HDF Town Hall **[1]**

Submitted by erinmr on Sun, 2014-06-29 16:47   Friday, July 11, 2014 - 09:15 to 10:45
Event: Summer Meeting 2014 [2]
Session Type: Breakout [3]
Expertise Level: Beginner [4]
Abstract/Agenda:
The HDF format is used across a broad spectrum of scientific disciplines and computing environments. We are planning our next major release for later this year and will discuss new features and how they might support Earth science requirements. We would also like to hear from you about how you are successfully using HDF or about challenges that we might work towards overcoming.

Notes:
HDF Townhall

Goal – facilitate two kinds of communications

1)   HDF group (Mike the visionary leader) will present where we are and some new idea-features

2)   Feedback from users

HDF Project Update – Mike Folk

·   It is an update on what the group is doing, project, and latest work in technical area

·   Mission – provide high quality software, services to support users of the software, focus on the entire data lifecycle

·   Not-for-profit based in Champaign – HDF group started in 1988

·   Group Services – core software maintenance and distribution, helpdesk and mailing list (free)

o   (not free) – priority support, enterprise support (work with dept/organization within a larger group), consulting, training special projects

·   Funding varies from year to year but Earth Science – EOS are consistent…. Others are high speed detectors, high performance computer, various

·   Revenues by source – DOE/other govt/academic 48%, NASA/other Earth Science 43%, Commercial/foreign 9%

·   Technical activities

o   4 technical areas – technical operations, core software, earth science, applications

o   Trying to grow applications

Earth Science Activities

ESDIS

·   http://www.hdfeos.net [5]

o   Maintain HDFeos website – provides services, user support, forum

o   Put more emphasis on demos and examples

o   HDF-EOS tools

o   3,500 visitors per month

·   Web services

o   OPeNDAP, THREDDS, ENVI service engine

o   Want to know from you – what web services would you like to see on HDF-EOS.org

o   Send demo codes too

o   New tool examples have been added

·   Slideshare

o

·   All workshop slides available through slideshare – 27,000 views in 2014

·   Follow on twitter @HDFEOS

·   EOS Tools maintained

o   Help convert to something that can be used by something else

o   OPeNDAP handler for HDF4 and 5 – continue to work and hand new data products

·   Other ESDIS – general maintenance, QA, and user support – this supports a lot of HDF

o   HDF5 Product Designer – excited about – came from a community member

o   CERES to help migrate HDF4 to 5 – think will provide a president or examples for others

o   Activities in ESDSWG working group

JPSS (Joint Polar Satellite System)

·   Lots of work on tool development

o   Augmentation, aggregation and attribute editor tools

·   Small project/studies – compression for NPP (National Polar Partnership) products, web services, and metadata conventions

·   Maintenance and test on their system

·   Ted has done a lot of expanded scope esp in metadata & standardization

o   GeoTiff, ISO TC 211, Ocean Observation initiative, CH2MHill, and EarthCube

General

·   HDF forum – active mailing list – relieves help desk, testing and configuration, can even lead to funding

·   Product maintenance is the key – library and tool releases…. Need input on priorities

·   Release schedule – HDF4 once a year (Feb), HDF5 twice a year (May and Nov) – no major releases in a while (2007) – continue to add capability – Java follows HDF5 release

Platform support

·   It is important to get feedback on the platform that HDF supports

·   Platforms to drop – windows and max … add mac and GNU … can add linux or windows

Recent and upcoming new Capabilities

·   Concurrent read/write file access – SWMR – single writer/multiple readers

·   H5watch tool – watches a data set – writes when records are appended

·   Virtual Object Layer (VOL) – idea from 1998, but never flew until others needed it – support HDF5 api but doesn't require HDF5 file format – plug in layer allows for file to be read/write to something other – current plugin is HF5 library which write HDF5, but can use file system where groups would be directories

·   Direct chunk write

o   If data in memory looks the same as how you want it to look in file

o   Bypasses parts of the HDF library

o   For 2+ GB compress as 1.5 GB

o   When know what the problem is, then can tune it for specific kinds of things

·   Fault tolerance through "journaling" – H5recover tool to restore metadata in a file

·   Faster i/o with "metadata aggregation" – when write in HDF lots of small pieces – now page it to create blocks that are pages and write it out together

·   Dynamically loadable filters – filter = module that does compression or encryption

·   Persistent file free space tracking/recovery – when delete it leaves a hole – new info goes into the end… now writer can write into holes

·   Asynchronous I/O – available in 1.10

·   H5repack and h5diff – performance improvements – very useful tool for ESIP community

HDF5 1.10.0 Roadmap

·   This is the next release

·   Talked about most of the capabilities

o   HDF5 append only capability – if you knows it is just append to datasets, then index structure is better

·   Hope is release within a year (or sooner) – working on little details for release

Conclusion – a hero application

- LBNL trillion particle simulation – plasma physics groups

o Problem – support I/o and analysis needs for state-of-the-art plasma physics code

o 120,000 core machine, 250 TB dataset

o Scalable write and analysis – novel indexing (Fastbit) for fast querying

§ Index dataset in 10 minutes; query in 3 seconds

- Q – list of platform – does that include binaries for each – No don't have binaries

o Choice of a binary or a download and build on the website

- Q – on the build package distribution topic – check NetCDF list serve – track how many questions people are stumped between HDF and NetCDF – a large % of the question are build NetCDF and having trouble couple HDF5 – built/maintained by 2 organization – hope sometime this gets address – as HDF gets more complex and NetCDF takes on more CF…

o This is second – NetCDF library is how Java are working with NetCDF and HDF4 – it is a troubling route

o A: regularly test NetCDF4 – but point is that there isn't coordination

o Alex - The built process – average user has to build HDF before NetCDF from two different locations – get a build with NetCDF in HDF

- Q problem is also because people install RPM system – really want those to be dialed in – if you have installed a different version of HDF5… then it is broken … maybe run on young repository – binaries

o It is out of their hands… not unidata… it is all

- Problem is when need interdependencies – gdal requires specific HDF5 shared version – it is always wrong in how you would load it – it is just coordination at the top – because it is so often repeated that we know so well

o Ted – maybe help some of the traffic to the HDF forum

o Problem shows up in openstack and generic user support forums

- James - Think about getting together with dbn and rpn and target it to populate with our stuff – want it totally dialed – maybe ping esip that runs a repository for these

- There is an analogy in numpy and scipy – superpack – much easier… bundle together a package

- OPeNDAP and HDF are type 2 are research – I think this happens in type 3 is application (mainly private sector entities) … type 3 ESIP members is where this would be discussed… there is no type 3 forum or place

o What is the barrier to getting into the main ecosystem – when working with shared object libraries – not the how set of rpn packages, … you have to be able to talk about programmer and object interface – means different mindset

o ACTION

- The HDF EOS might be a good place to deal with this

o   Move hdfeos forward – possibly in hdf

o   Part of the issue is amount that NASA is interested in it

Ted

o   Outreach – NSF and NASA metadata – outreach in data is important – this is only the second HDF workshop with ESIP (first one was last year). Want to continue communication with NASA – might need more days again – happy that OPeNDAP followed the lead

o   Have lots of outreach with users and DAAC – Mike mentioned tools and examples that were extended – interaction with science teams is super critical but haven't done very well – one is science developer tool kit – which is what the science team see…

o   Had 3 interactions this year – CERES has gone very well – started with Walt Baskin (uses JAVA and HDF) – CERES team came to Walt… early in their process first improving HDF4 – going to HDF4 with more conventions – next version is HDF4 and soon HDF5 and then migrate archive - got to us early – Kent wrote them a 2-3 pages samples with clear guidance

o   Q are they currently changing archive – Yes – Kent has wanted to, specifically with dimension scales- need to improve the conventionality of datasets

o   Mopit – went to HDF5 – Joe and Ted met with them – similar discussions… they already wrote all HDF5 files – but can't implement there – intent was right – they use IDL and IDL is their primary client and IDL is 4 or 5 releases behind and IDL doesn't use compression

o   OBPG (ocean biology processing group) – subtribe at Goddard – move to HDF and support CF, ACDD, and ISO – Ted has tools – found they weren't very compliant – thanked but they weren't that interested anyway – then Joe added non-invasive augmentation … made them compliant in the classic way of using NCML by putting convention on top of file

o   3 positive outcomes – ACTION – write-up as a 2 page

o   Another place where need interaction with science team is product designer

o   Jeff lee is the target audience and less technical user make easier to use

o   Need central community

o   SWOT, ISAR -… missions that are early in development are the target audience- want to get to the "jeff" or their bosses

o   Q how will it work with missions that have already launched

§ Kent is working with GPM files – patch conventionality back into the existing products

§ Conventionalization - Conventions get built in the product – they come out and work with all the libraries

o   As project goes on, the reprocessing goes down – it I an opportunity to fix file formats

o   Important task of migrating from HDF4 to HDF5 – been around for 15 yr… need to revisit this

o   Built a translator at HDF… went to NASA… then it will be expensive… then urgency is falling off… but becoming more urgent

o   Migration problems are really social problems not technical problems – how do we outreach to the

science teams

o   Can't access anything without conventions – work in broad industries – all over science

o   Think in vertical markets – ex Earth Science – need to spread across the disciplines

o   Have a proposal to work with Basic Plasma Science center – thinking about OPeNDAP... Andrew Collett, it guys there, and HDF will work together

o   Long running experiments – SWMR – run for 2 or 3 days – don't have a way to check if it is working – have 23 TB of HDF – tool = IDL... same as a lot of earth science

o   Hoping to get NSF funding

o   Plasma lab and Andrew's lab (dusty) have their own conventions around HDF5. They have a big facility and they have to share data

o   Also, Andrew & python – see more focus on python in HDF – want to achieve that without losing focus on other languages (esp JAVA)

o   Q – several people are interested in REST api – look into idea of presenting HDF5 data through REST interface. Bought both books by o'reily had a complete example with a rest service with a coupled rest client.

o   There are too many – better if we would consolidate on a single REST api

o   Have you seen the paper on the single REST api

o   ACTION – send Joe the paper

o   When people look on the internet – surprising how few have a complete but small REST server with a coupled REST client

o   Q – what are you looking for in a rest service – just pull a server side and full a service

§ Is that the only way to do it... do I have to install opendap to have a restful server

o   Problem is how to transfer the binary data?

o   Not much easier than pydap – THREDDS requires Tomcat – pydap only needs a pulse

o   Product Designer will include RESTful interface – write into a server somewhere

o   Exciting is 3 new people – Aleksander Jelenak from NOAA (satellite guy), John Ready, Joe Lee - also just won a ESIP funding Friday -

Session Leads:

Name: Ted Habermann [6]
Organization(s): The HDF Group [7]
Email: thabermann@hdfgroup.org [8]

Name: Mike Folk [9]
Organization(s): The HDF Group [7]
Email: mfolk@hdfgroup.org [10]

Participants:
online: David Coyle, Ellen Johnson

inroom: kelly monteleone, ted habermann, bob simons, h ramaphiyan, aleksandar jelenak, stephan

berrick, jeff walter, grace peng, ward flier, ajay krishnan, joe lee, zaihua J, joel plutchak, james gallagher, ken casy, john readey, stephan klene, joe glassy, roland schweitzer

Creative Common License: Creative Commons Attribution 3.0 License
Accepted:

Source URL: http://commons.esipfed.org/node/2566

Links:
[1] http://commons.esipfed.org/node/2566
[2] http://commons.esipfed.org/2014SummerMeeting
[3] http://commons.esipfed.org/session-type/breakout
[4] http://commons.esipfed.org/taxonomy/term/260
[5] http://www.hdfeos.net/
[6] http://commons.esipfed.org/node/1847
[7] http://commons.esipfed.org/taxonomy/term/822
[8] mailto:thabermann@hdfgroup.org
[9] http://commons.esipfed.org/node/1679
[10] mailto:mfolk@hdfgroup.org