



*Norwegian  
Meteorological Institute  
met.no*

**IT-infrastructure at met.no  
for EMEP, AEROCOM, HTAP**

**Jan Griesfeller**



# Overview

- Supercomputing facilities
- Norstore (data archive)
- Aeroicom infrastructure
- HTAP infrastructure



# Supercomputing facilities used for EMEP (not exclusively)

- Stallo (for smaller simulations)
  - HP BL460c Gen8, 304 nodes, 4864 cores, CPU: Intel Xeon E5-2670, 101Tflop, 12,8 TB RAM, 2.1 PB storage
  - Based on standard server technology (blade server with 'standard' interconnects between nodes)
  - No official top550 ranking, but at the end of the list comparing the 101 Tflop peak performance
  - Located at University of Tromsø



# Supercomputing facilities used for EMEP (not exclusively)

- Vilje (for big simulations > 127 CPUs)
  - SGI Altix ICE X, 1404 nodes, 22464 cores, 44Tb RAM, Intel Sandy bridge CPUs, 467 Tflop theoretical total peak performance
  - Special interconnect between nodes for higher speed (but based on std interconnection protocols)
  - #56 at Nov2012 Top500 list
  - Located at NTNU, Trondheim
  - Partly owned by met.no



# Supercomputing facilities used for EMEP (not exclusively)

- Hexagon (for climate model runs; NorESM)
  - Cray XE6, 696 nodes, 22272 cores, CPU: AMD Interlagos, 205 TFlops, 22 TB RAM, 540TB storage
  - Very special interconnect between nodes (Cray 2-D torus interconnect)
  - #157 at Nov2012 Top500 list
  - Located at UiB / Uni Research, Bergen

# Norstore



- «NorStore is a national infrastructure for the management, curation and long-term archiving of digital scientific data.»
- Supercomputers do **not** provide space for long term archiving of scientific data
- Provides 2x610TB + space on tapes for researchers in the national high performance computing facilities
- But 'only' up to 60MB/s data rate to supercomputers (~ 5TiB per day)
- Archive of 100TB on disks + 60TB on tape for EMEP

# Some numbers or why do we need those supercomputers?



- Amount of input data per simulated year: ~4TB  
==> data copy from Norstore takes nearly 1 day alone;  
biggest available single harddrive is 4TB, but unusable for servers (too unreliable), server HDDs up to 1TB
- Amount of output data: Depends on what you ask for
- Speed:  
# of simulations per EMEP report to run: ~1000  
(5 meteorological years, 4-5 pollutants, per country; results should be available ASAP)



# AeroCom: What is it?

- The AEROCOM-project is an open international initiative of scientists interested in the advancement of the understanding of the global aerosol and its impact on climate
- We do model-model comparisons and model-observations comparisons
- Observations are both satellite based (e.g. MODIS, MISR, Parasol, etc.) and ground based (e.g. Aeronet, EBAS)

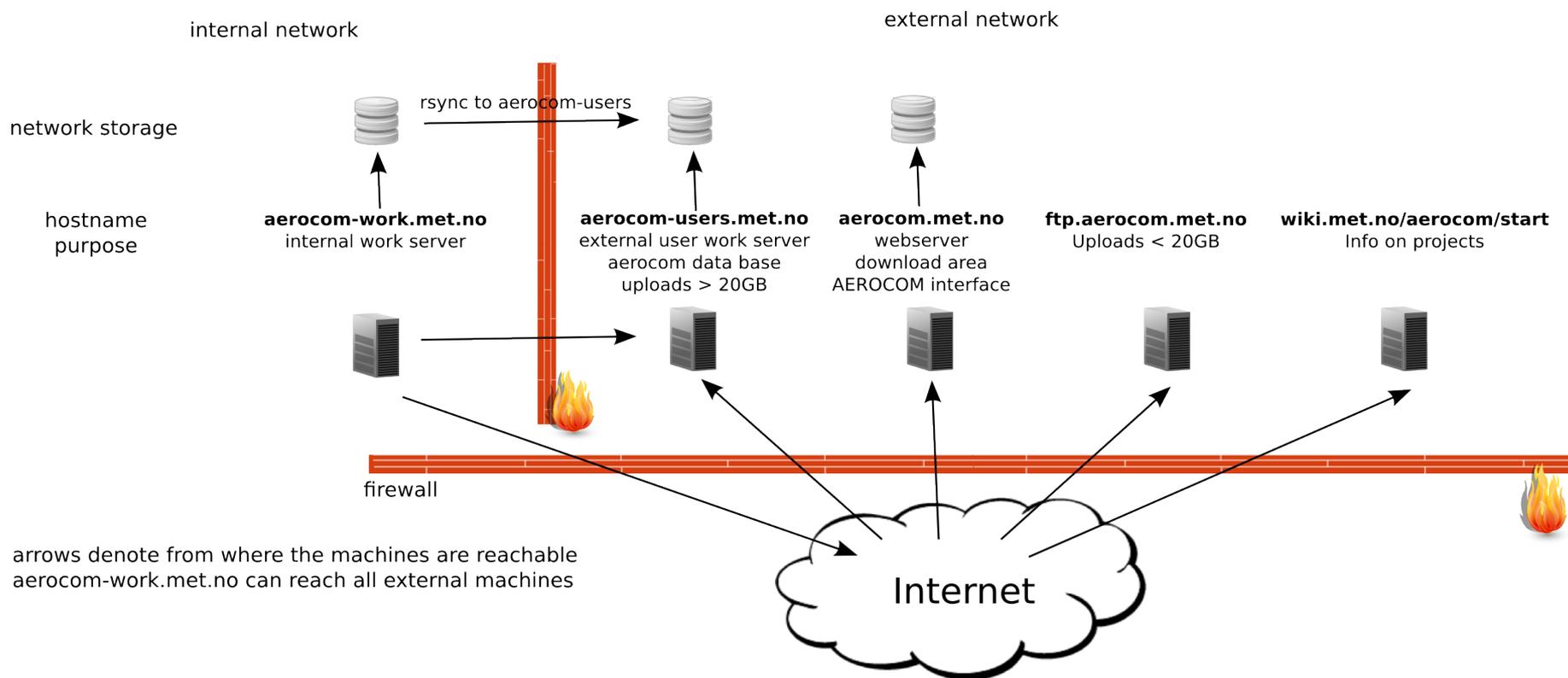


## AeroCom: What is it cont.?

- We also do satellite validation (e.g. CCI-Aerosol)
- Results are also presented via a web interface
- AeroCom data base: Exchange for different model data via a user server where external users can download data or work on data on the server



# Aerocom infrastructure: Hardware



# Description of the aerocom hardware



Aerocom-work Work server	Aerocom-users External users	Aerocom Web server	Ftp.aerocom Ftp server
<ul style="list-style-type: none"> <li>- 8 cores / 16 threads @ 2.93GHz</li> <li>- 48GB RAM</li> <li>- 13TB Storage</li> </ul>	<ul style="list-style-type: none"> <li>- 4 cores / 4 threads @2.4 GHz</li> <li>- 4 GB RAM</li> <li>- 12TB Storage</li> </ul>	<ul style="list-style-type: none"> <li>- virtual server</li> <li>- 2GB RAM</li> <li>- 1TB Storage</li> </ul>	<ul style="list-style-type: none"> <li>- virtual server</li> <li>- 3GB RAM</li> <li>- 100Gb for aerocom</li> <li>- alias for ftp.met.no</li> </ul>
<ul style="list-style-type: none"> <li>- Internal work server</li> </ul>	<ul style="list-style-type: none"> <li>- external work server</li> <li>- Host of the aerocom data base</li> <li>- Can accessed by external users via ssh</li> </ul>	<ul style="list-style-type: none"> <li>- web server</li> <li>- download area</li> </ul>	<ul style="list-style-type: none"> <li>- ftp server to send us data for the aerocom data base and other projects (HTAP, ECLIPSE, CCI, etc.)</li> </ul>



# AeroCom: Software

- Mainly open source and all GNU/Linux based  
nco, ncview, python, perl, apache, MySQL, cdo, ncl, ...
- The only exception is **IDL**.  
Could probably today be replaced by python and matplotlib, but we invested several man years into our tools.



# HTAP and its infrastructure

- The Task Force on Hemispheric Transport of Air Pollution (TF HTAP) is an international scientific cooperative effort to improve the understanding of the intercontinental transport of air pollution across the Northern Hemisphere. <http://www.htap.org>
- Infrastructure for HTAP
  - CF-Checker  
(NetCDF Climate and Forecast (CF) Metadata Convention;  
<http://cf-pcmdi.llnl.gov/>)
  - WCS (web coverage service) server



# CF-Checker: Introduction

- Users can upload their netcdf files via a web browser to check if they are CF compliant as HTAP requires
- Users get a detailed error and warning list to work on their data to make it 'HTAP compliant'
- Users get help of how to solve standard errors
- **Command line version exists as well.**

# CF-Checker: Root web page



**Task Force on Hemispheric  
Transport of Air Pollution**



File upload facility for TFHTAP model data

Help

## File and CF-Version

Select File(s) to Upload

Browse..

Select CF-version to validate

auto

## Test Results

Please select files to upload and test!



# CF-Checker: after check



**Task Force on Hemispheric Transport of Air Pollution**



Meteorologisk institutt  
met.no

File upload facility for TFHTAP model data

Help

**File and CF-Version**

Select File(s) to Upload  Select CF-version to validate

**Test Results**

File Name	File size	Upload Status
aerocom.LSCE-vRV.A2.CTRL.daily.od550aer.2006.nc	13144 kByte	failure

**CF-Convention Test**

- ✘ attribute name "\_Fillvalue" of variable "od550aer" does not follow naming conventions (2.3)
- ✘ od550aer: units "?" not recognized (3.1)
- ✘ od550aer: standard\_name '?' does not exist in standard name table (3.3)
- ✘ time: no or unknown calendar defined, so month\_lengths attribute is required (4.4.1)
- ✘ od550aer: following variables that are referenced in coordinates attribute do not exist: time\_counter (5)
- ✘ time: a coordinate variable must have values that are strictly monotonic (5)
- ℹ running CFchecker version 1.5.11 (INIT)
- ℹ checking compliance with convention CF-1.1 (INIT)
- ...

[click here](#) to list all errors!

This tool is developed and maintained by  JÜLICH



# WCS: Very quick introduction

- WCS server (web coverage service) for 'easy' data access for the HTAP community
- Standard protocol from Open Geospatial Consortium (OGC)
- Link:  
<http://www.opengeospatial.org/standards/wcs>

# WCS: Very quick introduction cont.



- A WCS provides access to coverage data in forms that are useful for client-side rendering, as input into scientific models, and for other clients.
- WCS allows clients to choose portions of a server's information holdings based on spatial constraints and other query criteria (the server is doing the data selection)



# Questions?