Big Data Research at DKRZ

Michael Lautenschlager and Colleagues from DKRZ and Scientific Computing Research Group

Symposium Big Data in Science
Karlsruhe October 7th, 2014
Big Data in Climate Research

- **Big data** is an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications. (Wikipedia)

- **Big data** usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big data "size" is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data. (Snijders, C., Matzat, U., & Reips, U.-D. (2012). ‘Big Data’: Big gaps of knowledge in the field of Internet. *International Journal of Internet Science, 7*, 1-5.)

- Gartner’s "3Vs" model for describing big data: increasing volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources). (Laney, Douglas. "3D Data Management: Controlling Data Volume, Velocity and Variety". Gartner. Retrieved 6 February 2001.)

- Big Data in climate research:
  - Satellite data from Earth observation (2 V)
  - Climate model data (1.5 V)
  - Data from observational networks (e.g. Tsunami and Earth quarks) (3V)
Outline

• The German Climate Computing Center (DKRZ)
• Infrastructure at DKRZ: present and future
• Climate model data characteristic

• DKRZ’s HPC infrastructure developments
• Climate model data production workflow

• Future Developments
The German Climate Computing Center (DKRZ)

Founded in 1987 as a national institution

Operated as a non-profit limited company with four shareholders

- Max Planck Society for Research (55%)
- The City of Hamburg represented by the University of Hamburg (27%)
- Alfred Wegener Research Institute in Bremerhaven (9%)
- Helmholtz Center for Research in Geesthacht (9%)
Mission

DKRZ – Partner for Climate Research

Maximum Compute Performance.
Sophisticated Data Management.
Competent Service.
Climate Modelling Support at DKRZ

Basic Workflows:
- Climate Model Development
- Climate Model Data Production

---

Analysis and Interpretation

Model & Experiment Development

Experiment Setup & Execution

Data Management

Data Sharing

Development

Production & Dissemination
HLRE-2: Architecture

Current machine: IBM Power6
8.500 cores, 158 TFLOPS, 20 TB main memory
6 PB disk space, 100 PB tape archive space

In 2015 replacement by HLRE-3
Increases every ten years:
Processor speed: 500x
Disk capacity: 100x
Disk speed: 11x (17x with SSDs)

### HLRE-2 to HLRE-3

<table>
<thead>
<tr>
<th></th>
<th>2009</th>
<th>2015</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodes</td>
<td>150 TF/s</td>
<td>3 PF/s</td>
<td>20x</td>
</tr>
<tr>
<td></td>
<td>264</td>
<td>2,500</td>
<td>9.5x</td>
</tr>
<tr>
<td><strong>Node performance</strong></td>
<td>0.6 TF/s</td>
<td>1.2 TF/s</td>
<td>2x</td>
</tr>
<tr>
<td><strong>System memory</strong></td>
<td>20 TB</td>
<td>170 TB</td>
<td>8.5x</td>
</tr>
<tr>
<td><strong>Storage capacity</strong></td>
<td>5.6 PB</td>
<td>45 PB</td>
<td>8x</td>
</tr>
<tr>
<td><strong>Storage throughput</strong></td>
<td>30 GB/s</td>
<td>400 GB/s</td>
<td>13.3x</td>
</tr>
<tr>
<td><strong>Disk drives</strong></td>
<td>7,200</td>
<td>8,500</td>
<td>1.2x</td>
</tr>
<tr>
<td><strong>Archive capacity</strong></td>
<td>53 PB</td>
<td>335 PB</td>
<td>6.3x</td>
</tr>
<tr>
<td><strong>Archive throughput</strong></td>
<td>9.6 GB/s</td>
<td>21 GB/s</td>
<td>2.2x</td>
</tr>
<tr>
<td><strong>Power consumption</strong></td>
<td>1.6 MW</td>
<td>1.4 MW</td>
<td>0.9x</td>
</tr>
<tr>
<td><strong>Investment</strong></td>
<td>30 M€</td>
<td>30 M€</td>
<td>1x</td>
</tr>
</tbody>
</table>
HLRE-2: Data Archive

Focus: climate model data and related observations

Content August 2014
HPSS: 36 PB
WDCC: 4 PB (fully documented)

HPSS: 8 PB/year (HLRE-3: 75 PB/y)
WDCC: 0.5 – 1.5 PB/year (HLRE-3: 8 PB/y)
Climate Modelling
Example

CMIP5 RCP Scenarios simulated with MPI-ESM
Global Mean 2m Temperature Change relative to 1986-2005

CMIP5 Release WS MPI-M/DKRZ (Feb. 2012)
Data Amounts CMIP3/CMIP5

  - Participation: 17 modelling centres with 25 models
  - In total 36 TB model data central at PCMDI and ca. ½ TB in IPCC DDC at WDCC/DKRZ as reference data

  - Participation: 29 modelling groups with 61 models
  - Produced data volume: ca. 10 PB with 640 TB from MPI-ESM
  - CMIP5 requested data volume: ca. 2 PB (in CMIP5 data federation)
  - Data volume for IPCC DDC: 1.6 PB (complete quality assurance process) with 60 TB from MPI-ESM

- Status CMIP5 data federated archive (August 2014):
  - 2.3 PB for 69000 data sets stored in 4.3 Mio Files in 23 data nodes
  - CMIP5 data is more than 50 times CMIP3

- Extrapolation for CMIP6 data federation:
  - Volume: 150 PB
  - Number of files: 280 Mio Files
Research: More efficient disk usage

- Files system: Lustre-ZFS
  - lz4 compression increases energy consumption by max. 1%
  - But provides compression of 30% in average
able to conduct production simulations on target grids as large as $10^{10}$ grid elements ($10^4 \times 10^4 \times 400$)

grid spacing finer than 400 m (100 m being the target)

capable of efficiently using a diversity of advanced high performance computing resources

scaling of the model to many tens, possibly hundreds, of thousands of cores
Parallelization on HPC: R2B07 GLOBAL 20km no I/O

Efficiency 83%

Parallel I/O under development
DKRZ offers
Services for End-to-End Workflows in Climate Modelling

Basic Workflows:
- Climate Model Development
- Climate Model Data Production

Poster:
ISC'14, Leipzig, Juni 2014

URL: https://www.dkrz.de/pdfs/poster/DataServices_ENG.pdf
www.dkrz.de
Climate Model Data Production Workflow

Separation:
- Scientific data analysis
- Long-term Archiving
1) Data Management Plan
The data time line as well as volumes, structures, access patterns and storage locations have to be defined as accurate as possible.

2) DKRZ Storage
Each DKRZ HPC project has to specify and to apply for compute and storage resources on an annual basis.

3) ESGF Standardization
Climate data integration into ESGF (Earth System Grid Federation) requires standardization in order to make data intercomparable in the federation.

4) ESGF Services
DKRZ offers a number of services to manage, discover and access climate data in the international Earth System Grid Federation (ESGF).

5) LTA DOKU
LTA DOKU stands for in house long-term archiving in the DOKU(mentation) section of DKRZ’s tape archive. Focus here is on internal data access from data providers.

6) LTA WDCC
LTA WDCC stands for long-term archiving in the World Data Center Climate (WDCC). This service is open for data from DKRZ HPC projects, data from ESGF but also for data from outside DKRZ. These data are fully integrated in the database system of the WDCC. The full set of metadata is provided in order to allow for data interpretation even after ten years or more without contacting the data author. Focus here is on inter-disciplinary data access.

7) DataCite Data Publication
DataCite is an international organization which focuses on publication of scientific data in context with and to be used in the scientific literature. After passing the DataCite Data Publication services these data entities are irrevocable, has been assigned a DOI (Digital Object Identifier) and key metadata including the citation reference are registered in the DataCite repository. The DOI allows for transparent data access and the citation reference allows for integration into the reference list of a scientific publication.
Climate Model Data Federation (ESGF)

ESGF P2P Architecture:
Data Node, Index Node, Compute Node, IdP
LTA: Data ingestion

Services 5 + 6

1. Scientist MD Upload
2. Scientist Data Upload
3. WDCC Data Ingest
4. Long Term Storage (dissemination)
LTA: Data access

Services 5 + 6

Midtier

TDS
(or the like)

Appl. Server

LobServer

Storage@DKRZ

Archive: files

HPSS

Container: Lobs

DB Layer

CERA

• What
• Where
• Who

• When
• How
Future Big Data Developments

• Seamless end-to-end workflow
  – Model data production – analysis – long-term archiving
• Model Code optimization
  – Many cores systems, GPUs
• File systems
  – Compression, parallel I/O, co-design of storage I/O APIs, object store
• Object Storage
  – OpenStack Swift
• Data access
  – Unique Identifier (DOI, Handle System) may replace directory structure
• Network
  – Bandwidth, SLAs, managed networks