xcube

Release 0.2.0

Brockmann Consult GmbH

Sep 25, 2019

GETTING STARTED

Warning: This documentation is a work in progress and currently less than a draft.

xcube has been developed to generate, manipulate, analyse, and publish data cubes from EO data.

CHAPTER ONE

OVERVIEW

xcube is an open-source Python package and toolkit that has been developed to provide Earth observation (EO) data in an analysis-ready form to users. xcube achieves this by carefully converting EO data sources into self-contained *data cubes* that can be published in the cloud.

1.1 Data Cube

The interpretation of the term *data cube* in the EO domain usually depends on the current context. It may refer to a data service such as Sentinel Hub, to some abstract API, or to a concrete set of spatial images that form a time-series.

This section briefly explains the specific concept of a data cube used in the xcube project - the *xcube dataset*.

1.2 xcube Dataset

1.2.1 Data Model

An xcube dataset contains one or more (geo-physical) data variables whose values are stored in cells of a common multi-dimensional, spatio-temporal grid. The dimensions are usually time, latitude, and longitude, however other dimensions may be present.

All xcube datasets are structured in the same way following a common data model. They are also self-describing by providing metadata for the cube and all cube's variables following the CF conventions. For details regarding the common data model, please refer to the *xcube Dataset Specification*.

A xcube dataset's in-memory representation in Python programs is an xarray.Dataset instance. Each dataset variable is represented by multi-dimensional xarray.DataArray that is arranged in non-overlapping, contiguous sub-regions called *data chunks*.

1.2.2 Data Chunks

Chunked variables allow for out-of-core computations of xcube dataset that don't fit in a single computer's RAM as data chunks can be processed independently from each other.

The way how dataset variables are sub-divided into smaller chunks - their *chunking* - has a substantial impact on processing performance and there is no single ideal chunking for all use cases. For time series analyses it is preferable to have chunks with a smaller spatial dimensions and larger time dimension, for spatial analyses and visualisation on using a map, the opposite is the case.

xcube provide tools for re-chunking of xcube datasets (*xcube chunk*, *xcube level*) and the xcube server (*xcube serve*) allows serving the same data cubes using different chunkings. For further reading have a look into the Chunking and Performance section of the xarray documentation.

1.2.3 Processing Model

When xcube datasets are opened, only the cube's structure and its metadata are loaded into memory. The actual data arrays of variables are loaded on-demand only, and only for chunks intersecting the desired sub-region.

Operations that generate new data variables from existing ones will be chunked in the same way. Therefore, such operation chains generate a processing graph providing a deferred, concurrent execution model.

1.2.4 Data Format

For the external, physical representation of xcube datasets we usually use the Zarr format. Zarr takes full advantage of data chunks and supports parallel processing of chunks that may originate from the local file system or from remote cloud storage such as S3 and GCS.

1.2.5 Python Packages

The xcube package builds heavily on Python's big data ecosystem for handling huge N-dimensional data arrays and exploiting cloud-based storage and processing resources. In particular, xcube's in-memory data model is provided by xarray, the memory management and processing model is provided through dask, and the external format is provided by zarr. xarray, dask, and zarr have increased their popularity for big data solutions over the last couple of years, for creating scalable and efficient EO data solutions.

1.3 Toolkit

On top of xarray, dask, zarr, and other popular Python data science packages, xcube provides various higher-level tools to generate, manipulate, and publish xcube datasets:

- *CLI* - access, generate, modify, and analyse xcube datasets using the `xcube` tool;
- *Python API* - access, generate, modify, and analyse xcube datasets via Python programs and notebooks;
- *Web API and Server* - access, analyse, visualize xcube datasets via an xcube server;
- *Viewer App* – publish and visualise xcube datasets using maps and time-series charts.

1.4 Workflows

The basic use case is to generate an xcube dataset and deploy it so that your users can access it:

1. generate an xcube dataset from some EO data sources using the *xcube gen* tool with a specific *input processor*.
2. optimize the generated xcube dataset with respect to specific use cases using the *xcube chunk* tool.
3. optimize the generated xcube dataset by consolidating metadata and elimination of empty chunks using *xcube optimize* and *xcube prune* tools.
4. deploy the optimized xcube dataset(s) to some location (e.g. on AWS S3) where users can access them.

Then you can:

5. access, analyse, modify, transform, visualise the data using the *Python API* and xarray API through Python programs or JupyterLab, or
6. extract data points by coordinates from a cube using the *xcube extract* tool, or
7. resample the cube in time to generate temporal aggregations using the *xcube resample* tool.

Another way to provide the data to users is via the *xcube server*, that provides a RESTful API and a WMTS. The latter is used to visualise spatial subsets of xcube datasets efficiently at any zoom level. To provide optimal visualisation and data extraction performance through the xcube server, xcube datasets may be prepared beforehand. Steps 8 to 10 are optional.

8. verify a dataset to be published conforms with the *xcube Dataset Specification* using the *xcube verify* tool.
9. adjust your dataset chunking to be optimal for generating spatial image tiles and generate a multi-resolution image pyramid using the *xcube chunk* and *xcube level* tools.
10. create a dataset variant optimal for time series-extraction again using the *xcube chunk* tool.
11. configure xcube datasets and publish them through the xcube server using the *xcube serve* tool.

You may then use a WMTS-compatible client to visualise the datasets or develop your own xcube server client that will make use of the xcube's REST API.

The easiest way to visualise your data is using the xcube *Viewer App*, a single-page web application that can be configured to work with xcube server URLs.

CHAPTER

TWO

EXAMPLES

When you follow the examples section you can build your first tiny xcube dataset and view it in the xcube-viewer by using the xcube server. The examples section is still growing and improving :)

Have fun exploring xcube!

Warning: This chapter is a work in progress and currently less than a draft.

2.1 Generating an xcube dataset

In the following example a tiny demo xcube dataset is generated.

2.1.1 Analysed Sea Surface Temperature over the Global Ocean

Input data for this example is located in the xcube repository. The input files contain analysed sea surface temperature and sea surface temperature anomaly over the global ocean and are provided by Copernicus Marine Environment Monitoring Service. The data is described in a dedicated Product User Manual.

Before starting the example, you need to activate the xcube environment:

```
$ conda activate xcube
```

If you want to take a look at the input data you can use cli/xcube dump to print out the metadata of a selected input file:

```
$ xcube dump examples/gen/data/20170605120000-UKMO-L4_GHRSST-SSTfnd-OSTIAanom-GLOB-
↪v02.0-fv02.0.nc
```

```
<xarray.Dataset>
Dimensions:       (lat: 720, lon: 1440, time: 1)
Coordinates:
  * lat           (lat) float32 -89.875 -89.625 -89.375 ... 89.375 89.625 89.875
  * lon           (lon) float32 0.125 0.375 0.625 ... 359.375 359.625 359.875
  * time          (time) object 2017-06-05 12:00:00
Data variables:
    sst_anomaly   (time, lat, lon) float32 ...
    analysed_sst  (time, lat, lon) float32 ...
Attributes:
    Conventions:              CF-1.4
    title:                    Global SST & Sea Ice Anomaly, L4 OSTIA, 0.25 ...
```

(continues on next page)

(continued from previous page)

```
    summary:                     A merged, multi-sensor L4 Foundation SST anom...
    references:                  Donlon, C.J., Martin, M., Stark, J.D., Robert...
    institution:                 UKMO
    history:                     Created from sst:temperature regridded with a...
    comment:                     WARNING Some applications are unable to prope...
    license:                     These data are available free of charge under...
    id:                          UKMO-L4LRfnd_GLOB-OSTIAanom
    naming_authority:            org.ghrsst
    product_version:             2.0
    uuid:                        5c1665b7-06e8-499d-a281-857dcbfd07e2
    gds_version_id:              2.0
    netcdf_version_id:           3.6
    date_created:                20170606T061737Z
    start_time:                  20170605T000000Z
    time_coverage_start:         20170605T000000Z
    stop_time:                   20170606T000000Z
    time_coverage_end:           20170606T000000Z
    file_quality_level:          3
    source:                      UKMO-L4HRfnd-GLOB-OSTIA
    platform:                    Aqua, Envisat, NOAA-18, NOAA-19, MetOpA, MSG1...
    sensor:                      AATSR, AMSR, AVHRR, AVHRR_GAC, SEVIRI, TMI
    metadata_conventions:        Unidata Observation Dataset v1.0
    metadata_link:               http://data.nodc.noaa.gov/NESDIS_DataCenters/...
    keywords:                    Oceans > Ocean Temperature > Sea Surface Temp...
    keywords_vocabulary:         NASA Global Change Master Directory (GCMD) Sc...
    standard_name_vocabulary:    NetCDF Climate and Forecast (CF) Metadata Con...
    westernmost_longitude:       0.0
    easternmost_longitude:       360.0
    southernmost_latitude:       -90.0
    northernmost_latitude:       90.0
    spatial_resolution:          0.25 degree
    geospatial_lat_units:        degrees_north
    geospatial_lat_resolution:   0.25 degree
    geospatial_lon_units:        degrees_east
    geospatial_lon_resolution:   0.25 degree
    acknowledgment:              Please acknowledge the use of these data with...
    creator_name:                Met Office as part of CMEMS
    creator_email:               servicedesk.cmems@mercator-ocean.eu
    creator_url:                 http://marine.copernicus.eu/
    project:                     Group for High Resolution Sea Surface Tempera...
    publisher_name:              GHRSST Project Office
    publisher_url:               http://www.ghrsst.org
    publisher_email:             ghrsst-po@nceo.ac.uk
    processing_level:            L4
    cdm_data_type:               grid
```

Below an example xcube dataset will be created, which will contain the variable analysed_sst. The metadata for a specific variable can be viewed by:

```
$ xcube dump examples/gen/data/20170605120000-UKMO-L4_GHRSST-SSTfnd-OSTIAanom-GLOB-
↪v02.0-fv02.0.nc --var analysed_sst
```

```
<xarray.DataArray 'analysed_sst' (time: 1, lat: 720, lon: 1440)>
[1036800 values with dtype=float32]
Coordinates:
  * lat       (lat) float32 -89.875 -89.625 -89.375 ... 89.375 89.625 89.875
```

(continues on next page)

(continued from previous page)

```
  * lon      (lon) float32 0.125 0.375 0.625 0.875 ... 359.375 359.625 359.875
  * time     (time) object 2017-06-05 12:00:00
Attributes:
    long_name:      analysed sea surface temperature
    standard_name:  sea_surface_foundation_temperature
    type:           foundation
    units:          kelvin
    valid_min:      -300
    valid_max:      4500
    source:         UKMO-L4HRfnd-GLOB-OSTIA
    comment:
```

For creating a toy xcube dataset you can execute the command-line below. Please adjust the paths to your needs:

```
$ xcube gen -o "your/output/path/demo_SST_xcube.zarr" -c examples/gen/config_files/
↪xcube_sst_demo_config.yml --sort examples/gen/data/*.nc
```

The configuration file specifies the input processor, which in this case is the default one. The output size is 10240, 5632. The bounding box of the data cube is given by `output_region` in the configuration file. The output format (`output_writer_name`) is defined as well. The chunking of the dimensions can be set by the `chunksizes` attribute of the `output_writer_params` parameter, and in the example configuration file the chunking is set for latitude and longitude. If the chunking is not set, a automatic chunking is applied. The spatial resampling method (`output_resampling`) is set to 'nearest' and the configuration file contains only one variable which will be included into the xcube dataset - 'analysed-sst'.

The Analysed Sea Surface Temperature data set contains the variable already as needed. This means no pixel masking needs to be applied. However, this might differ depending on the input data. You can take a look at a configuration file which takes Sentinel-3 Ocean and Land Colour Instrument (OLCI) as input files, which is a bit more complex. The advantage of using pixel expressions is, that the generated cube contains only valid pixels and the user of the data cube does not have to worry about something like land-masking or invalid values. Furthermore, the generated data cube is spatially regular. This means the data are aligned on a common spatial grid and cover the same region. The time stamps are kept from the input data set.

Caution: If you have input data that has file names not only varying with the time stamp but with e.g. A and B as well, you need to pass the input files in the desired order via a text file. Each line of the text file should contain the path to one input file. If you pass the input files in a desired order, then do not use the parameter `--sort` within the commandline interface.

2.1.2 Optimizing and pruning a xcube dataset

If you want to optimize your generated xcube dataset e.g. for publishing it in a xcube viewer via xcube serve you can use cli/xcube optimize:

```
$ xcube optimize demo_SST_xcube.zarr -C
```

By executing the command above, an optimized xcube dataset called demo_SST_xcube-optimized.zarr will be created. You can take a look into the directory of the original xcube dataset and the optimized one, and you will notice that a file called .zmetadata. .zmetadata contains the information stored in .zattrs and .zarray of each variable of the xcube dataset and makes requests of metadata faster. The option `-C` optimizes coordinate variables by converting any chunked arrays into single, non-chunked, contiguous arrays.

For deleting empty chunks cli/xcube prune can be used. It deletes all data files associated with empty (NaN-only) chunks of an xcube dataset, and is restricted to the ZARR format.

```
$ xcube prune demo_SST_xcube-optimized.zarr
```

The pruned xcube dataset is saved in place and does not need an output path. The size of the xcube dataset was 6,8 MB before pruning it and 6,5 MB afterwards. According to the output printed to the terminal, 30 block files were deleted.

The metadata of the xcube dataset can be viewed with cli/xcube dump as well:

```
$ xcube dump demo_SST_xcube-optimized.zarr
```

```
<xarray.Dataset>
Dimensions:        (bnds: 2, lat: 5632, lon: 10240, time: 3)
Coordinates:
  * lat            (lat) float64 62.67 62.66 62.66 62.66 ... 48.01 48.0 48.0
    lat_bnds       (lat, bnds) float64 dask.array<shape=(5632, 2), chunksize=(5632, 2)>
  * lon            (lon) float64 -16.0 -16.0 -15.99 -15.99 ... 10.66 10.66 10.67
    lon_bnds       (lon, bnds) float64 dask.array<shape=(10240, 2), chunksize=(10240,␣
↪2)>
  * time           (time) datetime64[ns] 2017-06-05T12:00:00 ... 2017-06-07T12:00:00
    time_bnds      (time, bnds) datetime64[ns] dask.array<shape=(3, 2), chunksize=(3,␣
↪2)>
Dimensions without coordinates: bnds
Data variables:
    analysed_sst  (time, lat, lon) float64 dask.array<shape=(3, 5632, 10240),␣
↪chunksize=(1, 704, 640)>
Attributes:
    acknowledgment:             Data Cube produced based on data provided by ...
    comment:
    contributor_name:
    contributor_role:
    creator_email:              info@brockmann-consult.de
    creator_name:               Brockmann Consult GmbH
    creator_url:                https://www.brockmann-consult.de
    date_modified:              2019-09-25T08:50:32.169031
    geospatial_lat_max:         62.666666666666664
    geospatial_lat_min:         48.0
    geospatial_lat_resolution:  0.002604166666666666
    geospatial_lat_units:       degrees_north
    geospatial_lon_max:         10.666666666666664
    geospatial_lon_min:         -16.0
    geospatial_lon_resolution:  0.0026041666666666665
    geospatial_lon_units:       degrees_east
    history:                    xcube/reproj-snap-nc
    id:                         demo-bc-sst-sns-l2c-v1
    institution:                Brockmann Consult GmbH
    keywords:
    license:                    terms and conditions of the DCS4COP data dist...
    naming_authority:           bc
    processing_level:           L2C
    project:                    xcube
    publisher_email:            info@brockmann-consult.de
    publisher_name:             Brockmann Consult GmbH
    publisher_url:              https://www.brockmann-consult.de
    references:                 https://dcs4cop.eu/
    source:                     CMEMS Global SST & Sea Ice Anomaly Data Cube
    standard_name_vocabulary:
    summary:
    time_coverage_end:          2017-06-08T00:00:00.000000000
```

(continues on next page)

(continued from previous page)

```
    time_coverage_start:        2017-06-05T00:00:00.000000000
    title:                      CMEMS Global SST Anomaly Data Cube
```

The metadata for the variable analysed_sst can be viewed:

```
$ xcube dump demo_SST_xcube-optimized.zarr --var analysed_sst
```

```
<xarray.DataArray 'analysed_sst' (time: 3, lat: 5632, lon: 10240)>
dask.array<shape=(3, 5632, 10240), dtype=float64, chunksize=(1, 704, 640)>
Coordinates:
  * lat      (lat) float64 62.67 62.66 62.66 62.66 ... 48.01 48.01 48.0 48.0
  * lon      (lon) float64 -16.0 -16.0 -15.99 -15.99 ... 10.66 10.66 10.66 10.67
  * time     (time) datetime64[ns] 2017-06-05T12:00:00 ... 2017-06-07T12:00:00
Attributes:
    comment:
    long_name:            analysed sea surface temperature
    source:               UKMO-L4HRfnd-GLOB-OSTIA
    spatial_resampling:   Nearest
    standard_name:        sea_surface_foundation_temperature
    type:                 foundation
    units:                kelvin
    valid_max:            4500
    valid_min:            -300
```

Warning: This chapter is a work in progress and currently less than a draft.

2.2 Publishing xcube datasets

This example demonstrates how to run an xcube server to publish existing xcube datasets.

2.2.1 Running the server

To run the server on default port 8080 using the demo configuration:

```
$ xcube serve --verbose -c examples/serve/demo/config.yml
```

To run the server using a particular xcube dataset path and styling information for a variable:

```
$ xcube serve --styles conc_chl=(0,20,"viridis") examples/serve/demo/cube-1-250-250.
↪zarr
```

2.2.2 Test it

After starting the server, check the various functions provided by xcube Web API.

- **Datasets:**
    - Get datasets
    - Get dataset details

  - Get dataset coordinates
- **Color bars:**
  - Get color bars
  - Get color bars (HTML)
- **WMTS:**
  - Get WMTS KVP Capabilities (XML)
  - Get WMTS KVP local tile (PNG)
  - Get WMTS KVP remote tile (PNG)
  - Get WMTS REST Capabilities (XML)
  - Get WMTS REST local tile (PNG)
  - Get WMTS REST remote tile (PNG)
- **Tiles**
  - Get tile (PNG)
  - Get tile grid for OpenLayers 4.x
  - Get tile grid for Cesium 1.x
  - Get legend for layer (PNG)
- **Time series service (preliminary & unstable, will likely change soon)**
  - Get time stamps per dataset
  - Get time series for single point
- **Places service (preliminary & unstable, will likely change soon>'_**
  - Get all features
  - Get all features of collection "inside-cube"
  - Get all features for dataset "local"
  - Get all features of collection "inside-cube" for dataset "local"

2.2.3 xcube Viewer

xcube datasets published through `xcube serve` can be visualised using the xcube-viewer web application. To do so, run `xcube serve` with the `--show` flag.

In order make this option usable, xcube-viewer must be installed and build:

1. Download and install yarn.
2. Download and build xcube-viewer:

```
$ git clone https://github.com/dcs4cop/xcube-viewer.git
$ cd xcube-viewer
$ yarn build
```

3. Configure `xcube serve` so it finds the xcube-viewer On Linux (please adjust path):

```
$ export XCUBE_VIEWER_PATH=/abs/path/to/xcube-viewer/build

On Windows (please adjust path):
```

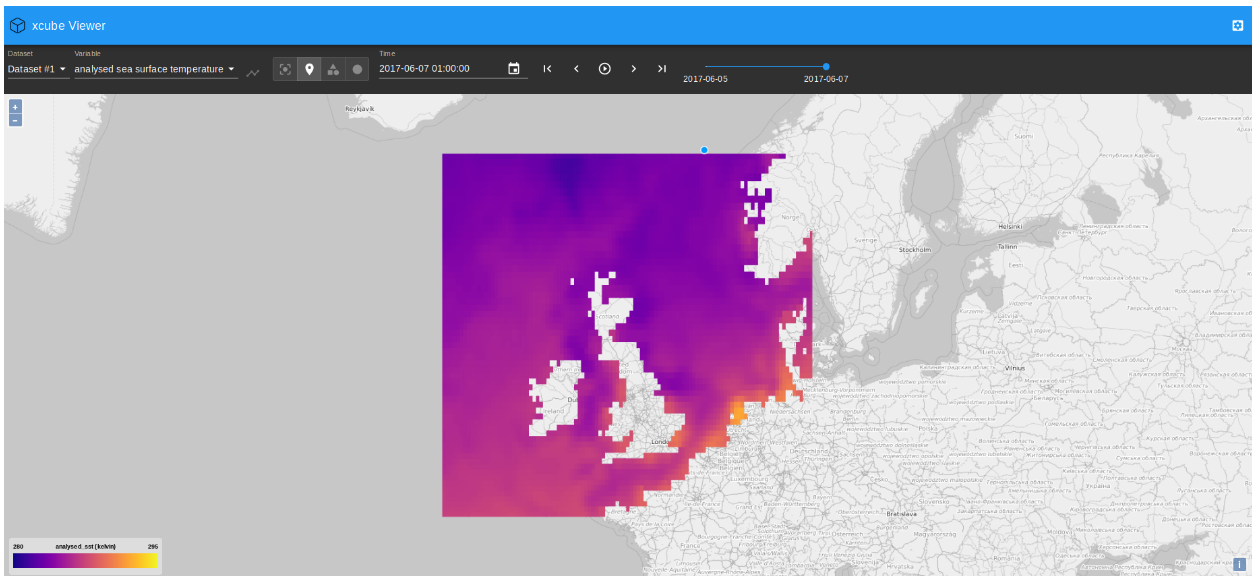
```
> SET XCUBE_VIEWER_PATH=/abs/path/to/xcube-viewer/build
```

4. Then run `xcube serve --show`:

```
$ xcube serve --show --styles conc_chl=(0,20,"viridis") examples/serve/demo/cube-1-
↪250-250.zarr
```

Viewing the generated xcube dataset described in the example *Generating an xcube dataset*:

```
$ xcube serve --show --styles "analysed_sst=(280,290,'plasma')" demo_SST_xcube-
↪optimized.zarr
```



In case you get an error message "cannot reach server" on the very bottom of the web app's main window, refresh the page.

You can play around with the value range displayed in the viewer, here it is set to min=280K and max=290K. The colormap used for mapping can be modified as well and the colormaps provided by matplotlib can be used.

2.2.4 Other clients

There are example HTML pages for some tile server clients. They need to be run in a web server. If you don't have one, you can use Node's `httpserver`:

```
$ npm install -g httpserver
```

After starting both the xcube server and web server, e.g. on port 9090:

```
$ httpserver -d -p 9090
```

you can run the client demos by following their links given below.

OpenLayers

- OpenLayers 4 Demo
- OpenLayers 4 Demo with WMTS

Cesium

To run the Cesium Demo first download Cesium and unpack the zip into the `xcube serve` source directory so that there exists an `./Cesium-x.y.z` sub-directory. You may have to adapt the Cesium version number in the demo's HTML file.

CHAPTER

THREE

INSTALLATION

3.1 Installation using conda

Into existing conda environment (>= Python 3.7)

```
$ git install -c conda-forge xcube
```

Into new conda environment

```
$ git create -c conda-forge -n xcube python3
$ git install -c conda-forge xcube
```

3.2 Installation from sources

First

```
$ git clone https://github.com/dcs4cop/xcube.git
$ cd xcube
$ conda env create
```

Then

```
$ activate xcube
$ python setup.py develop
```

Update

```
$ activate xcube
$ git pull --force
$ python setup.py develop
```

Run tests

```
$ pytest
```

with coverage

```
$ pytest --cov=xcube
```

with coverage report in HTML

```
$ pytest --cov-report html --cov=xcube
```

3.3 Docker

To start a demo using docker use the following commands

```
$ docker build -t [your name] .
$ docker run -d -p [host port]:8000 [your name]
```

Example:

```
$  docker build -t xcube:0.1.0dev6 .
$  docker run -d -p 8001:8000 xcube:0.1.0dev6
$  docker ps
```

CHAPTER

FOUR

CLI

The xcube command-line interface (CLI) is a single executable cli/xcube with several sub-commands comprising functions ranging from xcube dataset generation, over analysis and manipulation, to dataset publication.

4.1 Common Arguments and Options

Most of the commands operate on inputs that are xcube datasets. Such inputs are consistently named `CUBE` and provided as one or more command arguments. CUBE inputs may be a path into the local file system or a path into some object storage bucket, e.g. in AWS S3. Command inputs of other types are consistently called `INPUT`.

Many commands also output something, i.e. are writing files. The paths or names of such outputs are consistently provided by the `-o OUTPUT` or `--output OUTPUT` option. As the output is an option, there is usually a default value for it. If multiply file formats are supported, commands usually provide a `-f FORMAT` or `--format FORMAT` option. If omitted, the format may be guessed from the output's name.

4.2 Cube generation

4.2.1 `xcube gen`

Synopsis

Generate xcube dataset.

```
$ xcube gen --help
```

```
Usage: xcube gen [OPTIONS] [INPUT]...

  Generate xcube dataset. Data cubes may be created in one go or
  successively for all given inputs. Each input is expected to provide a
  single time slice which may be appended, inserted or which may replace an
  existing time slice in the output dataset. The input paths may be one or
  more input files or a pattern that may contain wildcards '?', '*', and
  '**'. The input paths can also be passed as lines of a text file. To do
  so, provide exactly one input file with ".txt" extension which contains
  the actual input paths to be used.

Options:
  -P, --proc INPUT-PROCESSOR      Input processor name. The available input
                                  processor names and additional information
```

(continues on next page)

(continued from previous page)

```
                                   about input processors can be accessed by
                                   calling xcube gen --info . Defaults to
                                   "default", an input processor that can deal
                                   with simple datasets whose variables have
                                   dimensions ("lat", "lon") and conform with
                                   the CF conventions.
  -c, --config CONFIG              xcube dataset configuration file in YAML
                                   format. More than one config input file is
                                   allowed.When passing several config files,
                                   they are merged considering the order passed
                                   via command line.
  -o, --output OUTPUT              Output path. Defaults to 'out.zarr'
  -f, --format FORMAT              Output format. Information about output
                                   formats can be accessed by calling xcube gen
                                   --info. If omitted, the format will be
                                   guessed from the given output path.
  -S, --size SIZE                  Output size in pixels using format
                                   "<width>,<height>".
  -R, --region REGION              Output region using format "<lon-min>,<lat-
                                   min>,<lon-max>,<lat-max>"
  --variables, --vars VARIABLES    Variables to be included in output. Comma-
                                   separated list of names which may contain
                                   wildcard characters "*" and "?".
  --resampling␣
↪[Average|Bilinear|Cubic|CubicSpline|Lanczos|Max|Median|Min|Mode|Nearest|Q1|Q3]
                                   Fallback spatial resampling algorithm to be
                                   used for all variables. Defaults to
                                   'Nearest'. The choices for the resampling
                                   algorithm are: ['Average', 'Bilinear',
                                   'Cubic', 'CubicSpline', 'Lanczos', 'Max',
                                   'Median', 'Min', 'Mode', 'Nearest', 'Q1',
                                   'Q3']
  -a, --append                     Deprecated. The command will now always
                                   create, insert, replace, or append input
                                   slices.
  --prof                           Collect profiling information and dump
                                   results after processing.
  --sort                           The input file list will be sorted before
                                   creating the xcube dataset. If --sort
                                   parameter is not passed, order of input list
                                   will be kept.
  -I, --info                       Displays additional information about format
                                   options or about input processors.
  --dry_run                        Just read and process inputs, but don't
                                   produce any outputs.
  --help                           Show this message and exit.
```

Below is the ouput of a `xcube gen --info` call showing five input processors installed via plugins.

```
$ xcube gen --info
```

```
input processors to be used with option --proc:
  default                          Single-scene NetCDF/CF inputs in xcube standard␣
↪format
  rbins-seviri-highroc-scene-l2    RBINS SEVIRI HIGHROC single-scene Level-2 NetCDF␣
↪inputs
```

(continues on next page)

(continued from previous page)

```
  rbins-seviri-highroc-daily-l2     RBINS SEVIRI HIGHROC daily Level-2 NetCDF inputs
  snap-olci-highroc-l2              SNAP Sentinel-3 OLCI HIGHROC Level-2 NetCDF inputs
  snap-olci-cyanoalert-l2           SNAP Sentinel-3 OLCI CyanoAlert Level-2 NetCDF␣
↪inputs
  vito-s2plus-l2                    VITO Sentinel-2 Plus Level 2 NetCDF inputs

For more input processors use existing "xcube-gen-..." plugins from the github␣
↪organisation DCS4COP or write own plugin.


output formats to be used with option --format:
  csv                       (*.csv)       CSV file format
  mem                       (*.mem)       In-memory dataset I/O
  netcdf4                   (*.nc)        NetCDF-4 file format
  zarr                      (*.zarr)      Zarr file format (http://zarr.readthedocs.io)
```

Configuration File

Configuration files passed to `xcube gen` via the `-c, --config` option use YAML format. Multiple configuration files may be given. In this case all configurations are merged into a single one. Parameter values will be overwritten by subsequent configurations if they are scalars. If they are objects / mappings, their values will be deeply merged.

The following parameters can be used in the configuration files:

`input_processor` [str] The name of an *input processor*. See `-P, --proc` option above.

Default The default value is `'default'`, xcube's default input processor. It can ingest and process inputs that

- use an `EPSG:4326` (or compatible) grid;
- have 1-D `lon` and `lat` coordinate variables using WGS84 coordinates and decimal degrees;
- have a decodable 1-D `time` coordinate or define the one of the following global attribute pairs `time_coverage_start` and `time_coverage_end`, `time_start` and `time_end` or `time_stop`;
- provide data variables with the dimensions `time`, `lat`, `lon`, in this order.
- conform to the **'CF Conventions'_**.

`output_size` [[int, int]] The spatial dimension sizes of the output dataset given as number of grid cells in longitude and latitude direction (width and height).

`output_region` [[float, float, float, float]] The spatial extent of output datasets given as a bounding box [lat-min, lat-min, lon-max, lat-max] using decimal degrees.

`output_variables` [[*variable-definitions*]] The definition of variables that will be included in the output dataset. Each variable definition may be just a name or a mapping from a name to variable attributes. If it is just a name it must be the name of an existing variable either in the INPUT or in `processed_variables`. If the variable definition is a mapping, some of the attributes affect the way how variables are processed. All but the `name` attributes become variable metadata in the output.

`name` [str] The new name of the variable in the output.

`valid_pixel_expression` [str] An expression used to mask this variable, see *Expressions*. The expression identifies all valid pixels in each INPUT.

`resampling` [str] The resampling method used. See `--resampling` option above.

Default By default, all variables in INPUT will occur in output.

processed_variables [[*variable-definitions*]] The definition of variables that will be produced or processed after reading each INPUT. The main purpose is to generate intermediate variables that can be referred to in the `expression` in other variable definitions in `processed_variables` and `valid_pixel_expression` in variable definitions in `output_variables`. The following attributes are recognised:

expression [str] An expression used to produce this variable, see *Expressions*.

output_writer_name [str] The name of a supported output format. May be one of `'zarr'`, `'netcdf4'`, `'mem'`.

Default `'zarr'`

output_writer_params [str] A mapping that defines parameters that are passed to output writer denoted by `output_writer_name`.

output_metadata [[*attribute-definitions*]] General metadata that will be present in the output dataset as global attributes. You can put any common CF attributes here.

Any attributes that are mappings will be "flattened" by concatenating the attribute names using the underscrore character. For example,:

```
publisher:
  name:  "Brockmann Consult GmbH"
  url:   "https://www.brockmann-consult.de"
```

will create the two entries:

```
publisher_name:  "Brockmann Consult GmbH"
publisher_url:   "https://www.brockmann-consult.de"
```

Expressions

Expressions are plain text values of the `expression` and `valid_pixel_expression` attributes of the variable definitions in the `processed_variables` and `output_variables` parameters. The expression syntax is that of standard Python. `xcube gen` uses expressions to produce new variables listed in `processed_variables` and to mask variables by the `valid_pixel_expression`.

An expression may refer any variables in the INPUT datasets and any variables defined by the `processed_variables` parameter. Expressions may make use of most of the standard Python operators and may apply all numpy ufuncs to referred variables. Also most of the xarray.DataArray API may be used on variables within an expression.

In order to utilise flagged variables, the syntax `variable_name.flag_name` can be used in expressions. According to the CF Conventions, flagged variables are variables whose metadata include the attributes `flag_meanings` and `flag_values` and/or `flag_masks`. The `flag_meanings` attribute enumerates the allowed values for `flag_name`. The flag attributes must be present in the variables of each INPUT.

Example

An example that uses a configuration file only:

```
$ xcube gen --config ./config.yml /data/eo-data/SST/2018/**/*.nc
```

An example that uses the default input processor and passes all other configuration via command-line options:

```
$ xcube gen -S 2000,1000 -R 0,50,5,52.5 --vars conc_chl,conc_tsm,kd489,c2rcc_flags,
↪quality_flags -o hiroc-cube.zarr /data/eo-data/SST/2018/**/*.nc
```

Some input processors have been developed for specific EO data sources used within the DCS4COP project. They may serve as examples how to develop input processor plug-ins:

- xcube-gen-rbins
- xcube-gen-bc
- xcube-gen-vito

Python API

The related Python API function is `xcube.api.gen_cube()`.

4.2.2 xcube grid

Attention: This tool will likely change in the near future.

Synopsis

Find spatial xcube dataset resolutions and adjust bounding boxes.

```
$ xcube grid --help
```

```
Usage: xcube grid [OPTIONS] COMMAND [ARGS]...

  Find spatial xcube dataset resolutions and adjust bounding boxes.

  We find suitable resolutions with respect to a possibly regional fixed
  Earth grid and adjust regional spatial bounding boxes to that grid. We
  also try to select the resolutions such that they are taken from a certain
  level of a multi-resolution pyramid whose level resolutions increase by a
  factor of two.

  The graticule at a given resolution level L within the grid is given by

      RES(L) = COVERAGE * HEIGHT(L)
      HEIGHT(L) = HEIGHT_0 * 2 ^ L
      LON(L, I) = LON_MIN + I * HEIGHT_0 * RES(L)
      LAT(L, J) = LAT_MIN + J * HEIGHT_0 * RES(L)

  With

      RES:      Grid resolution in degrees.
      HEIGHT:   Number of vertical grid cells for given level
      HEIGHT_0: Number of vertical grid cells at lowest resolution level.

  Let WIDTH and HEIGHT be the number of horizontal and vertical grid cells
  of a global grid at a certain LEVEL with WIDTH * RES = 360 and HEIGHT *
  RES = 180, then we also force HEIGHT = TILE * 2 ^ LEVEL.
```

(continues on next page)

(continued from previous page)

```
Options:
  --help  Show this message and exit.

Commands:
  abox    Adjust a bounding box to a fixed Earth grid.
  levels  List levels for a resolution or a tile size.
  res     List resolutions close to a target resolution.
```

Example: Find suitable target resolution for a ~300m (Sentinel 3 OLCI FR resolution) fixed Earth grid within a deviation of 5%.

```
$ xcube grid res 300m -D 5%
```

```
TILE     LEVEL    HEIGHT   INV_RES  RES (deg)        RES (m), DELTA_RES (%)
540      7        69120    384      0.0026041666666666665   289.9   -3.4
4140     4        66240    368      0.002717391304347826    302.5   0.8
8100     3        64800    360      0.002777777777777778    309.2   3.1
...
```

289.9m is close enough and provides 7 resolution levels, which is good. Its inverse resolution is 384, which is the fixed Earth grid identifier.

We want to see if the resolution pyramid also supports a resolution close to 10m (Sentinel 2 MSI resolution).

```
$ xcube grid levels 384 -m 6
```

```
LEVEL    HEIGHT   INV_RES  RES (deg)        RES (m)
0        540      3        0.3333333333333333      37106.5
1        1080     6        0.16666666666666666     18553.2
2        2160     12       0.08333333333333333     9276.6
...
11       1105920  6144     0.00016276041666666666  18.1
12       2211840  12288    8.138020833333333e-05   9.1
13       4423680  24576    4.0690104166666664e-05  4.5
```

This indicates we have a resolution of 9.1m at level 12.

Lets assume we have xcube dataset region with longitude from 0 to 5 degrees and latitudes from 50 to 52.5 degrees. What is the adjusted bounding box on a fixed Earth grid with the inverse resolution 384?

```
$ xcube grid abox  0,50,5,52.5  384
```

```
Orig. box coord. = 0.0,50.0,5.0,52.5
Adj. box coord.  = 0.0,49.21875,5.625,53.4375
Orig. box WKT    = POLYGON ((0.0 50.0, 5.0 50.0, 5.0 52.5, 0.0 52.5, 0.0 50.0))
Adj. box WKT     = POLYGON ((0.0 49.21875, 5.625 49.21875, 5.625 53.4375, 0.0 53.4375,
↪ 0.0 49.21875))
Grid size  = 2160 x 1620 cells
with
  TILE      = 540
  LEVEL     = 7
  INV_RES   = 384
  RES (deg) = 0.0026041666666666665
  RES (m)   = 289.89450727414993
```

Note, to check bounding box WKTs, you can use the handy Wicket tool.

4.3 Cube inspection

4.3.1 `xcube dump`

Synopsis

Dump contents of a dataset.

```
$ xcube dump --help
```

```
Usage: xcube dump [OPTIONS] INPUT

  Dump contents of an input dataset.

Options:
  --variable, --var VARIABLE
                                  Name of a variable (multiple allowed).
  -E, --encoding                  Dump also variable encoding information.
  --help                          Show this message and exit.
```

Example

```
$ xcube dump xcube_cube.zarr
```

4.3.2 `xcube verify`

Synopsis

Perform cube verification.

```
$ xcube verify --help
```

```
Usage: xcube verify [OPTIONS] CUBE

  Perform cube verification.

  The tool verifies that CUBE
  * defines the dimensions "time", "lat", "lon";
  * has corresponding "time", "lat", "lon" coordinate variables and that they
    are valid, e.g. 1-D, non-empty, using correct units;
  * has valid  bounds variables for "time", "lat", "lon" coordinate
    variables, if any;
  * has any data variables and that they are valid, e.g. min. 3-D, all have
    same dimensions, have at least dimensions "time", "lat", "lon".

  If INPUT is a valid xcube dataset, the tool returns exit code 0. Otherwise a
  violation report is written to stdout and the tool returns exit code 3.

Options:
  --help  Show this message and exit.
```

Python API

The related Python API functions are *xcube.api.verify_cube()* and *xcube.api.assert_cube()*.

4.4 Cube data extraction

4.4.1 xcube extract

Synopsis

Extract cube points.

```
$ xcube extract --help
```

```
Usage: xcube extract [OPTIONS] CUBE POINTS

  Extract data points from an xcube dataset.

  Extracts data cells from CUBE at coordinates given in each POINTS record
  and writes the resulting values to given output path and format.

  POINTS must be a CSV file that provides at least the columns "lon", "lat",
  and "time". The "lon" and "lat" columns provide a point's location in
  decimal degrees. The "time" column provides a point's date or date-time.
  Its format should preferably be ISO, but other formats may work as well.

Options:
  -o, --output OUTPUT    Output path. If omitted, output is written to stdout.
  -f, --format FORMAT    Output format. Currently, only 'csv' is supported.
  -C, --coords           Include cube cell coordinates in output.
  -B, --bounds           Include cube cell coordinate boundaries (if any) in
                         output.
  -I, --indexes          Include cube cell indexes in output.
  -R, --refs             Include point values as reference in output.
  --help                 Show this message and exit.
```

Example

```
$ xcube extract xcube_cube.zarr -o point_data.csv -Cb --indexes --refs
```

Python API

Related Python API functions are *xcube.api.get_cube_values_for_points()*, *xcube.api.get_cube_point_indexes()*, and *xcube.api.get_cube_values_for_indexes()*.

4.5 Cube manipulation

4.5.1 xcube chunk

Synopsis

(Re-)chunk xcube dataset.

```
$ xcube chunk --help
```

```
Usage: xcube chunk [OPTIONS] CUBE

  (Re-)chunk xcube dataset. Changes the external chunking of all variables
  of CUBE according to CHUNKS and writes the result to OUTPUT.

Options:
  -o, --output OUTPUT  Output path. Defaults to 'out.zarr'
  -f, --format FORMAT  Format of the output. If not given, guessed from
                       OUTPUT.
  -p, --params PARAMS  Parameters specific for the output format. Comma-
                       separated list of <key>=<value> pairs.
  -C, --chunks CHUNKS  Chunk sizes for each dimension. Comma-separated list of
                       <dim>=<size> pairs, e.g. "time=1,lat=270,lon=270"
  --help               Show this message and exit.
```

Example

```
$ xcube chunk input_not_chunked.zarr -o output_rechunked.zarr --chunks "time=1,
↪lat=270,lon=270"
```

Python API

The related Python API function is `xcube.api.chunk_dataset()`.

4.5.2 `xcube level`

Synopsis

Generate multi-resolution levels.

```
$ xcube level --help
```

```
Usage: xcube level [OPTIONS] INPUT

  Generate multi-resolution levels. Transform the given dataset by INPUT
  into the levels of a multi-level pyramid with spatial resolution
  decreasing by a factor of two in both spatial dimensions and write the
  result to directory OUTPUT.

Options:
  -o, --output OUTPUT             Output path. If omitted, "INPUT.levels" will
                                  be used.
  -L, --link                      Link the INPUT instead of converting it to a
                                  level zero dataset. Use with care, as the
                                  INPUT's internal spatial chunk sizes may be
                                  inappropriate for imaging purposes.
```

(continues on next page)

(continued from previous page)

```
  -t, --tile-size TILE-SIZE       Tile size, given as single integer number or
                                  as <tile-width>,<tile-height>. If omitted,
                                  the tile size will be derived from the
                                  INPUT's internal spatial chunk sizes. If the
                                  INPUT is not chunked, tile size will be 512.
  -n, --num-levels-max NUM-LEVELS-MAX
                                  Maximum number of levels to generate. If not
                                  given, the number of levels will be derived
                                  from spatial dimension and tile sizes.
  --help                          Show this message and exit.
```

Example

```
$ xcube level --link -t 720 data/cubes/test-cube.zarr
```

Python API

The related Python API function are *xcube.api.compute_levels()*, *xcube.api.read_levels()*, and *xcube.api.write_levels()*.

4.5.3 xcube optimize

Synopsis

Optimize xcube dataset for faster access.

```
$ xcube optimize --help
```

```
Usage: xcube optimize [OPTIONS] CUBE

  Optimize xcube dataset for faster access.

  Reduces the number of metadata and coordinate data files in xcube dataset
  given by CUBE. Consolidated cubes open much faster especially from remote
  locations, e.g. in object storage, because obviously much less HTTP
  requests are required to fetch initial cube meta information. That is, it
  merges all metadata files into a single top-level JSON file ".zmetadata".
  Optionally, it removes any chunking of coordinate variables so they
  comprise a single binary data file instead of one file per data chunk. The
  primary usage of this command is to optimize data cubes for cloud object
  storage. The command currently works only for data cubes using ZARR
  format.

Options:
  -o, --output OUTPUT  Output path. The placeholder "<built-in function
                       input>" will be replaced by the input's filename
                       without extension (such as ".zarr"). Defaults to
                       "{input}-optimized.zarr".
  -I, --in-place       Optimize cube in place. Ignores output path.
  -C, --coords         Also optimize coordinate variables by converting any
                       chunked arrays into single, non-chunked, contiguous
```

(continues on next page)

(continued from previous page)

```
                         arrays.
  --help                 Show this message and exit.
```

Examples

Write an cube with consolidated metadata to `cube-optimized.zarr`:

```
$ xcube optimize ./cube.zarr
```

Write an optimized cube with consolidated metadata and consolidated coordinate variables to `optimized/cube.zarr` (directory `optimized` must exist):

```
$ xcube optimize -C -o ./optimized/cube.zarr ./cube.zarr
```

Optimize a cube in-place with consolidated metadata and consolidated coordinate variables:

```
$ xcube optimize -IC ./cube.zarr
```

Python API

The related Python API function is *xcube.api.optimize_dataset()*.

4.5.4 xcube prune

Delete empty chunks.

Attention: This tool will likely be integrated into `xcube optimize` in the near future.

```
$ xcube prune --help
```

```
Usage: xcube prune [OPTIONS] CUBE

  Delete empty chunks. Deletes all data files associated with empty (NaN-
  only) chunks in given CUBE, which must have ZARR format.

Options:
  --dry-run  Just read and process input, but don't produce any outputs.
  --help     Show this message and exit.
```

A related Python API function is `xcube.api.get_empty_dataset_chunks()`.

4.5.5 xcube resample

Synopsis

Resample data along the time dimension.

```
$ xcube resample --help
```

```
Usage: xcube resample [OPTIONS] CUBE

  Resample data along the time dimension.

Options:
  -c, --config CONFIG             xcube dataset configuration file in YAML
                                  format. More than one config input file is
                                  allowed.When passing several config files,
                                  they are merged considering the order passed
                                  via command line.
  -o, --output OUTPUT             Output path. Defaults to 'out.zarr'.
  -f, --format [zarr|netcdf4|mem]
                                  Output format. If omitted, format will be
                                  guessed from output path.
  --variables, --vars VARIABLES   Comma-separated list of names of variables
                                  to be included.
  -M, --method TEXT               Temporal resampling method. Available
                                  downsampling methods are 'count', 'first',
                                  'last', 'min', 'max', 'sum', 'prod', 'mean',
                                  'median', 'std', 'var', the upsampling
                                  methods are 'asfreq', 'ffill', 'bfill',
                                  'pad', 'nearest', 'interpolate'. If the
                                  upsampling method is 'interpolate', the
                                  option '--kind' will be used, if given.
                                  Other upsampling methods that select
                                  existing values honour the '--tolerance'
                                  option. Defaults to 'mean'.
  -F, --frequency TEXT            Temporal aggregation frequency. Use format
                                  "<count><offset>" where <offset> is one of
                                  'H', 'D', 'W', 'M', 'Q', 'Y'. Defaults to
                                  '1D'.
  -O, --offset TEXT               Offset used to adjust the resampled time
                                  labels. Uses same syntax as frequency. Some
                                  Pandas date offset strings are supported as
                                  well.
  -B, --base INTEGER              For frequencies that evenly subdivide 1 day,
                                  the origin of the aggregated intervals. For
                                  example, for '24H' frequency, base could
                                  range from 0 through 23. Defaults to 0.
  -K, --kind TEXT                 Interpolation kind which will be used if
                                  upsampling method is 'interpolation'. May be
                                  one of 'zero', 'slinear', 'quadratic',
                                  'cubic', 'linear', 'nearest', 'previous',
                                  'next' where 'zero', 'slinear', 'quadratic',
                                  'cubic' refer to a spline interpolation of
                                  zeroth, first, second or third order;
                                  'previous' and 'next' simply return the
                                  previous or next value of the point. For
                                  more info refer to
                                  scipy.interpolate.interp1d(). Defaults to
                                  'linear'.
  -T, --tolerance TEXT            Tolerance for selective upsampling methods.
                                  Uses same syntax as frequency. If the time
                                  delta exceeds the tolerance, fill values
                                  (NaN) will be used. Defaults to the given
```

(continues on next page)

(continued from previous page)

```
                                frequency.
  --dry-run                     Just read and process inputs, but don't
                                produce any outputs.
  --help                        Show this message and exit.
```

Examples

Upsampling example:

```
$ xcube resample --vars conc_chl,conc_tsm -F 12H -T 6H -M interpolation -K linear␣
↪examples/serve/demo/cube.nc
```

Downsampling example:

```
$ xcube resample --vars conc_chl,conc_tsm -F 3D -M mean -M std -M count examples/
↪serve/demo/cube.nc
```

Python API

The related Python API function is *xcube.api.resample_in_time()*.

4.5.6 xcube vars2dim

Synopsis

Convert cube variables into new dimension.

```
$ xcube vars2dim --help
```

```
Usage: xcube vars2dim [OPTIONS] CUBE

  Convert cube variables into new dimension. Moves all variables of CUBE
  into into a single new variable <var-name> with a new dimension DIM-NAME
  and writes the results to OUTPUT.

Options:
  --variable, --var VARIABLE  Name of the new variable that includes all
                              variables. Defaults to "data".
  -D, --dim_name DIM-NAME     Name of the new dimension into variables.
                              Defaults to "var".
  -o, --output OUTPUT         Output path. If omitted, 'INPUT-vars2dim.INPUT-
                              FORMAT' will be used.
  -f, --format FORMAT         Format of the output. If not given, guessed from
                              OUTPUT.
  --help                      Show this message and exit.
```

Python API

The related Python API function is *xcube.api.vars_to_dim()*.

4.6 Cube publication

4.6.1 xcube serve

Synopsis

Serve data cubes via web service.

`xcube serve` starts a light-weight web server that provides various services based on xcube datasets:

- Catalogue services to query for xcube datasets and their variables and dimensions, and feature collections;
- Tile map service, with some OGC WMTS 1.0 compatibility (REST and KVP APIs);
- Dataset services to extract subsets like time-series and profiles for e.g. JavaScript clients.

```
$ xcube serve --help
```

```
Usage: xcube serve [OPTIONS] [CUBE]...

  Serve data cubes via web service.

  Serves data cubes by a RESTful API and a OGC WMTS 1.0 RESTful and KVP
  interface. The RESTful API documentation can be found at
  https://app.swaggerhub.com/apis/bcdev/xcube-server.

Options:
  -A, --address ADDRESS  Service address. Defaults to 'localhost'.
  -P, --port PORT        Port number where the service will listen on.
                         Defaults to 8080.
  --prefix PREFIX        Service URL prefix. May contain template patterns
                         such as "${version}" or "${name}". For example
                         "${name}/api/${version}".
  -u, --update PERIOD    Service will update after given seconds of
                         inactivity. Zero or a negative value will disable
                         update checks. Defaults to 2.0.
  -S, --styles STYLES    Color mapping styles for variables. Used only, if one
                         or more CUBE arguments are provided and CONFIG is not
                         given. Comma-separated list with elements of the form
                         <var>=(<vmin>,<vmax>) or
                         <var>=(<vmin>,<vmax>,"<cmap>")
  -c, --config CONFIG    Use datasets configuration file CONFIG. Cannot be
                         used if CUBES are provided.
  --tilecache SIZE       In-memory tile cache size in bytes. Unit suffixes
                         'K', 'M', 'G' may be used. Defaults to '512M'. The
                         special value 'OFF' disables tile caching.
  --tilemode MODE        Tile computation mode. This is an internal option
                         used to switch between different tile computation
                         implementations. Defaults to 0.
  -s, --show             Run viewer app. Requires setting the environment
                         variable XCUBE_VIEWER_PATH to a valid xcube-viewer
                         deployment or build directory. Refer to
                         https://github.com/dcs4cop/xcube-viewer for more
                         information.
  -v, --verbose          Delegate logging to the console (stderr).
  --traceperf            Print performance diagnostics (stdout).
  --help                 Show this message and exit.
```

Configuration File

The xcube server is used to configure the xcube datasets to be published.

xcube datasets are any datasets that

- that comply to Unidata's CDM and to the CF Conventions;
- that can be opened with the xarray Python library;
- that have variables that have at least the dimensions and shape (`time`, `lat`, `lon`), in exactly this order;
- that have 1D-coordinate variables corresponding to the dimensions;
- that have their spatial grid defined in the WGS84 (`EPSG:4326`) coordinate reference system.

The xcube server supports xcube datasets stored as local NetCDF files, as well as Zarr directories in the local file system or remote object storage. Remote Zarr datasets must be stored in publicly accessible, AWS S3 compatible object storage (OBS).

As an example, here is the configuration of the demo server.

To increase imaging performance, xcube datasets can be converted to multi-resolution pyramids using the cli/xcube_level tool. In the configuration, the format must be set to `'level'`. Leveled xcube datasets are configured this way:

```
Datasets:

  - Identifier: my_multi_level_dataset
    Title: "My Multi-Level Dataset"
    FileSystem: local
    Path: my_multi_level_dataset.level
    Format: level

  - ...
```

To increase time-series extraction performance, xcube datasets my be rechunked with larger chunk size in the `time` dimension using the cli/xcube_chunk tool. In the xcube server configuration a hidden dataset is given, and the it is referred to by the non-hidden, actual dataset using the `TimeSeriesDataset` setting:

```
Datasets:

  - Identifier: my_dataset
    Title: "My Dataset"
    FileSystem: local
    Path: my_dataset.zarr
    TimeSeriesDataset: my_dataset_opt_for_ts

  - Identifier: my_dataset_opt_for_ts
    Title: "My Dataset optimized for Time-Series"
    FileSystem: local
    Path: my_ts_opt_dataset.zarr
    Format: zarr
    Hidden: True

  - ...
```

Example

```
xcube serve --port 8080 --config ./examples/serve/demo/config.yml --verbose
```

```
xcube Server: WMTS, catalogue, data access, tile, feature, time-series services for
↪xarray-enabled data cubes, version 0.2.0
[I 190924 17:08:54 service:228] configuration file
↪'D:\\Projects\\xcube\\examples\\serve\\demo\\config.yml' successfully loaded
[I 190924 17:08:54 service:158] service running, listening on localhost:8080, try
↪http://localhost:8080/datasets
[I 190924 17:08:54 service:159] press CTRL+C to stop service
```

Web API

The xcube server has a dedicated Web API Documentation on SwaggerHub. It also lets you explore the API of existing xcube-servers.

The xcube server implements the OGC WMTS RESTful and KVP architectural styles of the OGC WMTS 1.0.0 specification. The following operations are supported:

- **GetCapabilities**: `/xcube/wmts/1.0.0/WMTSCapabilities.xml`
- **GetTile**: `/xcube/wmts/1.0.0/tile/{DatasetName}/{VarName}/{TileMatrix}/{TileCol}/{TileRow}.png`
- **GetFeatureInfo**: *in progress*

CHAPTER FIVE

PYTHON API

5.1 Cube I/O

`xcube.api.`**`read_cube`**(*input_path: str*, *format_name: str = None*, ***kwargs*) → xarray.core.dataset.Dataset

Read a xcube dataset from *input_path*. If *format* is not provided it will be guessed from *input_path*.

Parameters

- **`input_path`** – input path
- **`format_name`** – format, e.g. "zarr" or "netcdf4"
- **`kwargs`** – format-specific keyword arguments

Returns xcube dataset

`xcube.api.`**`open_cube`**(*input_path: str*, *format_name: str = None*, ***kwargs*) → xarray.core.dataset.Dataset

The `read_cube` function as context manager that auto-closes the cube read.

Parameters

- **`input_path`** – input path
- **`format_name`** – format, e.g. "zarr" or "netcdf4"
- **`kwargs`** – format-specific keyword arguments

Returns xcube dataset

5.2 Cube generation

`xcube.api.`**`gen_cube`**(*input_paths: Sequence[str] = None, input_processor_name: str = None, input_processor_params: Dict = None, input_reader_name: str = None, input_reader_params: Dict[str, Any] = None, output_region: Tuple[float, float, float, float] = None, output_size: Tuple[int, int] = [512, 512], output_resampling: str = 'Nearest', output_path: str = 'out.zarr', output_writer_name: str = None, output_writer_params: Dict[str, Any] = None, output_metadata: Dict[str, Any] = None, output_variables: List[Tuple[str, Optional[Dict[str, Any]]]] = None, processed_variables: List[Tuple[str, Optional[Dict[str, Any]]]] = None, profile_mode: bool = False, sort_mode: bool = False, append_mode: bool = None, dry_run: bool = False, monitor: Callable[[...], None] = None*) → bool

Generate a xcube dataset from one or more input files.

Parameters

- **sort_mode** –
- **input_paths** – The input paths.
- **input_processor_name** – Name of a registered input processor (xcube.api.gen.inputprocessor.InputProcessor) to be used to transform the inputs.
- **input_processor_params** – Parameters to be passed to the input processor.
- **input_reader_name** – Name of a registered input reader (xcube.api.util.dsio.DatasetIO).
- **input_reader_params** – Parameters passed to the input reader.
- **output_region** – Output region as tuple of floats: (lon_min, lat_min, lon_max, lat_max).
- **output_size** – The spatial dimensions of the output as tuple of ints: (width, height).
- **output_resampling** – The resampling method for the output.
- **output_path** – The output directory.
- **output_writer_name** – Name of an output writer (xcube.api.util.dsio.DatasetIO) used to write the cube.
- **output_writer_params** – Parameters passed to the output writer.
- **output_metadata** – Extra metadata passed to output cube.
- **output_variables** – Output variables.
- **processed_variables** – Processed variables computed on-the-fly.
- **profile_mode** – Whether profiling should be enabled.
- **append_mode** – Deprecated. The function will always either insert, replace, or append new time slices.
- **dry_run** – Doesn't write any data. For testing.
- **monitor** – A progress monitor.

Returns True for success.

xcube.api.**new_cube**(*title='Test Cube'*, *width=360*, *height=180*, *spatial_res=1.0*, *lon_start=-180.0*, *lat_start=-90.0*, *time_periods=5*, *time_freq='D'*, *time_start='2010-01-01T00:00:00'*, *inverse_lat=False*, *drop_bounds=False*, *variables=None*)

Create a new empty cube. Useful for creating cubes templates with predefined coordinate variables and metadata. The function is also heavily used by xcube's unit tests.

The values of the *variables* dictionary can be either constants, array-like objects, or functions that compute their return value from passed coordinate indexes. The expected signature is::

```
def my_func(time: int, lat: int, lon: int) -> Union[bool, int, float]
```

Parameters

- **title** – A title.
- **width** – Horizontal number of grid cells
- **height** – Vertical number of grid cells
- **spatial_res** – Spatial resolution in degrees
- **lon_start** – Minimum longitude value

- **lat_start** – Minimum latitude value
- **time_periods** – Number of time steps
- **time_freq** – Duration of each time step
- **time_start** – First time value
- **inverse_lat** – Whether to create an inverse latitude axis
- **drop_bounds** – If True, coordinate bounds variables are not created.
- **variables** – Dictionary of data variables to be added.

Returns A cube instance

5.3 Cube data extraction

xcube.api.**get_cube_values_for_points**(*cube: xarray.core.dataset.Dataset, points: Union[xarray.core.dataset.Dataset, pandas.core.frame.DataFrame, Mapping[str, Any]], var_names: Sequence[str] = None, include_coords: bool = False, include_bounds: bool = False, include_indexes: bool = False, index_name_pattern: str = '{name}_index', include_refs: bool = False, ref_name_pattern: str = '{name}_ref', method: str = 'nearest', cube_asserted: bool = False*) → xarray.core.dataset.Dataset

Extract values from *cube* variables at given coordinates in *points*.

Parameters

- **cube** – The cube dataset.
- **points** – Dictionary that maps dimension name to coordinate arrays.
- **var_names** – An optional list of names of data variables in *cube* whose values shall be extracted.
- **include_coords** – Whether to include the cube coordinates for each point in return value.
- **include_bounds** – Whether to include the cube coordinate boundaries (if any) for each point in return value.
- **include_indexes** – Whether to include computed indexes into the cube for each point in return value.
- **index_name_pattern** – A naming pattern for the computed index columns. Must include "{name}" which will be replaced by the index' dimension name.
- **include_refs** – Whether to include point (reference) values in return value.
- **ref_name_pattern** – A naming pattern for the computed point data columns. Must include "{name}" which will be replaced by the point's attribute name.
- **method** – "nearest" or "linear".
- **cube_asserted** – If False, *cube* will be verified, otherwise it is expected to be a valid cube.

Returns A new data frame whose columns are values from *cube* variables at given *points*.

`xcube.api.`**`get_cube_point_indexes`**(*cube: xarray.core.dataset.Dataset, points: Union[xarray.core.dataset.Dataset, pandas.core.frame.DataFrame, Mapping[str, Any]], dim_name_mapping: Mapping[str, str] = None, index_name_pattern: str = '{name}_index', index_dtype=<class 'numpy.float64'>, cube_asserted: bool = False*) → xarray.core.dataset.Dataset

Get indexes of given point coordinates *points* into the given *dataset*.

Parameters

- **`cube`** – The cube dataset.
- **`points`** – A mapping from column names to column data arrays, which must all have the same length.
- **`dim_name_mapping`** – A mapping from dimension names in *cube* to column names in *points*.
- **`index_name_pattern`** – A naming pattern for the computed indexes columns. Must include "{name}" which will be replaced by the dimension name.
- **`index_dtype`** – Numpy data type for the indexes. If it is a floating point type (default), then *indexes* will contain fractions, which may be used for interpolation. For out-of-range coordinates in *points*, indexes will be -1 if *index_dtype* is an integer type, and NaN, if *index_dtype* is a floating point types.
- **`cube_asserted`** – If False, *cube* will be verified, otherwise it is expected to be a valid cube.

Returns A dataset containing the index columns.

`xcube.api.`**`get_cube_values_for_indexes`**(*cube: xarray.core.dataset.Dataset, indexes: Union[xarray.core.dataset.Dataset, pandas.core.frame.DataFrame, Mapping[str, Any]], include_coords: bool = False, include_bounds: bool = False, data_var_names: Sequence[str] = None, index_name_pattern: str = '{name}_index', method: str = 'nearest', cube_asserted: bool = False*) → xarray.core.dataset.Dataset

Get values from the *cube* at given *indexes*.

Parameters

- **`cube`** – A cube dataset.
- **`indexes`** – A mapping from column names to index and fraction arrays for all cube dimensions.
- **`include_coords`** – Whether to include the cube coordinates for each point in return value.
- **`include_bounds`** – Whether to include the cube coordinate boundaries (if any) for each point in return value.
- **`data_var_names`** – An optional list of names of data variables in *cube* whose values shall be extracted.
- **`index_name_pattern`** – A naming pattern for the computed indexes columns. Must include "{name}" which will be replaced by the dimension name.
- **`method`** – "nearest" or "linear".

- **cube_asserted** – If False, *cube* will be verified, otherwise it is expected to be a valid cube.

Returns A new data frame whose columns are values from *cube* variables at given *indexes*.

`xcube.api.`**`get_dataset_indexes`**(*dataset: xarray.core.dataset.Dataset, coord_var_name: str, coord_values: Union[xarray.core.dataarray.DataArray, numpy.ndarray], index_dtype=<class 'numpy.float64'>*) → Union[xarray.core.dataarray.DataArray, numpy.ndarray]

Compute the indexes and their fractions into a coordinate variable *coord_var_name* of a *dataset* for the given coordinate values *coord_values*.

The coordinate variable's labels must be monotonic increasing or decreasing, otherwise the result will be nonsense.

For any value in *coord_values* that is out of the bounds of the coordinate variable's values, the index depends on the value of *index_dtype*. If *index_dtype* is an integer type, invalid indexes are encoded as -1 while for floating point types, NaN will be used.

Returns a tuple of indexes as int64 array and fractions as float64 array.

Parameters

- **dataset** – A cube dataset.
- **coord_var_name** – Name of a coordinate variable.
- **coord_values** – Array-like coordinate values.
- **index_dtype** – Numpy data type for the indexes. If it is floating point type (default), then *indexes* contain fractions, which may be used for interpolation. If *dtype* is an integer type out-of-range coordinates are indicated by index -1, and NaN if it is is a floating point type.

Returns The indexes and their fractions as a tuple of numpy int64 and float64 arrays.

`xcube.api.`**`get_time_series`**(*cube: xarray.core.dataset.Dataset, geometry: Union[shapely.geometry.base.BaseGeometry, Dict[str, Any], str, Sequence[Union[float, int]]] = None, var_names: Sequence[str] = None, start_date: Union[numpy.datetime64, str] = None, end_date: Union[numpy.datetime64, str] = None, include_count: bool = False, include_stdev: bool = False, use_groupby: bool = False, cube_asserted: bool = False*) → Optional[xarray.core.dataset.Dataset]

Get a time series dataset from a data *cube*.

geometry may be provided as a (shapely) geometry object, a valid GeoJSON object, a valid WKT string, a sequence of box coordinates (x1, y1, x2, y2), or point coordinates (x, y). If *geometry* covers an area, i.e. is not a point, the function aggregates the variables to compute a mean value and if desired, the number of valid observations and the standard deviation.

start_date and *end_date* may be provided as a numpy.datetime64 or an ISO datetime string.

Returns a time-series dataset whose data variables have a time dimension but no longer have spatial dimensions, hence the resulting dataset's variables will only have N-2 dimensions. A global attribute `max_number_of_observations` will be set to the maximum number of observations that could have been made in each time step. If the given *geometry* does not overlap the cube's boundaries, or if not output variables remain, the function returns `None`.

Parameters

- **cube** – The xcube dataset
- **geometry** – Optional geometry

- **var_names** – Optional sequence of names of variables to be included.
- **start_date** – Optional start date.
- **end_date** – Optional end date.
- **include_count** – Whether to include the number of valid observations for each time step. Ignored if geometry is a point.
- **include_stdev** – Whether to include standard deviation for each time step. Ignored if geometry is a point.
- **use_groupby** – Use group-by operation. May increase or decrease runtime performance and/or memory consumption.
- **cube_asserted** – If False, *cube* will be verified, otherwise it is expected to be a valid cube.

Returns A new dataset with time-series for each variable.

5.4 Cube manipulation

xcube.api.**resample_in_time**(*cube: xarray.core.dataset.Dataset, frequency: str, method: Union[str, Sequence[str]], offset=None, base: str = 0, tolerance=None, interp_kind=None, var_names: Sequence[str] = None, metadata: Dict[str, Any] = None*)

Resample a xcube dataset in the time dimension.

Parameters

- **cube** – The xcube dataset.
- **frequency** – Resampling frequency.
- **method** – Resampling method or sequence of resampling methods.
- **offset** – Offset used to adjust the resampled time labels. Some pandas date offset strings are supported.
- **base** – Resampling method.
- **var_names** – Variable names to include.
- **tolerance** – Time tolerance for selective upsampling methods. Defaults to *frequency*.
- **interp_kind** – Kind of interpolation if *method* is 'interpolation'.
- **metadata** – Output metadata.

Returns A new xcube dataset resampled in time.

xcube.api.**vars_to_dim**(*cube: xarray.core.dataset.Dataset, dim_name: str = 'var', var_name='data', cube_asserted: bool = False*)

Convert data variables into a dimension.

Parameters

- **cube** – The xcube dataset.
- **dim_name** – The name of the new dimension and coordinate variable. Defaults to 'var'.
- **var_name** – The name of the new, single data variable. Defaults to 'data'.
- **cube_asserted** – If False, *cube* will be verified, otherwise it is expected to be a valid cube.

Returns A new xcube dataset with data variables turned into a new dimension.

`xcube.api.`**`chunk_dataset`**(*dataset: xarray.core.dataset.Dataset*, *chunk_sizes: Dict[str, int] = None*, *format_name: str = None*) → xarray.core.dataset.Dataset

Chunk dataset and update encodings for given format.

Parameters

- **`dataset`** – input dataset
- **`chunk_sizes`** – mapping from dimension name to new chunk size
- **`format_name`** – format, e.g. "zarr" or "netcdf4"

Returns the re-chunked dataset

`xcube.api.`**`unchunk_dataset`**(*dataset_path: str*, *var_names: Sequence[str] = None*, *coords_only: bool = False*)

Unchunk dataset variables in-place.

Parameters

- **`dataset_path`** – Path to ZARR dataset directory.
- **`var_names`** – Optional list of variable names.
- **`coords_only`** – Un-chunk coordinate variables only.

`xcube.api.`**`vars_to_dim`**(*cube: xarray.core.dataset.Dataset*, *dim_name: str = 'var'*, *var_name='data'*, *cube_asserted: bool = False*)

Convert data variables into a dimension.

Parameters

- **`cube`** – The xcube dataset.
- **`dim_name`** – The name of the new dimension and coordinate variable. Defaults to 'var'.
- **`var_name`** – The name of the new, single data variable. Defaults to 'data'.
- **`cube_asserted`** – If False, *cube* will be verified, otherwise it is expected to be a valid cube.

Returns A new xcube dataset with data variables turned into a new dimension.

5.5 Cube subsetting

`xcube.api.`**`select_vars`**(*dataset: xarray.core.dataset.Dataset*, *var_names: Collection[str] = None*) → xarray.core.dataset.Dataset

Select data variable from given *dataset* and create new dataset.

Parameters

- **`dataset`** – The dataset from which to select variables.
- **`var_names`** – The names of data variables to select.

Returns A new dataset. It is empty, if *var_names* is empty. It is *dataset*, if *var_names* is None.

`xcube.api.`**`clip_dataset_by_geometry`**(*dataset: xarray.core.dataset.Dataset*, *geometry: Union[shapely.geometry.base.BaseGeometry, Dict[str, Any], str, Sequence[Union[float, int]]]*, *save_geometry_wkt: Union[str, bool] = False*) → Optional[xarray.core.dataset.Dataset]

Spatially clip a dataset according to the bounding box of a given geometry.

Parameters

- **dataset** – The dataset
- **geometry** – A geometry-like object, see py:function:*convert_geometry*.
- **save_geometry_wkt** – If the value is a string, the effective intersection geometry is stored as a Geometry WKT string in the global attribute named by *save_geometry*. If the value is True, the name "geometry_wkt" is used.

Returns The dataset spatial subset, or None if the bounding box of the dataset has a no or a zero area intersection with the bounding box of the geometry.

5.6 Cube masking

`xcube.api.`**`mask_dataset_by_geometry`**(*dataset: xarray.core.dataset.Dataset, geometry: Union[shapely.geometry.base.BaseGeometry, Dict[str, Any], str, Sequence[Union[float, int]]], excluded_vars: Sequence[str] = None, no_clip: bool = False, save_geometry_mask: Union[str, bool] = False, save_geometry_wkt: Union[str, bool] = False*) → Optional[xarray.core.dataset.Dataset]

Mask a dataset according to the given geometry. The cells of variables of the returned dataset will have NaN-values where their spatial coordinates are not intersecting the given geometry.

Parameters

- **dataset** – The dataset
- **geometry** – A geometry-like object, see py:function:*convert_geometry*.
- **excluded_vars** – Optional sequence of names of data variables that should not be masked (but still may be clipped).
- **no_clip** – If True, the function will not clip the dataset before masking, this is, the returned dataset will have the same dimension size as the given *dataset*.
- **save_geometry_mask** – If the value is a string, the effective geometry mask array is stored as a 2D data variable named by *save_geometry_mask*. If the value is True, the name "geometry_mask" is used.
- **save_geometry_wkt** – If the value is a string, the effective intersection geometry is stored as a Geometry WKT string in the global attribute named by *save_geometry*. If the value is True, the name "geometry_wkt" is used.

Returns The dataset spatial subset, or None if the bounding box of the dataset has a no or a zero area intersection with the bounding box of the geometry.

class `xcube.api.`**`MaskSet`**(*flag_var: xarray.core.dataarray.DataArray*)

A set of mask variables derived from a variable *flag_var* with CF attributes "flag_masks" and "flag_meanings".

Each mask is represented by an *xarray.DataArray* and has the name of the flag, is of type *numpy.unit8*, and has the dimensions of the given *flag_var*.

Parameters **flag_var** – an *xarray.DataArray* that defines flag values. The CF attributes "flag_masks" and "flag_meanings" are expected to exists and be valid.

5.7 Cube optimization

xcube.api.**optimize_dataset**(*input_path: str*, *output_path: str = None*, *in_place: bool = False*, *unchunk_coords: bool = False*, *exception_type: Type[Exception] = <class 'ValueError'>*)

Optimize a dataset for faster access.

Reduces the number of metadata and coordinate data files in xcube dataset given by given by *dataset_path*. Consolidated cubes open much faster from remote locations, e.g. in object storage, because obviously much less HTTP requests are required to fetch initial cube meta information. That is, it merges all metadata files into a single top-level JSON file ".zmetadata". If *unchunk_coords* is set, it removes any chunking of coordinate variables so they comprise a single binary data file instead of one file per data chunk. The primary usage of this function is to optimize data cubes for cloud object storage. The function currently works only for data cubes using ZARR format.

Parameters

- **input_path** – Path to input dataset with ZARR format.
- **output_path** – Path to output dataset with ZARR format. May contain "{input}" template string, which is replaced by the input path's file name without file name extentsion.
- **in_place** – Whether to modify the dataset in place. If False, a copy is made and *output_path* must be given.
- **unchunk_coords** – Whether to also consolidate coordinate chunk files.
- **exception_type** – Type of exception to be used on value errors.

5.8 Cube metadata

xcube.api.**update_dataset_attrs**(*dataset: xarray.core.dataset.Dataset*, *global_attrs: Dict[str, Any] = None*, *update_existing: bool = False*, *in_place: bool = False*) → xarray.core.dataset.Dataset

Update spatio-temporal CF/THREDDS attributes given *dataset* according to spatio-temporal coordinate variables time, lat, and lon.

Parameters

- **dataset** – The dataset.
- **global_attrs** – Optional global attributes.
- **update_existing** – If `True`, any existing attributes will be updated.
- **in_place** – If `True`, *dataset* will be modified in place and returned.

Returns A new dataset, if *in_place* if `False` (default), else the passed and modified *dataset*.

xcube.api.**update_dataset_spatial_attrs**(*dataset: xarray.core.dataset.Dataset*, *update_existing: bool = False*, *in_place: bool = False*) → xarray.core.dataset.Dataset

Update spatial CF/THREDDS attributes of given *dataset*.

Parameters

- **dataset** – The dataset.
- **update_existing** – If `True`, any existing attributes will be updated.
- **in_place** – If `True`, *dataset* will be modified in place and returned.

Returns A new dataset, if *in_place* if `False` (default), else the passed and modified *dataset*.

`xcube.api.`**`update_dataset_temporal_attrs`**(*dataset: xarray.core.dataset.Dataset*, *update_existing: bool = False*, *in_place: bool = False*) → xarray.core.dataset.Dataset

Update temporal CF/THREDDS attributes of given *dataset*.

Parameters

- **`dataset`** – The dataset.
- **`update_existing`** – If `True`, any existing attributes will be updated.
- **`in_place`** – If `True`, *dataset* will be modified in place and returned.

Returns A new dataset, if *in_place* is `False` (default), else the passed and modified *dataset*.

5.9 Cube verification

`xcube.api.`**`assert_cube`**(*dataset: xarray.core.dataset.Dataset*, *name=None*) → xarray.core.dataset.Dataset

Assert that the given *dataset* is a valid xcube dataset.

Parameters

- **`dataset`** – The dataset to be validated.
- **`name`** – Optional parameter name.

Raise ValueError, if dataset is not a valid xcube dataset

`xcube.api.`**`verify_cube`**(*dataset: xarray.core.dataset.Dataset*) → List[str]

Verify the given *dataset* for being a valid xcube dataset.

The tool verifies that *dataset* * defines the dimensions "time", "lat", "lon"; * has corresponding "time", "lat", "lon" coordinate variables and that they

are valid, e.g. 1-D, non-empty, using correct units;

- has valid bounds variables for "time", "lat", "lon" coordinate variables, if any;
- has any data variables and that they are valid, e.g. min. 3-D, all have same dimensions, have at least dimensions "time", "lat", "lon".

Returns a list of issues, which is empty if *dataset* is a valid xcube dataset.

Parameters **`dataset`** – A dataset to be verified.

Returns List of issues or empty list.

5.10 Multi-resolution pyramids

`xcube.api.compute_levels`(*dataset: xarray.core.dataset.Dataset, spatial_dims: Tuple[str, str] = None, spatial_shape: Tuple[int, int] = None, spatial_tile_shape: Tuple[int, int] = None, var_names: Sequence[str] = None, num_levels_max: int = None, post_process_level: Callable[[xarray.core.dataset.Dataset, int, int], Optional[xarray.core.dataset.Dataset]] = None, progress_monitor: Callable[[xarray.core.dataset.Dataset, int, int], Optional[xarray.core.dataset.Dataset]] = None*) → List[xarray.core.dataset.Dataset]

Transform the given *dataset* into the levels of a multi-level pyramid with spatial resolution decreasing by a factor of two in both spatial dimensions.

It is assumed that the spatial dimensions of each variable are the inner-most, that is, the last two elements of a variable's shape provide the spatial dimension sizes.

Parameters

- **dataset** – The input dataset to be turned into a multi-level pyramid.
- **spatial_dims** – If given, only variables are considered whose last to dimension elements match the given *spatial_dims*.
- **spatial_shape** – If given, only variables are considered whose last to shape elements match the given *spatial_shape*.
- **spatial_tile_shape** – If given, chunking will match the provided *spatial_tile_shape*.
- **var_names** – Variables to consider. If None, all variables with at least two dimensions are considered.
- **num_levels_max** – If given, the maximum number of pyramid levels.
- **post_process_level** – If given, the function will be called for each level and must return a dataset.
- **progress_monitor** – If given, the function will be called for each level.

Returns A list of dataset instances representing the multi-level pyramid.

`xcube.api.read_levels`(*dir_path: str, progress_monitor: Callable[[xarray.core.dataset.Dataset, int, int], Optional[xarray.core.dataset.Dataset]] = None*) → List[xarray.core.dataset.Dataset]

Read the of a multi-level pyramid with spatial resolution decreasing by a factor of two in both spatial dimensions.

Parameters

- **dir_path** – The directory path.
- **progress_monitor** – An optional progress monitor.

Returns A list of dataset instances representing the multi-level pyramid.

`xcube.api.write_levels`(*output_path: str, dataset: xarray.core.dataset.Dataset = None, input_path: str = None, link_input: bool = False, progress_monitor: Callable[[xarray.core.dataset.Dataset, int, int], Optional[xarray.core.dataset.Dataset]] = None, **kwargs*) → List[xarray.core.dataset.Dataset]

Transform the given dataset given by a *dataset* instance or *input_path* string into the levels of a multi-level pyramid with spatial resolution decreasing by a factor of two in both spatial dimensions and write them to *output_path*.

One of *dataset* and *input_path* must be given.

Parameters

- **output_path** – Output path
- **dataset** – Dataset to be converted and written as levels.
- **input_path** – Input path to a dataset to be transformed and written as levels.
- **link_input** – Just link the dataset at level zero instead of writing it.
- **progress_monitor** – An optional progress monitor.
- **kwargs** – Keyword-arguments accepted by the `compute_levels()` function.

Returns A list of dataset instances representing the multi-level pyramid.

5.11 Utilities

`xcube.api.`**`convert_geometry`**(*geometry: Union[shapely.geometry.base.BaseGeometry, Dict[str, Any], str, Sequence[Union[float, int]], None]*) → Optional[shapely.geometry.base.BaseGeometry]

Convert a geometry-like object into a shapely geometry object (`shapely.geometry.BaseGeometry`).

A geometry-like object is may be any shapely geometry object, * a dictionary that can be serialized to valid GeoJSON, * a WKT string, * a box given by a string of the form "<x1>,<y1>,<x2>,<y2>"

or by a sequence of four numbers x1, y1, x2, y2,

- a point by a string of the form "<x>,<y>" or by a sequence of two numbers x, y.

Handling of geometries crossing the antimeridian:

- If box coordinates are given, it is allowed to pass x1, x2 where x1 > x2, which is interpreted as a box crossing the antimeridian. In this case the function splits the box along the antimeridian and returns a multi-polygon.
- In all other cases, 2D geometries are assumed to _not cross the antimeridian at **all_**.

Parameters **geometry** – A geometry-like object

Returns Shapely geometry object or None.

CHAPTER

SIX

WEB API AND SERVER

xcube's RESTful web API is used to publish data cubes to clients. Using the API, clients can

- List configured xcube datasets;
- Get xcube dataset details including metadata, coordinate data, and metadata about all included variables;
- Get cube data;
- Extract time-series statistics from any variable given any geometry;
- Get spatial image tiles from any variable;
- Get places (GeoJSON features including vector data) that can be associated with xcube datasets.

Later versions of API will also allow for xcube dataset management including generation, modification, and deletion of xcube datasets.

The complete description of all available functions is provided in the in the xcube Web API reference.

The web API is provided through the *xcube server* which is started using the *xcube serve* CLI command.

CHAPTER

SEVEN

VIEWER APP

The xcube viewer app is a simple, single-page web application to be used with the xcube server.

7.1 Demo

To test the viewer app, you can use the xcube viewer demo. When you open the page a message “cannot reach server” will appear. This is normal as the demo is configured to run with an xcube server started locally on default port 8080, see *Web API and Server*. Hence, you can either run an xcube server instance locally then reload the viewer page, or configure the viewer with an an existing xcube server. To do so open the viewer’s settings panels, select “Server”. A “Select Server” panel is opened, click the “+” button to add a new server. Here are two demo servers that you may add for testing:

- DCS4COP Demo Server (`https://xcube2.dcs4cop.eu/dcs4cop-dev/api/0.1.0.dev6/`) providing ocean color variables in the North Sea area for the Data Cube Service for Copernicus (DCS4COP) EU project;
- ESDL Server (`https://xcube.earthsystemdatalab.net`) providing global essential climate variables (ECVs) variables for the ESA Earth System Data Lab.

7.2 Functionality

Coming soon…

7.3 Build and Deploy

You can also build and deploy your own viewer instance. In the latter case, visit the xcube-viewer repository on GitHub and follow the instructions provides in the related README file.

CHAPTER

EIGHT

XCUBE DATASET SPECIFICATION

This document provides a technical specification of the protocol and format for *xcube datasets*, data cubes in the xcube sense.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119.

8.1 Document Status

This is the latest version, which is still in development.

Version: 1.0, draft

Updated: 31.05.2018

8.2 Motivation

For many users of Earth observation data, multivariate coregistration, extraction, comparison, and analysis of different data sources is difficult, while data is provided in various formats and at different spatio-temporal resolutions.

8.3 High-level requirements

xcube datasets

- SHALL be time series of gridded, geo-spatial, geo-physical variables.
- SHALL use a common, equidistant, global or regional geo-spatial grid.
- SHALL shall be easy to read, write, process, generate.
- SHALL conform to the requirements of analysis ready data (ARD).
- SHALL be compatible with existing tools and APIs.
- SHALL conform to standards or common practices and follow a common data model.
- SHALL be formatted as self-contained datasets.
- SHALL be "cloud ready", in the sense that subsets of the data can be accessed by individual URIs.

ARD links:

- http://ceos.org/ard/

- https://landsat.usgs.gov/ard
- https://medium.com/planet-stories/analysis-ready-data-defined-5694f6f48815

8.4 xcube Dataset Schemas

8.4.1 Basic Schema

- Attributes metadata convention
  - SHALL be CF >= 1.7
  - SHOULD adhere to Attribute Convention for Data Discovery
- Dimensions:
  - SHALL be at least `time`, `bnds`, and MAY be any others.
  - SHALL all be greater than zero, but `bnds` must always be two.
- Temporal coordinate variables:
  - SHALL provide time coordinates for given time index.
  - MAY be non-equidistant or equidistant.
  - `time[time]` SHALL provide observation or average time of *cell centers*.
  - `time_bnds[time, bnds]` SHALL provide observation or integration time of *cell boundaries*.
  - Attributes:
    * Temporal coordinate variables MUST have `units`, `standard_name`, and any others.
    * `standard_name` MUST be `"time"`, `units` MUST have format `"<deltatime> since <datetime>"`, where `datetime` must have ISO-format. `calendar` may be given, if not, `"gregorian"` is assumed.
- Spatial coordinate variables
  - SHALL provide spatial coordinates for given spatial index.
  - SHALL be equidistant in either angular or metric units
- Cube variables:
  - SHALL provide *cube cells* with the dimensions as index.
  - SHALL have shape
    * `[time, ..., lat, lon]` (see WGS84 schema) or
    * `[time, ..., y, x]` (see Generic schema)
  - MAY have extra dimensions, e.g. `layer` (of the atmosphere), `band` (of a spectrum).
  - SHALL specify the `units` metadata attribute.
  - SHOULD specify metadata attributes that are used to identify missing values, namely `_FillValue` and / or `valid_min`, `valid_max`, see notes in CF conventions on these attributes.
  - MAY specify metadata attributes that can be used to visualise the data:
    * `color_bar_name`: Name of a predefined colour mapping. The colour bar is applied between a minimum and a maximum value.

    * `color_value_min`, `color_value_max`: Minimum and maximum value for applying the colour bar. If not provided, minimum and maximum default to `valid_min`, `valid_max`. If neither are provided, minimum and maximum default to `0` and `1`.

8.4.2 WGS84 Schema (extends Basic)

- Dimensions:
  - SHALL be at least `time`, `lat`, `lon`, `bnds`, and MAY be any others.
- Spatial coordinate variables:
  - SHALL use WGS84 (EPSG:4326) CRS.
  - SHALL have `lat[lat]` that provides observation or average latitude of *cell centers* with attributes: `standard_name="latitude"` `units="degrees_north"`.
  - SHALL have `lon[lon]` that provides observation or average longitude of *cell centers* with attributes: `standard_name="longitude"` and `units="degrees_east"`.
  - SHOULD HAVE `lat_bnds[lat, bnds]`, `lon_bnds[lon, bnds]`: provide geodetic observation or integration coordinates of *cell boundaries*.
- Cube variables:
  - SHALL have shape `[time, ..., lat, lon]`.

8.4.3 Generic Schema (extends Basic)

- Dimensions: `time`, `y`, `x`, `bnds`, and any others.
  - SHALL be at least `time`, `y`, `x`, `bnds`, and MAY be any others.
- Spatial coordinate variables:
  - Any spatial grid and CRS.
  - `y[y]`, `x[x]`: provide spatial observation or average coordinates of *cell centers*.
    * Attributes: `standard_name`, `units`, other units describe the CRS / projections, see CF.
  - `y_bnds[y, bnds]`, `x_bnds[x, bnds]`: provide spatial observation or integration coordinates of *cell boundaries*.
  - MAY have `lat[y,x]`: latitude of *cell centers*.
    * Attributes: `standard_name="latitude"`, `units="degrees_north"`.
  - `lon[y,x]`: longitude of *cell centers*.
    * Attributes: `standard_name="longitude"`, `units="degrees_east"`.
- Cube variables:
  - MUST have shape `[time, ..., y, x]`.

8.5 xcube EO Processing Levels

This section provides an attempt to characterize xcube datasets generated from Earth Observation (EO) data according to their processing levels as they are commonly used in EO data processing.

8.5.1 Level-1C and Level-2C

- Generated from Level-1A, -1B, -2A, -2B EO data.
- Spatially resampled to common grid
  - Typically resampled at original resolution.
  - May be down-sampled: aggregation/integration.
  - May be upsampled: interpolation.
- No temporal aggregation/integration.
- Temporally non-equidistant.

8.5.2 Level-3

- Generated from Level-2C or -3 by temporal aggregation.
- No spatial processing.
- Temporally equidistant.
- Temporally integrated/aggregated.

CHAPTER NINE

XCUBE DEVELOPER GUIDE

Version 0.1, draft

IMPORTANT NOTE: Any changes to this doc must be reviewed by dev-team through pull requests.

9.1 Preface

> *Gedacht ist nicht gesagt.Gesagt ist nicht gehört.Gehört ist nicht verstanden.Verstanden ist nicht einverstanden.Einverstanden ist nicht umgesetzt.Umgesetzt ist nicht beibehalten.*

by Konrad Lorenz (translation is left to the reader)

9.2 Table of Contents

9.3 Versioning

We adhere to PEP-440.

The current software version is in `xcube/version.py`.

While developing a version, we append version suffix `.dev<N>`. Before the release, we remove the suffix.

9.4 Coding Style

We try adhering to PEP-8.

9.5 Main Packages

- `xcube.cli` - Here live CLI commands that are required by someone. CLI command implementations should be lightweight. Move implementation code either into `api` or `util`.CLI commands must be maintained w.r.t. backward compatibility. Therefore think twice before adding new or change existing CLI commands.
- `xcube.api` - Here live API functions that are required by someone or that exists because a CLI command is implemented here. API code must be maintained w.r.t. backward compatibility. Therefore think twice before adding new or change existing API.
- `xcube.webapi` - Here live Web API functions that are required by someone. Web API command implementations should be lightweight. Move implementation code either into `api` or `util`.Web API interface must be maintained w.r.t. backward compatibility. Therefore think twice before adding new or change existing API.
- `xcube.util` - Mainly implementation helpers. Comprises classes and functions that are used by `cli`, `api`, `webapi` in order to maximize modularisation and testability but to minimize code duplication.The code in here must not be dependent on any of `cli`, `api`, `webapi`. The code in here may change often and in any way as desired by code implementing the `cli`, `api`, `webapi` packages.

The following sections will guide you through extending or changing the main packages that form xcube's public interface.

9.5.1 Package `xcube.cli`

Checklist

Make sure your change

1. is covered by unit-tests (package `test/api`);
2. is reflected by the CLI's doc-strings and tools documentation (currently in `README.md`);
3. follows existing xcube CLI conventions;
4. follows PEP8 conventions;
5. is reflected in API and WebAPI, if desired;
6. is reflected in `CHANGES.md`.

Hints

Make sure your new CLI command is in line with the others commands regarding command name, option names, as well as metavar arguments names. The CLI command name shall ideally be a verb.

Avoid introducing new option arguments if similar options are already in use for existing commands.

In the following common arguments and options are listed.

Input argument:

```
@click.argument('input')
```

If input argument is restricted to an xcube dataset:

```
@click.argument('cube')
```

Output (directory) option:

```
@click.option('--output', '-o', metavar='OUTPUT',
              help='Output directory. If omitted, "INPUT.levels" will be used.')
```

Output format:

```
@click.option('--format', '-f', metavar='FORMAT', type=click.Choice(['zarr', 'netcdf
↪']),
              help="Format of the output. If not given, guessed from OUTPUT.")
```

Output parameters:

```
@click.option('--param', '-p', metavar='PARAM', multiple=True,
              help="Parameter specific for the output format. Multiple allowed.")
```

Variable names:

```
@click.option('--variable',--var', metavar='VARIABLE', multiple=True,
              help="Name of a variable. Multiple allowed.")
```

For parsing CLI inputs, use helper functions that are already in use. In the CLI command implementation code, raise `click.ClickException(message)` with a clear `message` for users.

Common xcube CLI options like `-f` for FORMAT should be lower case letters and specific xcube CLI options like `-S` for SIZE in `xcube gen` are recommended to be uppercase letters.

Extensively validate CLI inputs to avoid that API functions raise `ValueError`, `TypeError`, etc. Such errors and their message texts are usually hard to understand by users. They are actually dedicated to to developers, not CLI users.

There is a global option `--traceback` flag that user can set to dump stack traces. You don't need to print stack traces from your code.

9.5.2 Package `xcube.api`

Checklist

Make sure your change

1. is covered by unit-tests (package `test/api`);
2. is covered by API documentation;
3. follows existing xcube API conventions;
4. follows PEP8 conventions;
5. is reflected in xarray extension class `xcube.api.api.API`;
6. is reflected in CLI and WebAPI if desired;
7. is reflected in `CHANGES.md`.

Hints

Create new module in `xcube.api` and add your functions. For any functions added make sure naming is in line with other API. Add clear doc-string to the new API. Use Sphinx RST format.

Decide if your API methods requires *xcube datasets* as inputs, if so, name the primary dataset argument `cube` and add a keyword parameter `cube_asserted: bool = False`. Otherwise name the primary dataset argument `dataset`.

Reflect the fact, that a certain API method or function operates only on datasets that conform with the xcube dataset specifications by using `cube` in its name rather than `dataset`. For example `compute_dataset` can operate on any xarray datasets, while `get_cube_values_for_points` expects a xcube dataset as input or `read_cube` ensures it will return valid xcube datasets only.

In the implementation, if `not cube_asserted`, we must assert and verify the `cube` is a cube. Pass `True` to `cube_asserted` argument of other API called later on:

```
from .verify import assert_cube

def frombosify_cube(cube: xr.Dataset, ..., cube_asserted: bool = False):
    if not cube_asserted:
        assert_cube(cube)
    ...
    result = bibosify_cube(cube, ..., cube_asserted=True)
    ...
```

If `import xcube.api` is used in client code, any `xarray.Dataset` object will have an extra property `xcube` whose interface is defined by the class `xcube.api.XCubeAPI`. This class is an xarray extension that is used to reflect `xcube.api` functions and make it directly applicable to the `xarray.Dataset` object.

Therefore any xcube API shall be reflected in this extension class.

9.5.3 Package `xcube.webapi`

Checklist

Make sure your change

1. is covered by unit-tests (package `test/webapi`);
2. is covered by Web API specification and documentation (currently in `webapi/res/openapi.yml`);
3. follows existing xcube Web API conventions;
4. follows PEP8 conventions;
5. is reflected in CLI and API, if desired;
6. is reflected in `CHANGES.md`.

9.5.4 Hints

- The Web API is defined in `webapi.app` which defines mapping from resource URLs to handlers
- All handlers are implemented in `webapi.handlers`. Handler code just delegates to dedicated controllers.
- All controllers are implemented in `webapi.controllers.*`. They might further delegate into `api.*`

9.6 Development Process

1. Make sure there is an issue ticket for your code change work item

2. Select issue, priorities are as follows
    1. "urgent" and ("important" and "bug")
    2. "urgent" and ("important" or "bug")
    3. "urgent"
    4. "important" and "bug"
    5. "important" or "bug"
    6. others
3. Make sure issue is assigned to you, if unclear agree with team first.
4. Add issue label "in progress".
5. Create development branch named "developer-issue#-title".
6. Develop, having in mind the checklists and implementation hints above.
    1. In your first commit, refer the issue so it will appear as link in the issue history
    2. Develop, test, and push to the remote branch as desired.
    3. In your last commit, utilize checklists above. (You can include the line "closes #" in your commit message to auto-close the issue once the PR is merged.)
7. Create PR if build servers succeed on your branch. If not, fix issue first.For the PR assign the team for review, agree who is to merge. Also reviewers must have checklist in mind!
8. Merge PR after all reviewers are accepted your change. Otherwise go back.
9. Remove issue label "in progress".
10. Delete the development branch "developer-issue#-title".
11. If the PR is only partly solving an issue:
    1. Make sure the issue contains a to-do list (checkboxes) to complete the issue.
    2. Do not include the line "closes #" in your last commit message.
    3. Add "relates to issue#" in PR.
    4. Make sure to check the corresponding to-do items (checkboxes) *after* the PR is merged.
    5. Remove issue label "in progress".
    6. Leave issue open.

CHAPTER TEN

INDICES AND TABLES

- genindex
- modindex
- search

INDEX