



The HDF Group

100101010010101000101010  
01001010101010001010101010100  
0101010010010101010101000101010



# Introduction to HDF

## The Company



# Outline

- The HDF Group company
  - Mission and vision
  - Products and services
    - Support to JPSS
  - HDF software evolution
- Intro to HDF5



# THE HDF GROUP COMPANY



# Champaign, Illinois, USA





# The HDF Group

---

- [www.hdfgroup.org](http://www.hdfgroup.org)
- Not-for-profit company (since 2006), ex-NCSA at University of Illinois
- About 40 employees (more than 50% growth in the past 6 years)
- Mission-driven
  - Sustainable development of HDF technologies
  - Guarantee continual accessibility of HDF data

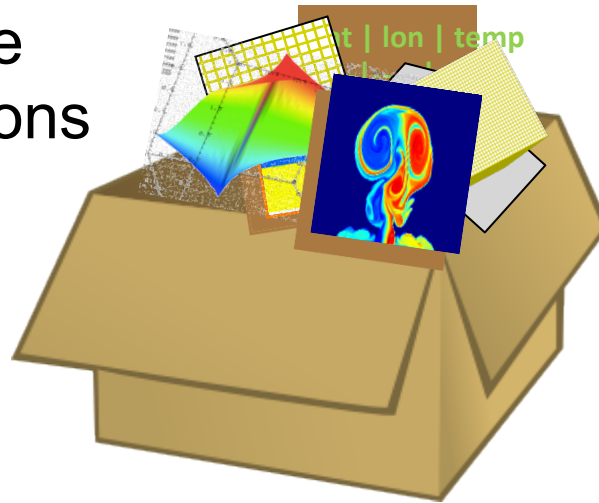


# **DATA CHALLENGES ADDRESSED BY HDF**

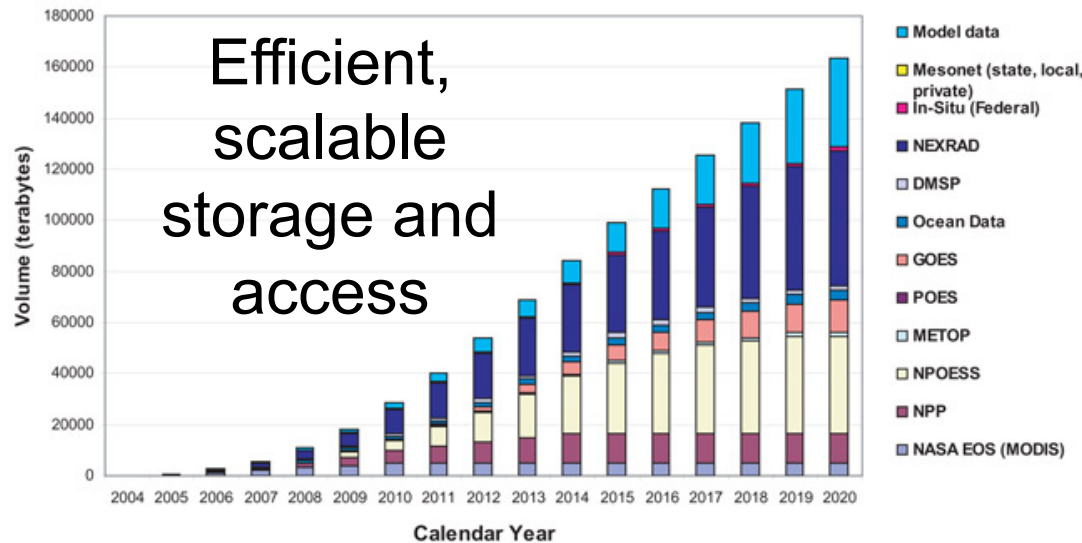
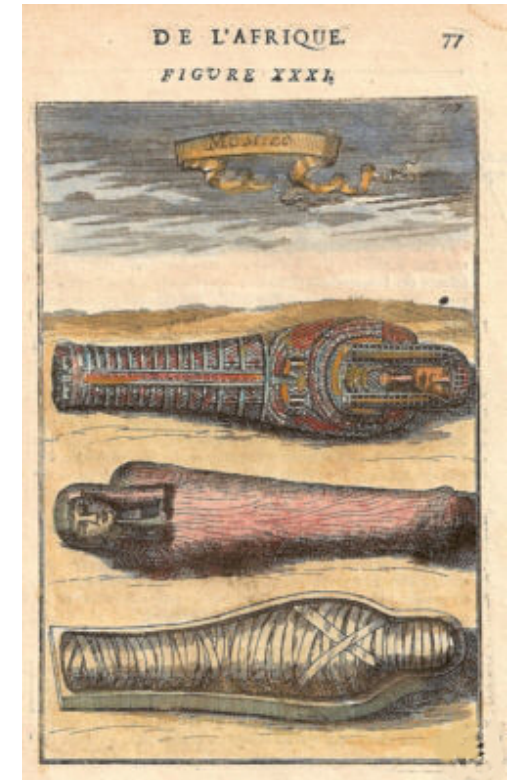


# Data Organization and Preservation

Need to organize complex collections of data



Long term data preservation





## Success stories

---

- Petabytes of NASA remote sensing data in HDF4 and HDF5 file formats
  - New NASA/JPSS missions chose HDF5 format for data archiving





# Data Variety and Complexity

## Describing Data Is Challenging

### Element Types

?

### Basis Functions and Interpolation Schemes

?

### sparse and dense fields

?

### Field value types

$[\rho]$   $\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$   $\begin{bmatrix} s_{xx} & s_{xy} & s_{xz} \\ & s_{yy} & s_{yz} \\ & & s_{zz} \end{bmatrix}$

?

$$\begin{bmatrix} 1s-1 & 1s-2/2s-1 \\ 1s-2/2p-1 & 1s-2/2s-2 \\ \dots & \dots \end{bmatrix}$$

### Mesh Types

?

### Coordinate Systems

?

### Storage Conventions And Data Structures

?

### Mesh Decompositions

?

### Compression

?

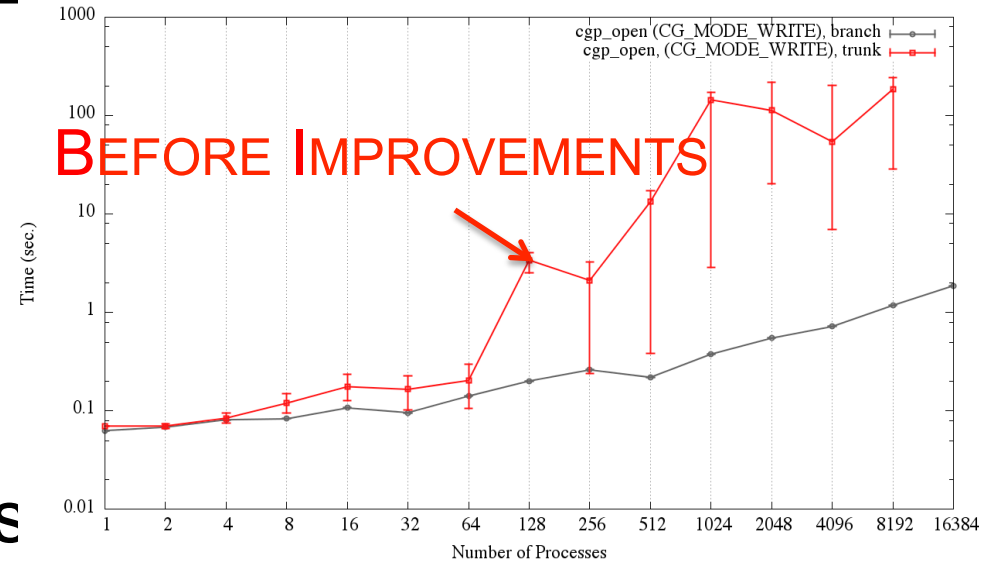
3

Thanks to Mark Miller, LLNL



## Success stories

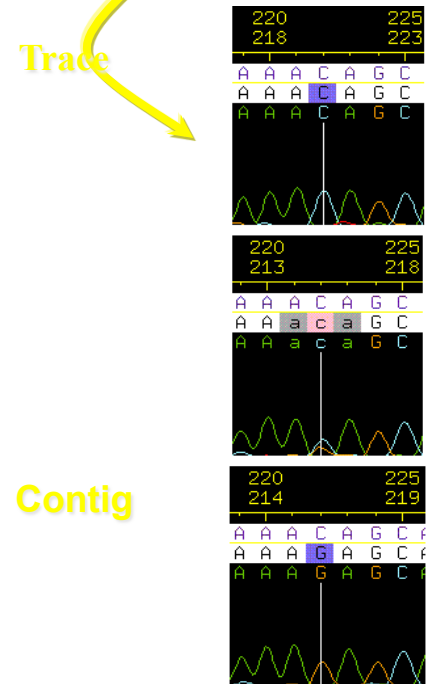
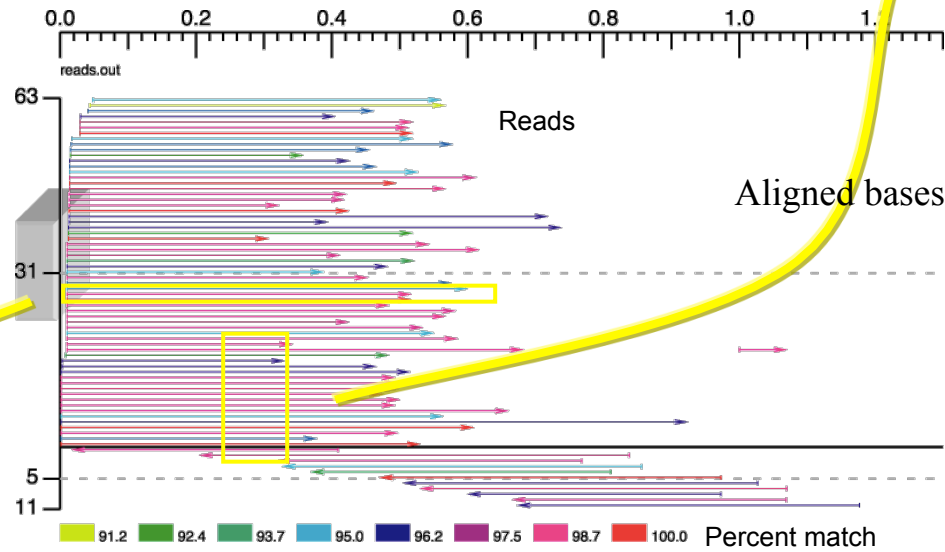
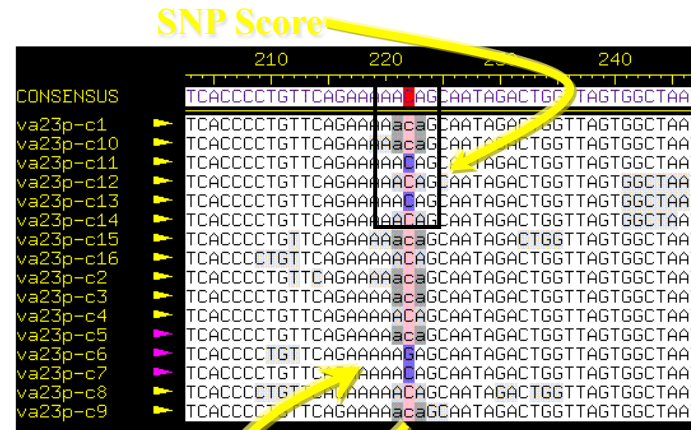
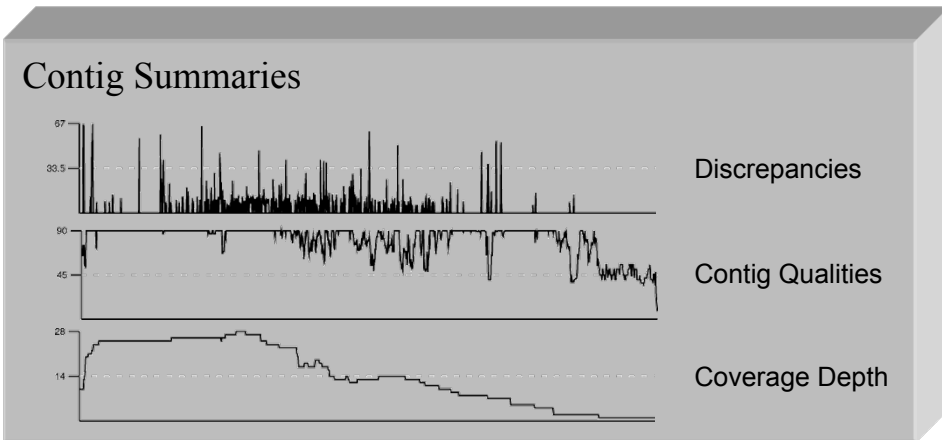
- CGNS – CFD standard
- The HDF Group helped to tune CGNS to solve large scale problems
- Computational mesh size
  - ~33 million elements
  - ~200 million nodes
- Efficiently handles large I/O from Exascale CFD simulations



Thanks to Scot Breitenfeld, THG

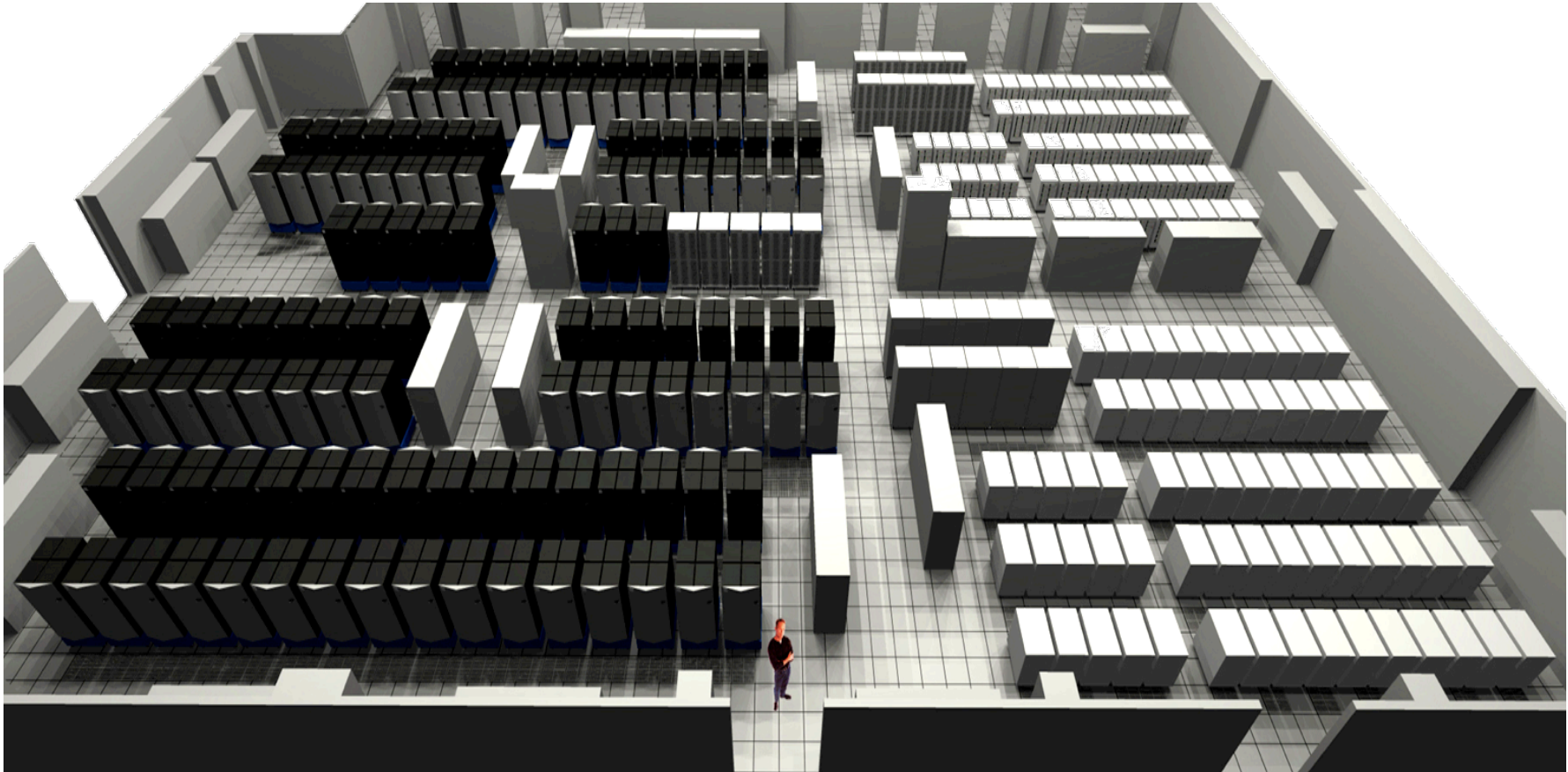


# Complex Data Relationships





# Data Access on Big Computers



... and small computers ...  
and FAST!



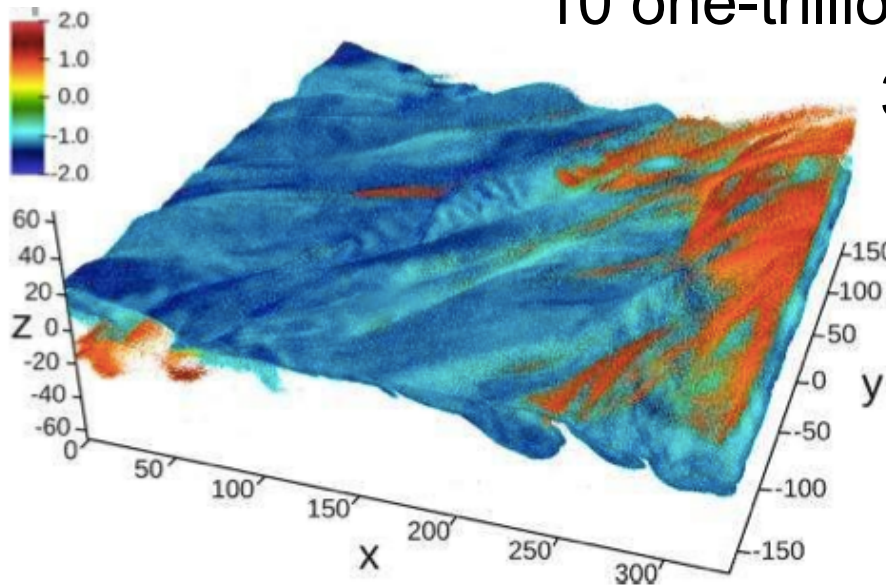


## Success story: Trillion Particle Simulation

- Physics plasma simulation at NERSC Cray XE6
- Simulation ran on 120,000 cores using
  - 80% of computing resources
  - 90% of available memory
  - 50% of Lustre scratch system and writing

10 one-trillion particle dumps of

30-42 TBs in HDF5 files;  
sustained ~ 27 GB/sec;  
total 350 TBs in HDF5







# The HDF Group philosophy

---

- Committed to Open Source
  - HDF software is free
  - BSD type of license
  - Community involvement
    - Testing
    - Patches
    - New features (e.g., CMake support)
- Serving diverse user base
  - Remote sensing, HPC, non-destructive testing, medical records, scientific modeling, etc.



# Brief History of HDF

1987 At NCSA (University of Illinois), a task force formed to create an architecture-independent format and library:  
AEHOO (All Encompassing Hierarchical Object Oriented format)  
Became HDF  

Early 1990's NASA adopted HDF for Earth Observing System project

1996 DOE's ASC (Advanced Simulation and Computing) Project began collaborating with the HDF group (NCSA) to create "Big HDF" (Increase in computing power of DOE systems at LLNL, LANL and Sandia National labs, required bigger, more complex data files).

"Big HDF" became HDF5



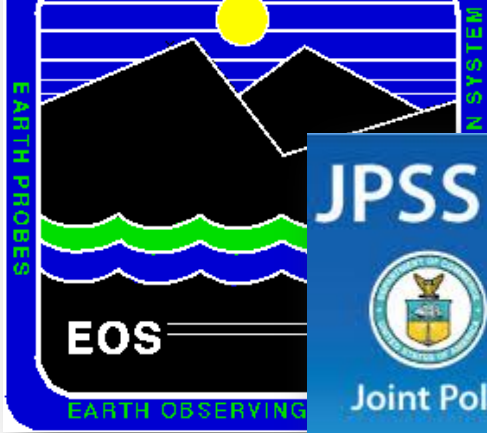
HDF5 was released with support from National Labs, NASA, NCSA

2006 The HDF Group spun off from University of Illinois as non-profit corporation





# Members of the HDF community



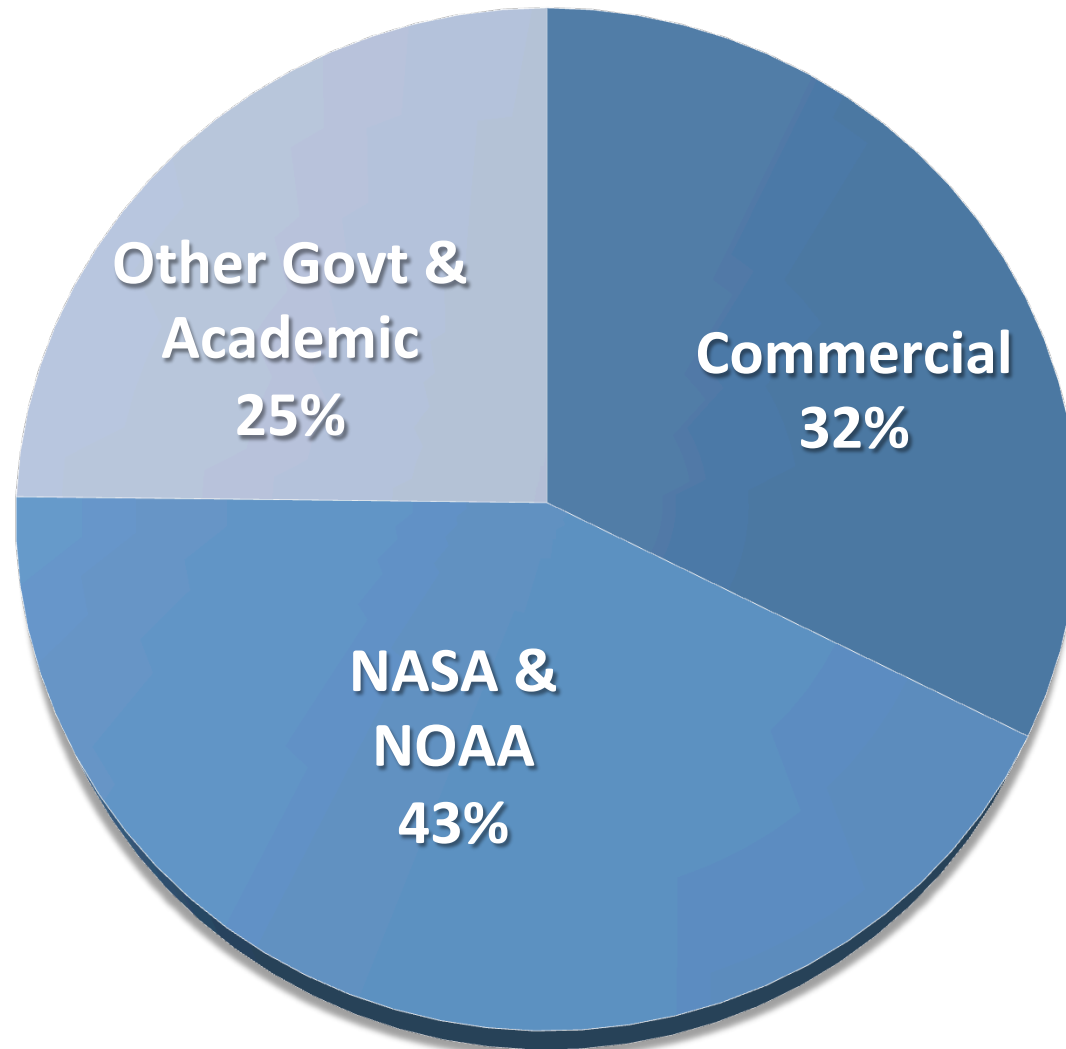
china eu india japan korea russia usa







## Revenues by source





# PRODUCTS AND SERVICES



# The HDF Group products

---

- Main product: **HDF Technology Suite**
  - For managing high volume complex, heterogeneous data
  - Flagship: HDF5 data store
    - Flexible and efficient storage and I/O
    - Portable
    - Highly customizable
    - Misc. tools
  - Specialized software and tools (e.g., JPSS)



# HDF5 Software

- HDF5 runs on all flavors of Linux, Mac OS X, Windows, AIX, Solaris, FreeBSD, Cray, etc.
- GNU, Intel, PGI compilers
- Platforms and configurations tested for each release are listed at <https://www.hdfgroup.org/HDF5/release/platforms5.html>
- Talk to us if you need help to port HDF5 to unsupported platform!



# The HDF Group services

---

- Helpdesk and mailing lists
  - [help@hdfgroup.org](mailto:help@hdfgroup.org)
  - [hdf-forum@hdfgroup.org](mailto:hdf-forum@hdfgroup.org)
  - Open to all users of HDF



# The HDF Group services

---

- Standard support
  - Assistance in general areas of HDF usage
- Premium support
  - Access to our consulting and training resources
  - Limited consulting hours are included
- Enterprise support
  - Help with developing common strategies for managing HDF data within organization
  - Organization shares consulting/troubleshooting services
- Training
- Consulting and custom development



# SUPPORT FOR JPSS



## Goal

Provide HDF5 risk-reduction support for the distribution of NPOESS VIIRS, OMPS and other sensor and environmental data products.





# Focus Areas

- Maintain JPSS specific software and features developed by The HDF Group
  - Tools: HDFView, nagg, h5augjpss
  - Library: HL library for handling HDF5 references
- Perform regular testing of the HDF5 and JPSS specific software on the platforms critical to JPSS users including AIX and Red Hat Enterprise Linux
- Provide user support
  - [help@hdfgroup.org](mailto:help@hdfgroup.org) (add JPSS to Subject line)
- Perform special maintenance tasks (e.g., bug fixes, new features implementations)
  - Troubleshooting performance issues when using compression with JPSS data
- Perform special research projects as requested
  - Serving S-NPP Data and Metadata



# More information on the project

- <https://www.hdfgroup.org/projects/jpss/>

The screenshot shows the HDF Group website's JPSS Downloads page. The navigation menu at the top includes Home, Why HDF?, Products, Services, About Us, News, Blog, and Contact Us. A search bar is located in the top right corner. The main content area is titled "DOWNLOADS" and contains the following information:

HOME > PROJECTS > JPSS

## DOWNLOADS

The latest versions of the software developed for the JPSS project are available for download. See the list below.

### nagg Release 1.6.2 April 14, 2015

nagg is a tool for aggregating JPSS data granules from existing files into new files with a different number of granules per file or different combinations of compatible products than in the original files. The tool was created to provide individual users the ability to rearrange NPP product data granules from downloaded files into new files with aggregations or packaging that are better suited as input for a particular application.

The 1.6.2-beta release removes the limit on the total number of granules processed and adds two new command options:

- nagg now sorts input files by the timestamps in the filename and processes granules from each subset of files according to the timestamps.
- --onefile command option to aggregate granules from all input files into a single aggregation, creating one output file for packaged output or one output file for each product for unpackaged output
- --nofill command option to suppress creation of fill granules when granules are missing from a time sequence or to match a compatible product

The tool was tested on Linux 32 and 64-bit systems. For more information on this release, see the [RELEASE.txt](#) file.

[Source code and binaries](#)

(For earlier versions, see: [All Releases](#))

### High-level Library for handling HDF5 object and region references Release 1.1.3 December 9, 2014

The library contains C and Fortran APIs to:

- Get information and read data pointed to by region references
- Create an array of region references using paths to datasets and corner coordinates of hyperslabs
- Create a dataset and write data pointed to by region references



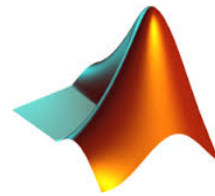
# OTHER HDF5 SOFTWARE



## Other software

- The HDF Group is working closely with the makers of commercial and Open Source software

- MATLAB
- IDL
- h5py
- PyTables
- netCDF-4
- HDF-EOS2(5)
- OPeNDAP





# HDF SOFTWARE EVOLUTION



## HDF5 status

- Current version is 1.8
- HDF5 1.8.15 released on May 15, 2015
- Maintenance releases are every 6 months
  - November and May
  - Bug fixes, new features that do not require file format change
- Major new release in works: 1.10
  - New features: SWMR, VDS, Parallel improvements
  - Alpha-beta series starts on October 31, 2015
  - Full release is March 31, 2016

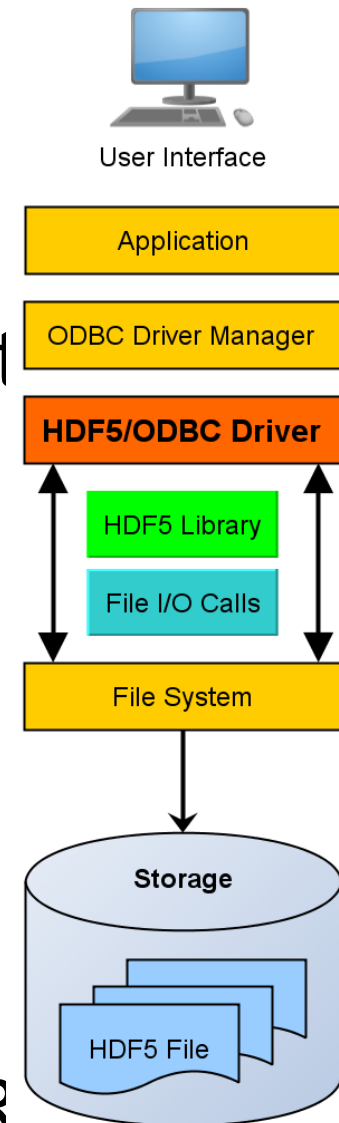


# NEW DIRECTIONS



# HDF5 ODBC Driver

- Open DataBase Connectivity (ODBC)
- Industry standard middleware API for accessing database management sys.
- All analytics apps. have an ODBC client
- **HiFive** – ODBC driver for HDF5
  - Windows, [Linux, MacOS X]
  - Client & Client/Server
  - Accessing HDF5 files from Excel & R



Thanks to Gerd Heber, THG





# HDF5 for the Web

- Can I access HDF5 files remotely?
- API? My (mobile) client speaks HTTP!
- What is a file system? Who uses files anymore?
- Cloud computing w/ HDF5



**OAuth**<sub>PUT</sub>  
**JSON**<sub>GET POST</sub> **XML**  
**REST**  
**Basic-Auth**  
**DELETE**

Thanks to John Readey, THG



# Acknowledgement

---

This work was supported by SGT under Prime Contract No. NNG12CR31C, funded by NASA.

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of SGT or NASA.



The HDF Group



# Thank You!

Questions?