Improvement possibilities for future MIP data processing

Luis Kornblueh and Reinhard Budich
Max Planck Institute for Meteorology

MARS for fast HPC data access

Based on a talk by Carsten Maass, ECMWF
Incentive is to adapt the MARS language for accessing MIP model output for HPC based users.

- Base on the **Meteorological Archival and Retrieval System** of ECMWF
- Meteorological data (GRIB and CF-Convention based netCDF4)
- Large amount of data (size of archive and number of fields)
- Research environment
- Batch and interactive use model
- Large number of users with different requirements
- Heterogeneous environment

MARS language - Mechanism to name archived fields

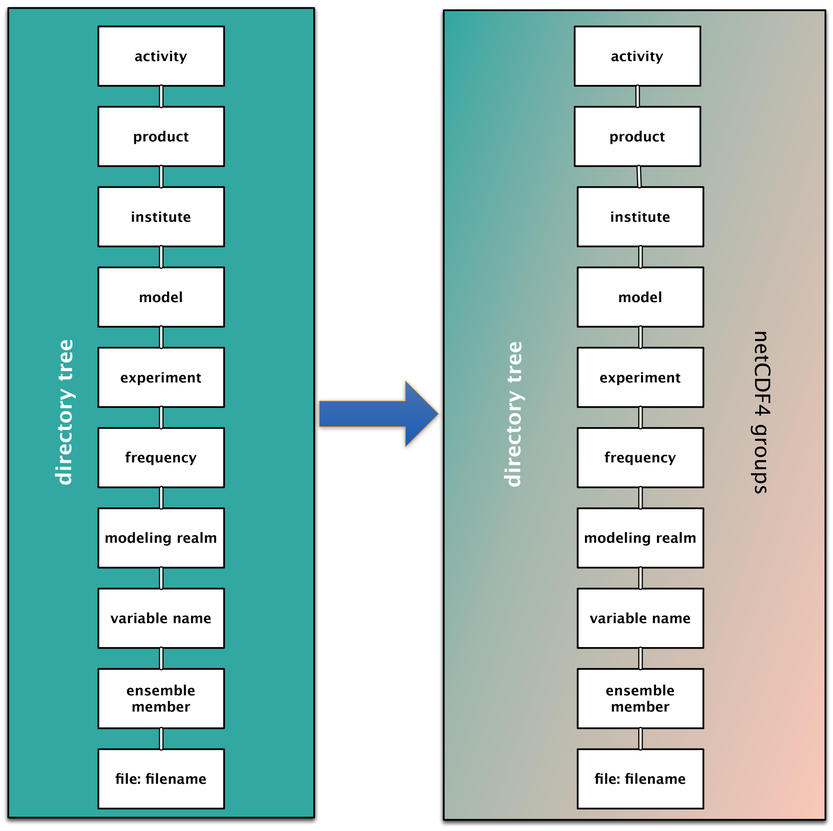
Request syntax:

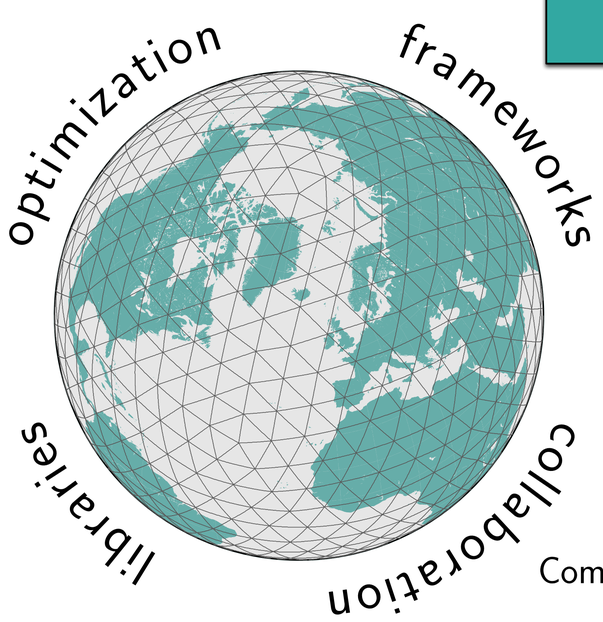
```
verb,
     keyword1 = value1,
     ...   = value2,
     keywordN = valueN
```

verb : action to be taken (e.g. retrieve, list, read)
keyword : a known MARS variable, e.g. type or date
value : value assigned to the keyword, e.g. analysis or temperature

Asynchroneous I/O

Flexibility of CMOR directory structure for HPC filesystem access by moving tree components from directory levels to netCDF4 groups.





Workflow activities

EUDAT Activities:
• Workflow Working Group
t.b.c. in EUDAT2020
• Conceptual Generic Execution Framework GEF
See http://www.eudat.eu/workflows
contact Christian Page christian.page@cerfacs.fr
or Reinhard Budich reinhard.budich@mpimet.mpg.de

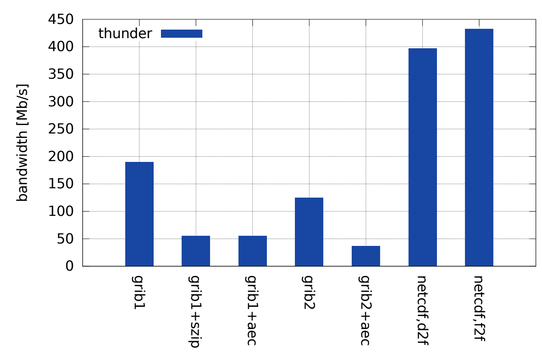
IS-ENES2 Activities:
• Workshops on:
Configuration Management Sep 2013
MD Generation Feb 2014
Workflow Solution initial workshop June 2014
See www.enes.org
contact Reinhard Budich reinhard.budich@mpimet.mpg.de

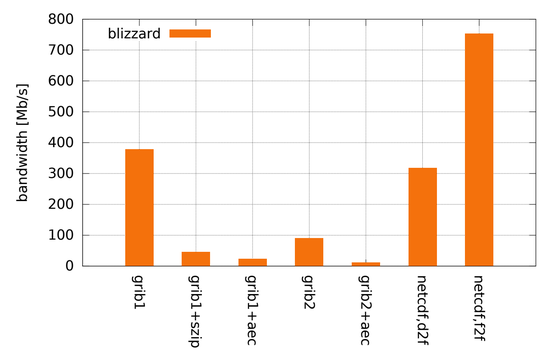
MPI-M Activities

Python based provenance collecting workflow management
contact Luis Kornblueh, luis.kornblueh@mpimet.mpg.de
or Deike Kleberg, deike.kleberg@mpimet.mpg.de

Compression

Average compression rate: 2.45
Compression saves: network bandwidth and disk capacity





- grib1, grib encoding, grib + szip encoding (ICON R03B07 transformed to r3072x960)
- grib1, grib encoding, grib + aec encoding (ICON R03B07 transformed to r3072x960)
- grib2, grib encoding, grib + aec encoding (ICON R03B07)
- netcdf, no encoding, but float and double input comparision (ICON R03B07)

Joint effor with Mathis Rosenhauer (DKRZ), and Shahram Najm (ECMWF)

Max-Planck-Institut für Meteorologie

Contact: luis.kornblueh@mpimet.mpg.de

MAX-PLANCK-GESELLSCHAFT