HDF Project Update

Mike Folk, Elena Pourmal, Kent Yang, Joe Lee
The HDF Group
Briefing to ESDIS
April 4, 2013
Project Goals

1. Support activities for NASA HDF data providers and data users, and the EOSDIS Core System (ECS)
2. Maintenance and quality assurance for HDF4 and HDF5 libraries and utilities
3. Evolving the HDF4 and HDF5 libraries and utilities to meet new NASA EOS requirement
4. Integration with complementary technologies and application domains
5. Prototype study of web technologies to access HDF data
Support activities for NASA HDF data providers and data users, and the EOSDIS Core System (ECS)
Support Activities

• Outreach
  • Mailing lists, Website
  • Conferences, Workshops, Other
• EOS support
Outreach – Mailing lists

• Mailing lists and archives
  news@hdfgroup.org
  http://hdfgroup.org/news/

  hdf-forum@hdfgroup.org
  http://mail.hdfgroup.org/pipermail/hdf-forum_hdfgroup.org/

• New mailing for NASA DAACs
  hdf-nasa-daac@hdfgroup.org
HDF ESDIS PROJECT GOALS

The primary goals of this project for the period between July 15, 2011 and March 31, 2015 are in the following areas:

Support activities for NASA HDF data providers and data users and for the ESDIS Core System (ECS)

As ESDIS matures and increasing amounts of data are archived, current users will be joined by users with an ever-broadening range of applications. Usability and accessibility of EOS data are not only crucial towards global climate change research but also become increasingly important for policy makers, educators, the private sector and the general public. The HDF Group will provide continuing support in the form of the HDF helpdesk, workshops, training and documentation. The HDF Group will also work with the ECS, DAACs and other data centers to improve usability and accessibility of EOS data by means of on-site visits, consulting and tutorials.

Maintenance and quality assurance for HDF4 and HDF5 libraries and utilities

Maintenance includes making minor feature changes to address ESDIS requirements, correcting errors, keeping the software, test suites, configurations, and documentation current, and conducting periodic releases of the software. Quality assurance involves upgrading and extending software testing, reviewing and revising documentation, improving the software development process, and strengthening software development standards.

Evolving the HDF4 and HDF5 library and utilities to meet new NASA EOS requirements

Two focused areas include enhancing HDF5 backward and forward compatibility support and developing tools to improve the accessibility and usability of data stored in HDF5.

Integration with complementary technologies and application domains

Foremost is to operate well with HDF-EOS technologies, which means making sure that the two perform efficiently together and that the HDF-EOS library and tools use HDF as effectively as possible. Other technologies, such as OPeNDAP, netCDF, XML, and GIS can add tremendous value when effectively integrated with HDF.

Prototype study of web technologies to access HDF data

Investigate other web technologies that can benefit the access of NASA HDF data.

It’s updated: http://hdfgroup.org/projects/esdis
Outreach – Conferences, Workshops

• 11\textsuperscript{th} NASA ESDSWG Conference, Nov. 2012
  • Poster: HDF Support for NASA EOSDIS Data Centers and Users

  • Poster: Lead by Examples: Serving NASA HDF User Communities with Examples at hdfeos.org Website
  • Poster: Challenges of Meeting the CF conventions in NASA HDF/HDF-EOS Data
Outreach - Others

- Participated in the ESDS HDF5 conventions WG telecons and contributed material for CF and interoperability issues.
- Participated in the ESDS User Needs WG telecons.
- Worked with ESDIS staff to write an article for Earth Observer. The article introduces the comprehensive example page on the hdfeos.org website.
- Advised ICESat-2 on HDF5 Product Builder.
- Held HDF-DAAC teleconference (3/7/13)
EOS Support

- EOS2 and EOS5 are tested daily with HDF4 and HDF5 development code.
- HDF-EOS website now has:
  - MEaSUREs VIP and NPP VIIRS IDL/MATLAB/NCL examples.
  - descriptions and a search box to help users access the website easily.
Goal #2

Maintenance and Quality Assurance for HDF4 and HDF5 libraries and utilities
Maintenance and QA Activities

- **Perform maintenance releases** of HDF software:
  - Support **OSs and compilers** identified by ESDIS.
  - **Address issues** and software evolution requests from ESDIS.

- Perform SQE and SQA activities:
  - Put in place missing processes.
  - Improve existing processes.
  - Build in software quality by performing testing and using the best software development practices.
## Maintenance Releases 2012–2013

<table>
<thead>
<tr>
<th>Year</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDF4</td>
<td></td>
<td>4.2.7</td>
<td>patch1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.2.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDF5</td>
<td>1.8.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CF</td>
<td>1.8.10</td>
</tr>
<tr>
<td>HDF-Java</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CF</td>
<td>2.9</td>
</tr>
<tr>
<td>h4h5 tools</td>
<td>CF</td>
<td>2.2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDF4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDF5</td>
<td>1.8.10-patch1</td>
<td>CF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDF-Java</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CF</td>
<td>2.10</td>
</tr>
<tr>
<td>h4h5 tools</td>
<td>CF</td>
<td>2.2.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CF – Code Freeze
6 weeks for HDF5
4 weeks for other products
HDF4 maintenance releases

HDF 4.2.9 (February 2013)

• Support for Mac 10.8 with Intel and Clang compilers
• Support for Cygwin version 1.7.7 and higher
• Documentation updates
HDF5 maintenance releases

HDF5 1.8.10 (Nov 2012)
HDF5 1.8.10-patch1 (Jan 2013)

- Many improvements in tools
  - Interoperability between h5dump and h5import
  - Performance improvements in h5diff for the files with many attributes
- Support for I/O sizes bigger than 2GB on Mac OS X
HDF5 maintenance releases

• HDF5 1.8.11 (**May 2013**)  
• ESDIS requests  
  • Provide clean output of h5dump (suppress metadata)  
  • Memory leaks exposed by netCDF-4  
  • Cygwin support  
  • CMake and Windows enhancements
HDF5 maintenance releases

- HDF5 1.8.12 (**November 2013**)
- ESDIS requests
  - **No specific requests** from the users yet
  - 12 general maintenance issues (Windows support and Fortran, C++ wrappers)
Future releases

• Request to support wide character filenames (MathWorks)
• Request to support UTF-32 encoding (H5Py)
• Request to support parallel compression
• h4toh5 2.2.2 (June 2013)
• Based on HDF 4.2.9 and HDF5 1.8.11
  • New binaries for Linux and Windows
Java maintenance releases

2.9 release (December 2012)

- Based on HDF 4.2.8 and HDF5 1.8.10
- **ESDIS requests** (total 9) including:
  - Show groups/attributes in creation order
  - Export data to a binary/ASCII file without having to open the object in the TableView
  - Reload feature to close/open file
  - Improvements for installation
Java maintenance releases

2.10 release (**December 2013**)

- Based on HDF 4.2.9 and HDF5 1.8.12
- **ESDIS requests** (total 31 as of today) including:
  - 0 or 1-based indexing when displaying arrays
  - Displaying long names of the files ("…" in the names)
  - Ability to modify HDF4 compressed dataset
  - Support netCDF-4 files with VL attributes
  - Many items from the October’s 2012 briefing wish list and users’ survey
User support issues and QA

- In 2012 HelpDesk received 1404 emails and sent 1241 emails
- Total issues 612
  - 82 issues from EOS priority users
  - All issues were resolved
- QA
  - Continue extensive daily regression testing including profiling tools and performance framework
  - Require peer code review for all changes in HDF5 and HDF4
Goal #3

Evolving the HDF4 and HDF5 libraries and utilities to meet new NASA EOS requirement
Support New OSs and Compilers

- HDF software is now supported on
  - SunOS 5.11 (Sparc) with Studio 12 compilers
  - CentOS 6 with GCC and Intel compilers
    - We will add PGI compilers in May
  - Mac OS X 10.8.* with Clang and Fortran, Java 1.7 (Oracle)
  - Cygwin 1.7.7
  - Windows 7 with VS 12 and Intel 13
  - Windows 8 with VS 12 and Intel 13
HDF5/JSON

JavaScript Object Notation
• Text encoding of JavaScript object and array literals
• Use cases similar to DDL and XML
  • Text representation
  • Diagnostic
  • HDF5 “blueprints”
  • Catalog records
  • Exchange format
  • Web services (REST)
  • NoSQL document stores

Advantages:
• Less noise (XML tags)
• Multi-dimensional arrays
• Binary encoding (BSON)
• Programmable (JavaScript)
• Browser support
• NoSQL document stores

Tools:
• BNF grammar
• h5json HDF5 → JSON
• jsonh5 JSON → HDF5
• Release date in Q2/2013
To evaluate the HDF Group’s data viewing tools and user needs, and to explore, recommend, and prioritize improvements.
Goal #4

Integration with complementary technologies and application domains
HDF and netCDF interoperability tools

- HDF4/HDF-EOS2 to CF conversion toolkit (hot!)
  - Release on June 15th, 2013
- HDF-EOS5 augmentation tool (maintenance)
  - Release on December 15th, 2013
- HDF-EOS2 dumper tool (maintenance)
  - Release in every other year
- HDF-EOS5 to netCDF-4 conversion tool (retired)
- HDF4 Handler
  - Possibly release on September 15th, 2013
- HDF5 Handler
  - Release in June, 2013 to synchronize with the Hyrax release
Prototype study of web technologies to access HDF data
Prototype Study

- THREDDS vs. Hyrax (OPeNDAP)
- Apache Open Source Incubator Pilot Project
- Digital Object Identifier (DOI) support in HDF5
Thank You!

- Later **for details**: Update on NASA Wish List
- Questions?
- Comments?
Acknowledgements

This work was supported by Subcontract number 114820 under Raytheon Contract number NNG10HP02C, funded by the National Aeronautics and Space Administration (NASA). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Raytheon or the National Aeronautics and Space Administration.