load_time_histmth

May 14, 2021

1 Investigating Iris load time for loading a single variable form a NetCDF multi-variable file

1.1 We mean 300 variables

```python
[15]: import iris
      import time
      import cProfile,io,pstats
      import os
```

```python
[17]: from IPython.core.display import display, HTML
      display(HTML("<style>.container { width:100% !important; }</style>"))
```

```
<IPython.core.display.HTML object>
```

```python
[2]: ddir  = "../histmth_data"

     # a file of size 260 Mo and with ~ 300 variables covering a period of one month
     # Format is NetCDF4 classic
     dhist = ddir+"/CM61-LR-hist-03.1950_20000101_20000131_1M_histmth.nc"

     # a single variable file of size 100 Mo and covering a period of 164 years
     dts   = ddir+"/CM61-LR-hist-03.1950_18500101_20141231_1M_t2m.nc"
```

1.2 A helper function for timing

```python
[3]: def testit(file,func,constraints) :
         # Call FUNC on FILE with arg CONSTRAINS and print execution time
         tim1 = time.time()
         lc=func(file,constraints)
         duration = time.time() - tim1
         print(file.split('_')[-1], func.__name__, "%.2g s"%duration, constraints)
```

1.3 Load time is quite small for a single variable file, whatever the load function version

```
[6]: testit(dts,func=iris.load,constraints=None)
```

t2m.nc load 0.083 s None

/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/cf.py:374: UserWarning: Missing CF-netCDF boundary
variable 'time_counter_bounds', referenced by netCDF variable 'time_counter'
  warnings.warn(message % (name, nc_var_name))

```
[7]: testit(dts,func=iris.load,constraints=iris.NameConstraint(var_name='t2m'))
```

t2m.nc load 0.083 s NameConstraint(var_name='t2m')

/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/cf.py:374: UserWarning: Missing CF-netCDF boundary
variable 'time_counter_bounds', referenced by netCDF variable 'time_counter'
  warnings.warn(message % (name, nc_var_name))

```
[8]: testit(dts,func=iris.load_raw,constraints=iris.NameConstraint(var_name='t2m'))
```

t2m.nc load_raw 0.079 s NameConstraint(var_name='t2m')

/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/cf.py:374: UserWarning: Missing CF-netCDF boundary
variable 'time_counter_bounds', referenced by netCDF variable 'time_counter'
  warnings.warn(message % (name, nc_var_name))

1.4 Load time is very large on the multi-variable file : ~ 100s

```
[9]: testit(dhist,func=iris.load,constraints=iris.NameConstraint(var_name='t2m'))
```

/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'epmax' invalid units 'su'
  warnings.warn(msg)
/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'sza' invalid units 'deg'
  warnings.warn(msg)
/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'R_incl' invalid units 'deg'
  warnings.warn(msg)

histmth.nc load 99 s NameConstraint(var_name='t2m')

```
[10]: testit(dhist,func=iris.load_raw, constraints=iris.
      ↪NameConstraint(var_name='t2m'))
```

/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'epmax' invalid units 'su'
  warnings.warn(msg)
/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'sza' invalid units 'deg'
  warnings.warn(msg)
/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'R_incl' invalid units 'deg'
  warnings.warn(msg)

histmth.nc load_raw 99 s NameConstraint(var_name='t2m')

```
[11]: testit(dhist,func=iris.load_cube, constraints=iris.
      ↪NameConstraint(var_name='t2m'))
```

/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'epmax' invalid units 'su'
  warnings.warn(msg)
/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'sza' invalid units 'deg'
  warnings.warn(msg)
/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'R_incl' invalid units 'deg'
  warnings.warn(msg)

histmth.nc load_cube 99 s NameConstraint(var_name='t2m')

1.5 Another tool takes much less time, even when including writing data to disk

```
[31]: tim1 = time.time()
      os.system("cdo -selvar,t2m "+dhist+" tmp.nc")
      iris.load("tmp.nc")
      duration = time.time() - tim1
      print( "same with cdo", "%.2g s"%duration)
```

same with cdo 0.71 s

1.6 Profiling the execution shows that NetCDFDataProxy.___getitem___ is the main time consuming function

```
[29]: pr = cProfile.Profile()
      pr.enable()
      iris.load(dhist,constraints=iris.NameConstraint(var_name='t2m'))
      s = io.StringIO()
      ps = pstats.Stats(pr, stream=s).sort_stats('cumulative')
      ps.print_stats()
      # Just un-comment next line to get the profile on stdout
      a=s.getvalue()
      for line in a.splitlines()[0:25] :
          print(line)
```

```
/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'epmax' invalid units 'su'
  warnings.warn(msg)
/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'sza' invalid units 'deg'
  warnings.warn(msg)
/data/ssenesi/conda/esmvaltool-dev/lib/python3.9/site-
packages/iris/fileformats/_pyke_rules/compiled_krb/fc_rules_cf_fc.py:2206:
UserWarning: Ignoring netCDF variable 'R_incl' invalid units 'deg'
  warnings.warn(msg)
        4771193 function calls (4757624 primitive calls) in 103.197 seconds

   Ordered by: cumulative time

   ncalls  tottime  percall  cumtime  percall filename:lineno(function)
        3    0.000    0.000  103.202   34.401 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/IPython/core/interactiveshell.py:3396(run_code)
      4/3    0.000    0.000  103.202   34.401 {built-in method builtins.exec}
        1    0.000    0.000  103.202  103.202 <ipython-
input-29-3dd963c4780d>:3(<module>)
        1    0.000    0.000  103.202  103.202 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/iris/__init__.py:279(load)
        1    0.000    0.000  103.198  103.198 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/iris/__init__.py:268(_load_collection)
        1    0.008    0.008  103.198  103.198 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/iris/cube.py:107(from_cubes)
      315    0.001    0.000  103.173    0.328 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/iris/__init__.py:245(_generate_cubes)
      315    0.002    0.000  103.173    0.328 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/iris/io/__init__.py:175(load_files)
      315    0.003    0.000  103.140    0.327 /data/ssenesi/conda/esmvaltool-
```

```
dev/lib/python3.9/site-packages/iris/fileformats/netcdf.py:774(load_cubes)
     314    0.014    0.000  102.007     0.325 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/iris/fileformats/netcdf.py:587(_load_cube)
     314    0.005    0.000   95.462     0.304 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/iris/fileformats/netcdf.py:565(_get_cf_var_data)
     314    0.005    0.000   94.997     0.303 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/iris/_lazy_data.py:150(as_lazy_data)
     314    0.010    0.000   94.894     0.302 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/dask/array/core.py:2934(from_array)
     314    0.013    0.000   94.688     0.302 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/dask/array/core.py:1121(__new__)
     628    0.013    0.000   94.600     0.151 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/dask/array/utils.py:32(meta_from_array)
     314   90.717    0.289   94.547     0.301 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/iris/fileformats/netcdf.py:434(__getitem__)
     314    0.001    0.000    6.171     0.020 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/pyke/knowledge_engine.py:291(activate)
     314    0.002    0.000    6.169     0.020 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/pyke/rule_base.py:151(activate)
     314    0.015    0.000    5.986     0.019 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/pyke/rule_base.py:144(run_fc_rules)
   19468    0.044    0.000    5.971     0.000 /data/ssenesi/conda/esmvaltool-
dev/lib/python3.9/site-packages/pyke/fc_rule.py:88(run)
```

[ ]: