HPC in Scientific Computing and beyond: Climate Use Case

Yevhen Alforov

EU-Forscher
Abteilung Forschung
Deutsches Klimarechenzentrum

08 July 2016

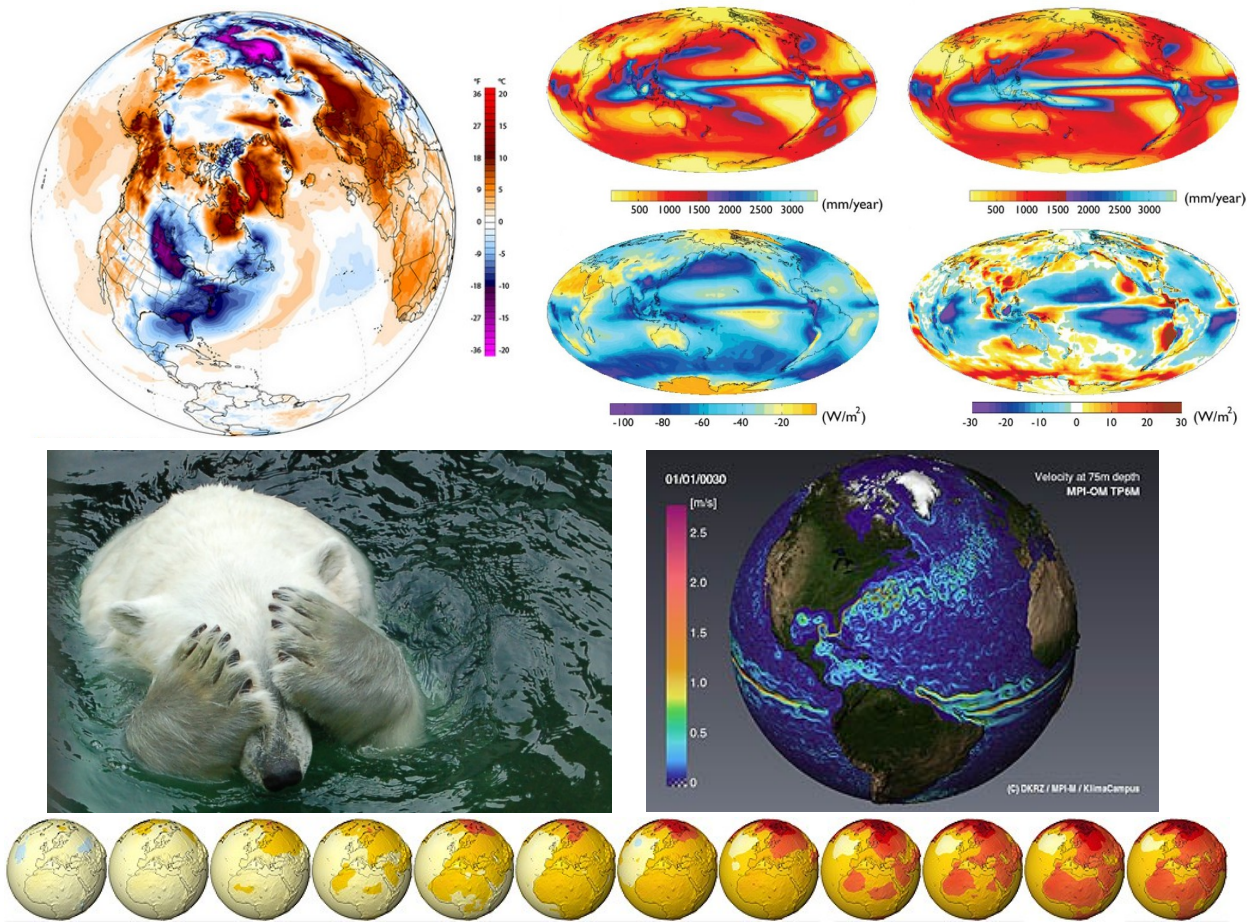
Big Data and Internet of Things

Earth System Research

- General objective
  - To understand how physical, chemical, and biological processes, as well as human behavior contribute to the dynamics of the Earth system, and specifically how they relate to global and regional climate changes.
  - To observe, monitor, analyze, understand, and predict in order to better manage the Earth system.

Climate Science models

Climate Science studies long-term trends of meteorological conditions as well as of their changes over time.



Climate Science models

- Climate Models and Applications
  - Emulate complex interactions in the Climate System
  - Provide simulations to better predict and understand climate variations and change
  - Help to quantify more accurately properties of earth materials and processes
  - Support large-scale experiments
  - Have complex computations
  - Under enhanced development
  - Their computer based simulations drive scientific progress
  - Generate extremely large amounts of data
  - Typical users of HPC infrastructure

Climate applications: example

ECOHAM
(ECOsystem Model Hamburg)
Version 5

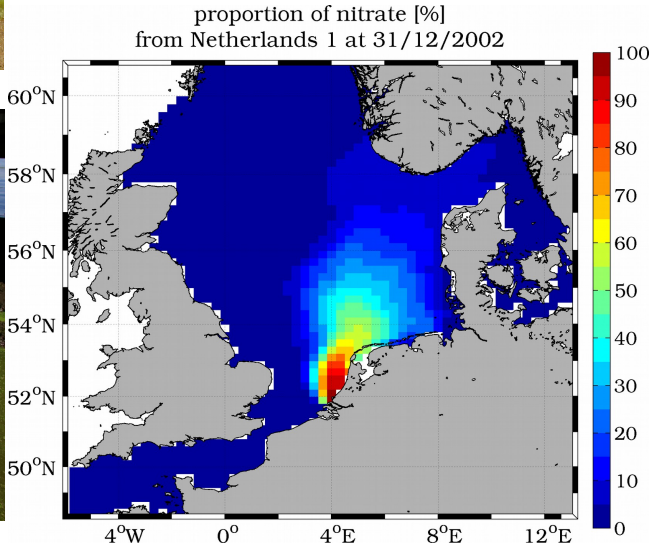
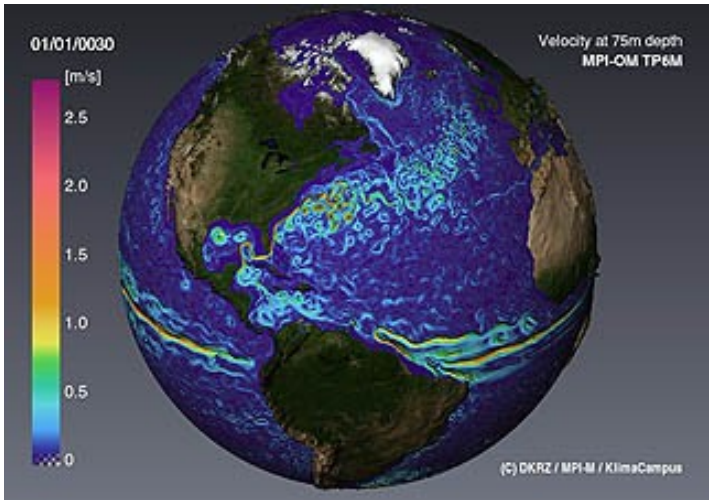


Figure: A three-dimensional ecosystem model for the North Sea

Climate Science models

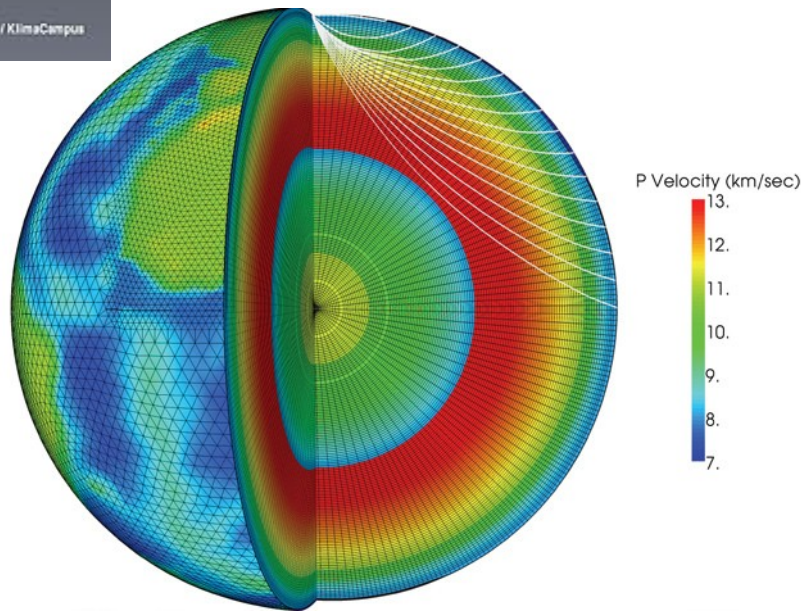
- Climate Models and Applications needs:
  - Stable and uniform consistent access
  - Long runs
  - Access to high volume data archiving
  - Long-term availability
  - Very high inter-processor and I/O bandwidths
- Challenges:
  - Processing and Production
  - Results analysis, interpretation and dissemination
  - Visual comprehension
  - Data management
  - Enormous data storage
  - Huge number of files
  - Program optimization for computer system used

Parallel climate applications in HPC



Discrete three-dimensional grid

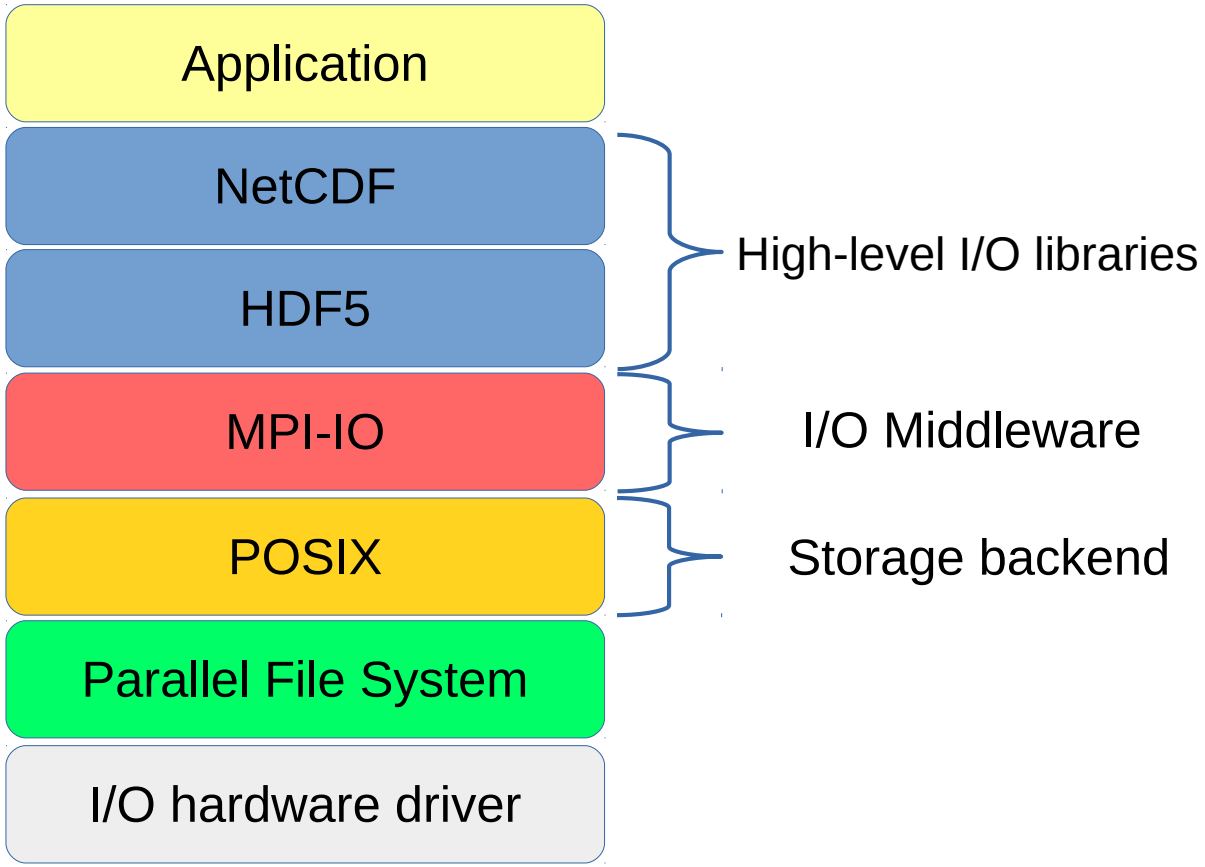
Each point has various quantities



High-level I/O interfaces

- NetCDF (Network Common Data Form)
  - Self-describing
  - System independent
  - Supports creation, access and sharing of scientific data
  - Provides Data portability
  - Provides API for many languages (C, Fortran, C++, Java, etc)
  - Parallel I/O support since release 4.0
- HDF (Hierarchical Data Format)
  - Is a set of file formats (HDF4, HDF5), surrounded by I/O interface
  - No limits on the number or size of data objects
  - Storing data in a hierachical manner: groups and datasets
  - Provides (de)compression of data that is read/written

I/O software stack



High-level I/O interfaces

- CDI (Climate Data Interface)
  - Provides access to Climate and forecast model Data
  - Provides API for C and Fortran
  - Fast and machine independent access to GRIB and NetCDF datasets

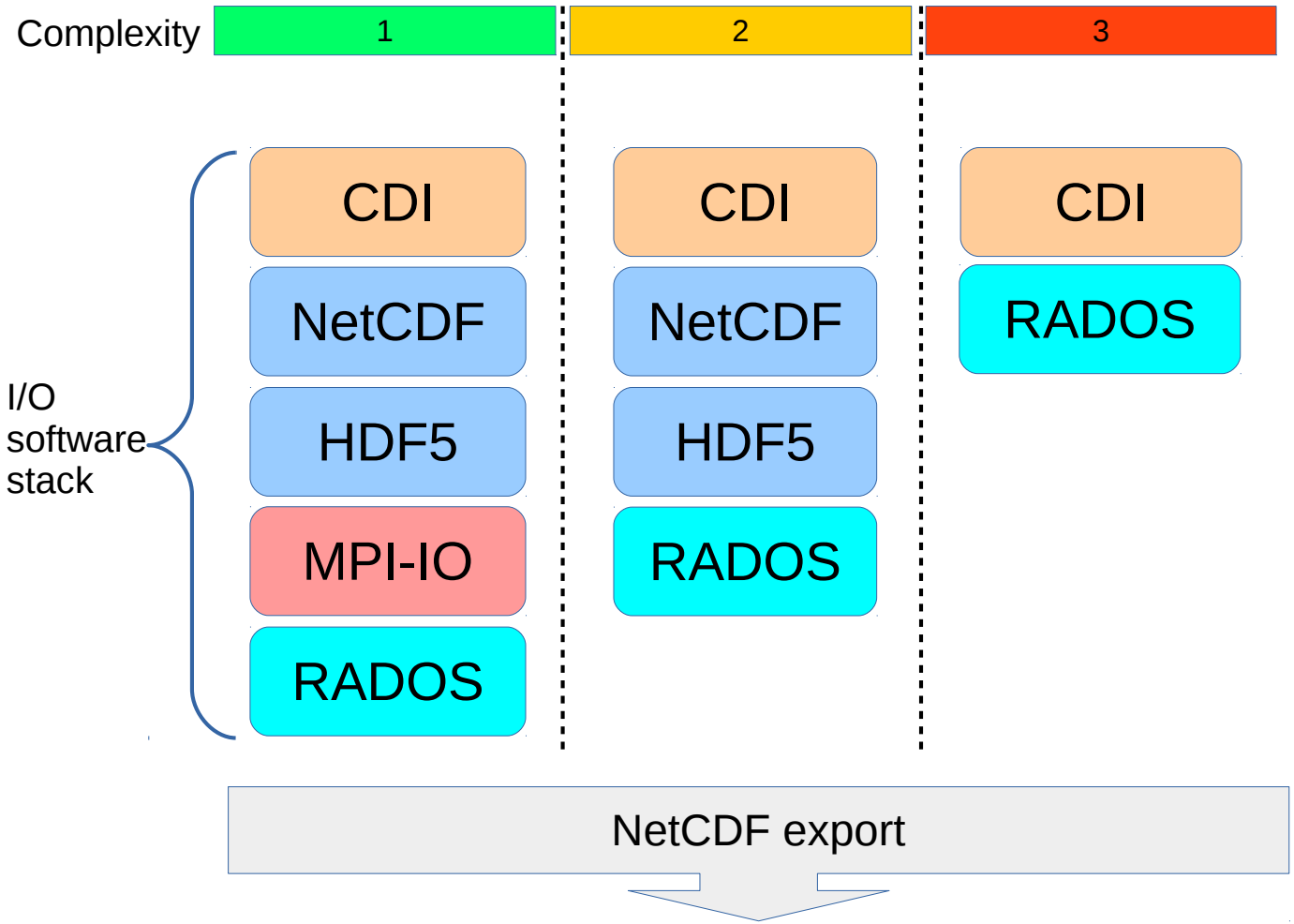
Max-Planck-Institut
für Meteorologie

Ceph

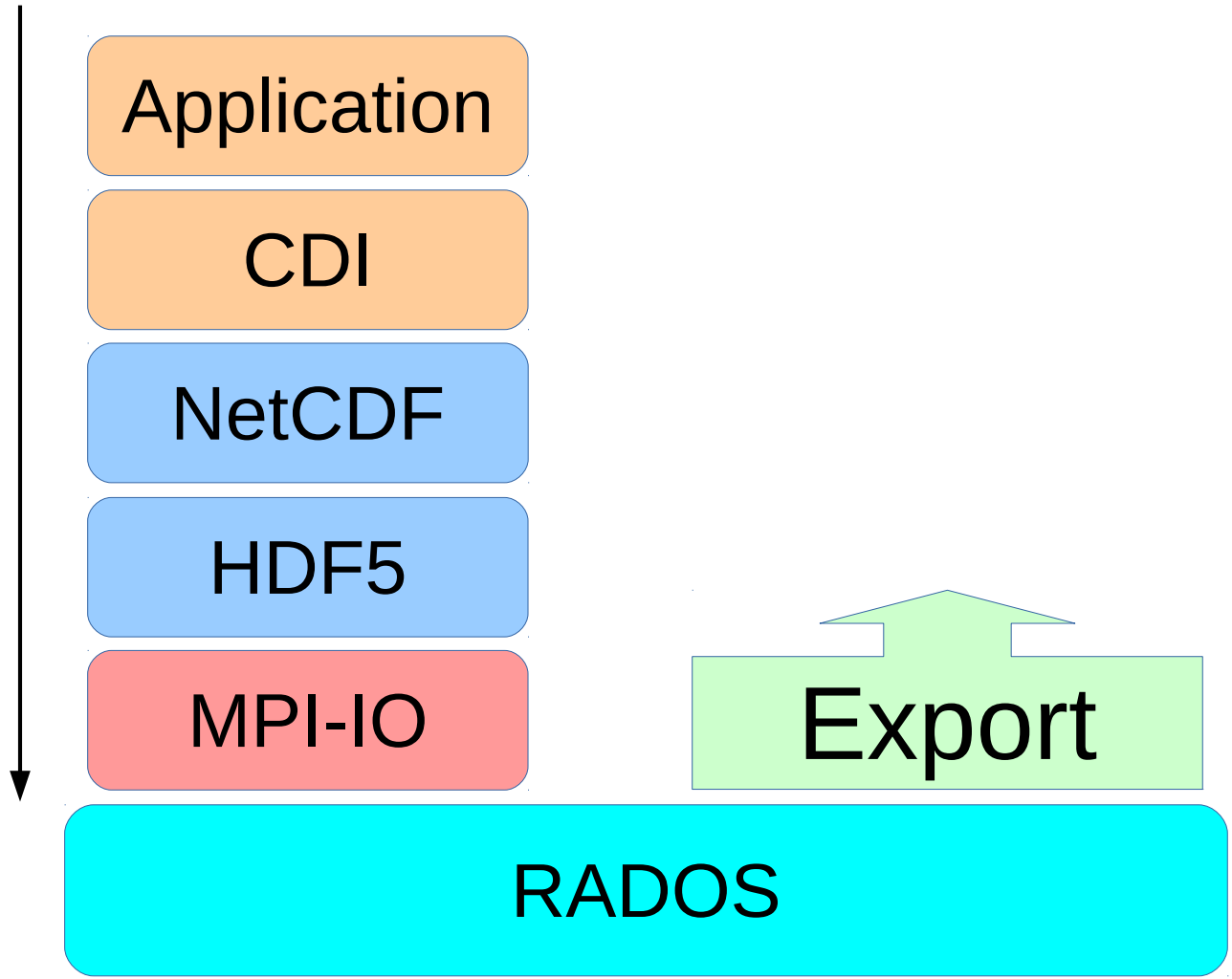
Ceph provides object, block, and file storage in one unified system

- Massively scalable parallel distributed storage system
- Open sourced and portable
- Runs on commodity hardware
- Self-healing and self-managing
- Fault tolerant
- CephFS provides near-POSIX semantics
- Has RADOS (reliable autonomic distributed object store in Ceph) as a backend

NetCDF export feature: Ceph, RADOS, CDI



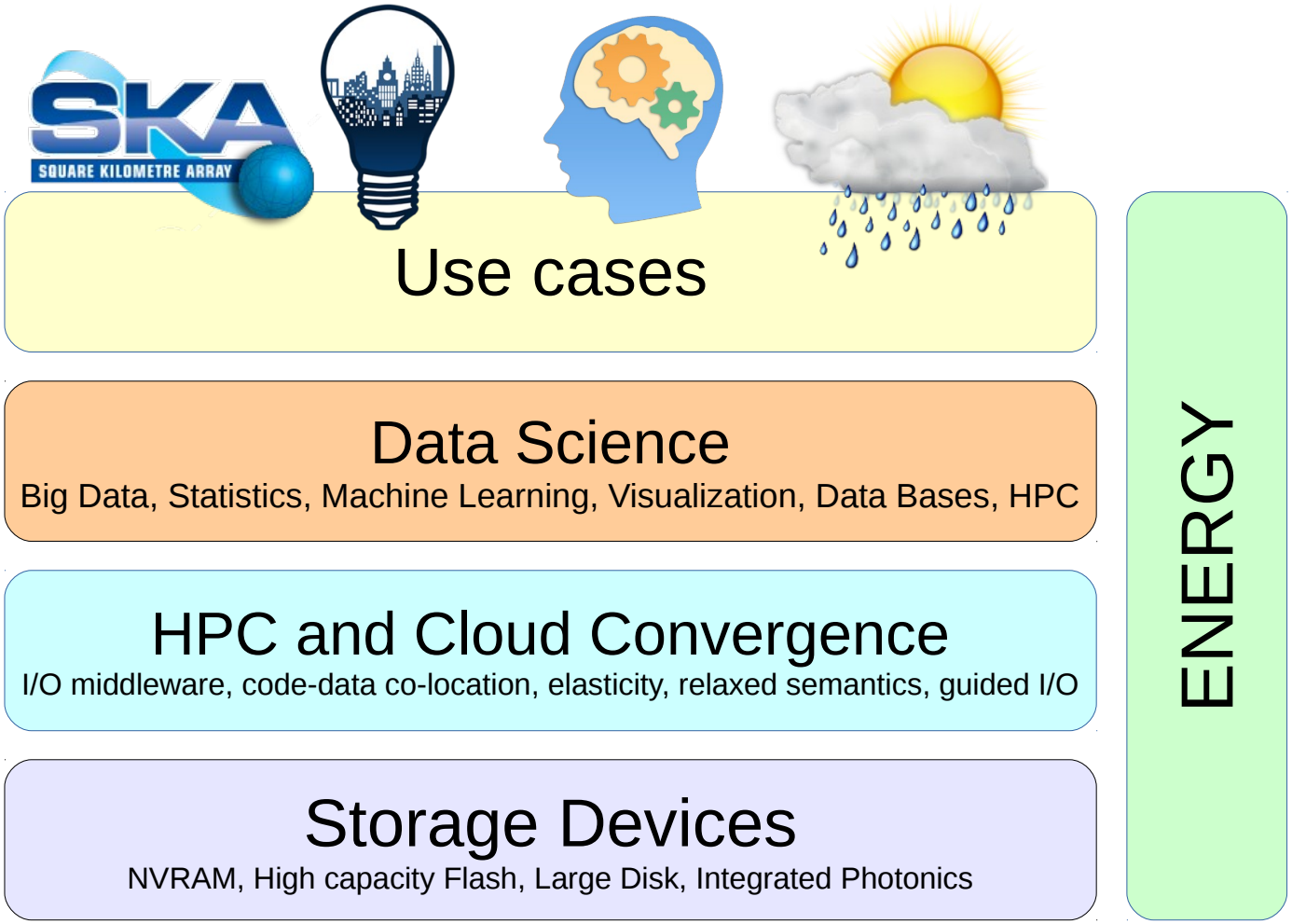
NetCDF export feature: Ceph, RADOS, CDI



BigStorage project work

- Research work
  - NetCDF export functionality on top of RADOS (reliable autonomic distributed object store in Ceph)
  - HPC and Cloud convergence
- Work within WG on Climate use case
  - Analysis and gathering I/O requirements for Climate applications
  - Benchmarks extraction for using by other partners

Big Storage project



Projects in Big Data World

- BigStorage
  - Unifying HPC and Cloud storage architectures
  - Data reduction through novel I/O interfaces
  - Multicriteria Decision Support
  - Energy-impact of data consistency management
  - Big Data processing and management
- SENECA
  - Quality in cloud-related software development
  - Process quality in cloud-related software development
  - Energy-efficiency of cloud computing systems

BigStorage