
- **Example:**
  
  `ls /gpfs/DATA4/www/Sentinel2/2018/01/29`
  
  `S2A_MSIL1C_20180129T002701_N0206_R016_T54HUC_20180129T051552.SAFE.zip`
  
  `S2A_MSIL1C_20180129T002701_N0206_R016_T54HVC_20180129T051552.SAFE.zip`
  
  `...`

- **Auxiliary data**
  
  `ls /calvalus/auxiliary/*`
  
  `/calvalus/auxiliary/dem:
  
  SRTM 3Sec`
  
  `/calvalus/auxiliary/meris_l2:
  
  atmosphere case1 cloud landaero lv2conf meris_l2_config.xml`
  
  `/calvalus/auxiliary/watermask:
  
  150m.zip  50m.zip  GC_water_mask.zip  MODIS_north_water_mask.zip`
  
  `MODIS_south_water_mask.zip`
Access to data

- Software package installation location for use on the cluster
  /calvalus/home/<user>/software/<package>-<version>/

- Example:
  
  ls -l /calvalus/home/cvuser1/software/workshop-ndvi-demo-1.0
  s2tbx-radiometric-indices-6.0.0.jar

- Software packages pre-installed
  
  ls /calvalus/software/1.0
  sen2cor-2.3.1
  sen2cor-2.4.0
  fmask-python-0.4.0
  idepix-6.0
  snap-buildin-6.0
  urbantep-timescan-1.0
  fub-wew-6.0
  s2-granules-1.0
  calvalus-2.14
  calvalus-2.14.1
  calvalus-2.14.2
  snap-5-cv-2.14.1
  snap-5-cv-2.15-SNAPSHOT
Access to data

- Working directory
  - for input, intermediates, output of a single processing task
  - on SSD of compute node
  - managed by Hadoop, available for the runtime of the task

- Distributed cache
  - processor software automatically deployed and unpacked before use
  - on SSD of compute node
  - managed by Hadoop on a LRU scheme

Sentinel data archive [PB]
Shared file system [TB] for software and data
Home directory [GB]
Tools and examples
Cluster task working dirs [GB]
## Submission of processing requests to the cluster

<table>
<thead>
<tr>
<th>productionType: L2</th>
</tr>
</thead>
<tbody>
<tr>
<td>productionName: Sen2cor test</td>
</tr>
<tr>
<td>processorBundleName: sen2cor</td>
</tr>
<tr>
<td>processorBundleVersion: 2.4.0</td>
</tr>
<tr>
<td>processorName: sen2cor</td>
</tr>
<tr>
<td>inputPath:</td>
</tr>
<tr>
<td>- /gpfs/DATA4/www/Sentinel2/2017/04/10/\n  S2A_MSIL1C_20170410T103021_N0204_R108_T32UNE_20170410T103020.SAFE.zip</td>
</tr>
<tr>
<td>- /gpfs/DATA4/www/Sentinel2/2017/06/29/\n  S2A_MSIL1C_20170629T103021_N0205_R108_T32UNE_20170629T103020.SAFE.zip</td>
</tr>
<tr>
<td>calvalus.output.dir: /calvalus/home/cvuser1/s2-sen2cor-test</td>
</tr>
</tbody>
</table>

s2-sen2cor-test.yaml

- Requests in XML or YAML format
- Production types L2, L3, pixel extraction (MA)
- Processor selection
- Input datasets
- Processing parameters
Submission of processing requests to the cluster

ssh cvuser1@cd-project
cd example-inst
  . myexample

submitproductionrequest special-requests/s2-sen2cor-test.yaml

# or in the background:
# nohup submitproductionrequest special-requests/s2-sen2cor-test.yaml > sen2cor.out &

ls /calvalus/home/cvuser1/s2-sen2cor-test/
yarn application –list
yarn logs –applicationId <applicationId>
Submission of processing requests to the cluster

productionType: L2Plus
productionName: S2 Idepix test
processorBundleName: urbantep-timescan
processorBundleVersion: 1.0
processorName: Idepix.Sentinel2

inputPath:
- /gpfs/DATA4/www/Sentinel2/${yyyy}/${MM}/${dd}/S2.*T32UNE.*.zip

minDate: 2017-04-01
maxDate: 2017-06-30

calvalus.input.regionName: Hamburg

calvalus.input.regionWKT: POLYGON((9.836883544921875 53.60587002495288,
10.443878173828125 53.60587002495288,10.443878173828125
9.836883544921875,9.836883544921875 53.39020833928862,
9.836883544921875 53.60587002495288))
calvalus.input.format: CALVALUS-SENTINEL-2-MSI-20M

processorParameters: |
<parameters>
  <computeCloudBuffer>true</computeCloudBuffer>
  <cloudBufferWidth>5</cloudBufferWidth>
</parameters>

calvalus.output.dir: /calvalus/home/cvuser1/s2-idepix-test
outputFormat: NetCDF4
Data-driven production

- event for new product
- repeated query for new products
- processing request executed on cluster
- result made available online
Processor bundles

User

installation package

request

Hadoop Cluster

output data

master

feeder

external data source or destination

test server

tests

tests

tests

node 1

local disk

node 2

local disk

node n

local disk

node 3

local disk

node 4

local disk
Spatial and temporal aggregation

L3 parameters

- target resolution
- nearest neighbour resampling, or binning
- temporal compositing period
- valid pixel filter expression
- variables (virtual bands)
- aggregators
Spatial and temporal aggregation

Examples of aggregators

- AVG
- MIN_MAX
- PERCENTILE (MEDIAN)
- ON_MAX_SET (with maximised variable and output bands)
- SNAP Toolbox API to implement more

Examples of valid pixel filters

- for Fmask outputs:
  - fmask == 1 or
  - fmask == 4 or
  - fmask == 5

- for Idepix outputs:
  - not pixel_classif_flags.IDEPIC_CLOUD and
  - not pixel_classif_flags.IDEPIC_INVALID and
  - not pixel_classif_flags.IDEPIC_CLOUD_AMBIGUOUS and
  - not pixel_classif_flags.IDEPIC_CLOUD_BUFFER

- Sen2Cor outputs:
  - quality_scene_classification &
  - (8+128+256+512+1024) == 0
Mean composite (with simple cloud screening)
Max-NDVI composite