Hera

astrophysical analysis over internet

Volodymyr Savchenko

SVOM, Strasbourg, October 2015

Standard way of running analysis

- 1. download software (~many Gb), dependencies, compile, setup environment
- 2. download IC, CalDB, (~many Gb), models, most of it probably is not used
- 3. select data to analyse
- 4. download data possibly next-to-raw, so much of them will be discarded
- 5. perhaps design parallel execution on a cluster
- 6. run your high level scientific analysis scripts

Hera - Astronomical Data Analysis over the Internet

1. download small (<10Mb) interface tool

2. ..

3. select data to analyse

4. ...

5. ...

6. run your high level scientific analysis scripts

the command line interface of Hera

the interface tool is familiar "FV" - part of HEASoft, can be downloaded separately

\$ fv -r fsort infile=intable.fits outfile=temp.fits method=insert

commands can be organized in scripts:

ftcopy 'rate.fit[col rate]' %1 ftcopy '%1[rate > 100]' %2 ftlist %1 hc outfile=ftlist.rate.out ftlist %2 hck outfile=ftlist.rate_gt_100.out

\$ fv -r runsc script_file

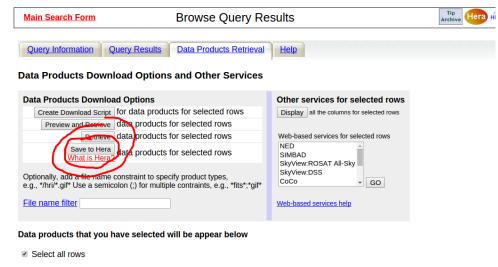
Hera is HEASARC

Any query can be used to save data to HERA instead of downloading.

It can be accessed and reduced near the data source, saving transfer time.

Some local space is provided for each user.

familiar results of HEASARC query



INTEGRAL Science Window Data

	scw id	scw ver	scw type	status	data in heasarc	ra	dec	good spi	good isgri	good picsit	good jemx1	good jemx2	good om
1	042200810021	001	slew	public	Y	05 34 31.95	+22 00 52.1	81	10	8	0	0	0
1	003900050010	001	pointing	public	Y	05 34 31.96	+22 00 52.4	0	3443	2322	8998	8998	8969
1	003900050020	001	pointing	public	Y	05 34 31.96	+22 00 52.4	0	3374	2202	8999	9000	9000
1	003900050030	001	pointing	public	Y	05 34 31.96	+22 00 52.4	0	3398	2307	8999	9000	9000
1	003900050040	001	pointing	public	Y	05 34 31.96	+22 00 52.4	0	3344	2311	8999	9000	9000
1	003900050050	001	pointing	public	Y	05 34 31.96	+22 00 52.4	0	3352	2170	9000	9000	9000
1	003900050060	001	pointing	public	Y	05 34 31.96	+22 00 52.4	0	3335	2209	9000	8999	9000
1	003900050070	001	pointing	public	Y	05 34 31.96	+22 00 52.4	0	3431	2198	8997	8990	9000
1	003900050080	001	pointing	public	Y	05 34 31.96	+22 00 52.4	0	3301	1908	5719	5727	9000
1	003900050090	001	pointing	public	Y	05 34 31.96	+22 00 52.4	0	142	62	0	0	1535
1	003900050100	001	pointing	public	Y	05 34 31.96	+22 00 52.4	0	0	0	0	0	0

http://heasarc.gsfc.nasa.gov/webHera/

the Web interface

web interface gives alternative access to the same tools

they have installed and keep up-to-date a lot of common tools, consider effort of installing all them locally?..

XMM-SAS, Chandra, Fermi, all Ftools and more

mage	- 🛃 👔	🕆 🔶	Name	Size	Date	Actions				
tils		4220086001	0.001	84Kb	09/18/2015 08:15					
asarc atools 1e	s 📄	ample.image	e.fits	276,480	04/25/2002 11:33	00 +5/				
onos ole	Choose a file to upload: Choose Files No file chosen Upload File use "Ctrl" or "Shift" key to select multi files									
SSION SCA/ titude/ nstein/	Tool Pa	arameter	Task Name: Get							
KOSAT/ ERMI/ ERMI/ EAOI/ EAOI/ ISTAR/ SO8/ Irrameter/ DSAT/ JZAKU/ Wiff/ ELA5B/ Innadu/ STAR/			vill be displayed here.							
TE/ andra IM-SAS ripting										

example of gbmrspgen

generates response for Fermi/GBM

requires a lot of calibration data, takes considerable time

the only way to access it is to use Hera

can be run in the GUI

can be integrated in the local scripts/pipeline

Hera vs Software and data distribution

security - private data is uploaded

if infrastructure is private for the project - no problem.

provider of software gets control of the software and some data can be used, but not fully disclosed (gbmrspgen)

network latency - constant network access is required

most of the data/calibration files are not transferred - only smaller reduced ones are

Hera vs Software and data distribution

control over software - impossible to install own software on there, apart from hera scripts

with command line interface one can run complicated local scripts/software calling remote tools when necessary

software provider get better control - can fix, update it more efficiently

many versions can be maintained for legacy and reproducibility

performance?

performance! you can "run" (run) hundreds of hera tasks from your laptop, they will be run on a remote cluster, available resources

Hera vs Software and data distribution: for developer

need to develop **software**, **interface**, **and batch system** to handle users possibly breaking it all completely

don't need to develop and maintain a lot of offline software, easier to monitor issues and update online infrastructure

Hera vs Computing center with shared filesystem

In limited collaboration (~<100 users?) one can use a large computing center as shared resources: like CC-IN2P3 Lyon

Hera is a computing center with shared software and data

The difference is that it provides not a linux shell but a set of interfaces with benefits of:

- security: users are less likely to break shared system
- integration with data exploration web interfaces

Conclusions

- Hera itself is public, but is managed by HEASARC and supported for a selection of missions of interest - not directly useful for private missions - but the approach might be useful
- Command-line delegating of some demanding tasks, integrating with local pipeline and scripts.
- Web interface provides nice way to select the data and run the reduction on it before downloading and processing the rest locally