Creating and managing very large HiPS: the Pan-STARRS case

Thomas Boch,[1] and Pierre Fernique[2]

[1]*CNRS, Observatoire de Strasbourg, Strasbourg, France;*
thomas.boch@astro.unistra.fr

[2]*CNRS, Observatoire de Strasbourg, Strasbourg, France;*
pierre.fernique@astro.unistra.fr

Abstract.
HiPS (Hierarchical Progressive Surveys) is a proven Virtual Observatory standard which enables an efficient way to deliver easily potentially huge images collection and allows for fast visualisation, exploration and science applications. CDS has recently published the HiPS for Pan-STARRS g and z-bands images, covering three quarter of the sky at a resolution of 250mas per pixel. We will describe in this paper the challenges we faced and the lessons learnt in generating and distributing these HiPS made of 47 million FITS tiles, amounting to 10 trillion pixels and more than 20TB per band. In particular, we will detail the methods we developed to optimize the generation, the storage and the transfer of the HiPS. In addition, a color HiPS, based on the two already available HiPS, has been made available and can be visualised from HiPS clients, like Aladin Desktop (Bonnarel et al. 2000) or Aladin Lite (Boch & Fernique 2014).

Pan-STARRS survey key figures

The Pan-STARRS PS1 survey covers three quarter of the sky in five photometric bands: g, r, i, z and y. Original images come as RICE compressed FITS files whose pixel size is 250 mas.
Each band is made up of 200,000 images for a total size of 15 TB.

In the next sections, we will cover the different steps allowing one the creation of the HiPS (**?**) for one Pan-STARRS band images.

1. Original files download

Downloading of images from the Space Telescope Science with a single wget was not sufficient to take full advantage of the network link between STScI and CDS: the transfer rate was only 12 MB/s. Using several wget connections in parallel increased the transfer rate to 46MB/s in average. Total transfer time for one band went down from 12 to 3.5 days.

2. FITS tiles generation

Pan-STARRS images come as RICE_1 tile-compressed FITS files. In our initial tests, FITS images had first to be uncompressed launching parallel instances of funpack. The Hipsgen tool has been upgraded to support RICE_1 images as input. This improvement saved us 4.5 days and has brought down the FITS tiles generation time to 20 days (on a 5 years old server with 128 GB RAM and 32 hyper-threaded cores)

3. JPEG tiles generation

Generation of JPEG tiles is a 20 days long process and result to generate 6TB of additional HiPS data. We created a Python service, based on the Falcon framework [1] and using astropy.visualization (Astropy Collaboration 2018) library for the heavy-lifting, which generates on-the-fly JPEG tiles from the existing FITS tiles. Thanks to Apache rewriting rules, access to those JPEG tiles is totally transparent for HiPS clients.

4. Transfer to production server

Our initial transfer process, based on rsync solely, was slow, averaging 12MB/s (1TB/day) on an internal Gigabit link. using parsync (a parallel rsync wrapper), we increased the transfer rate to 100 MB/s. The transfer time has been reduced from 20 days to less than 3 days.

5. RGB color tiles generation

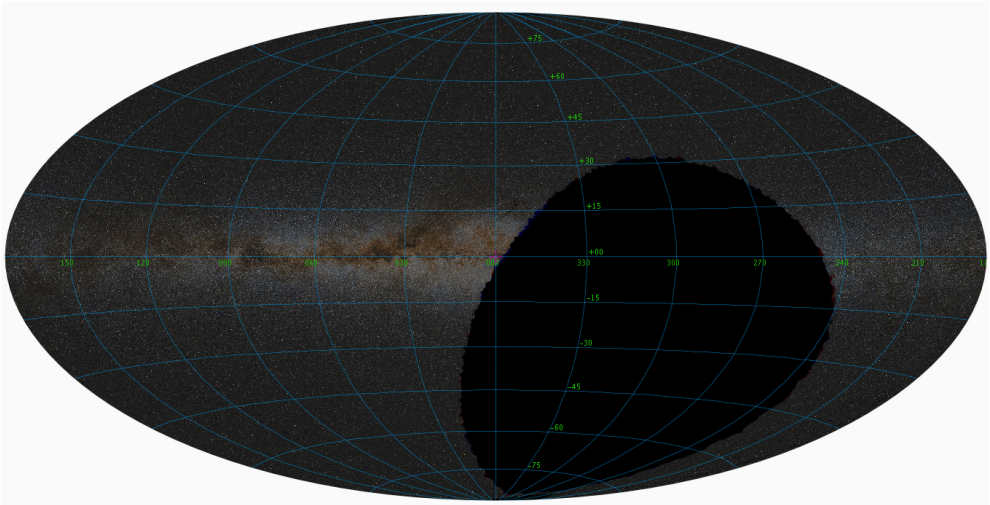

Figure 1. Color HiPS for PanSTARRS data, generated from z and g bands

[1] https://falconframework.org/

Red: z band, Green: 0.5 * (z+g), Blue: g band. Python script to generate color JPEG tiles from FITS tiles, with Lupton-like arcsinh stretch to maximize contrast. The figure /refaitoff above is this generated color HiPS visualised in Aladin Desktop.

Pan-STARRS HiPS key figures

As of today, three Pan-STARRS bands (g, z and y) are available as HiPS, described and available from the HiPS network.
Each HiPS has been created with a resolution of 200 mas (HEALPix order 20), slightly oversampled from the original 250 mas resolution. This represents 10 trillion pixels per band, divided into 47 million FITS tiles, and amounting to 25 TB.

Conclusion and links

Total HiPS generation time (from original FITS download to release) of one Pan-STARRS band has been reduced from 80 to 30 days. This new streamlined process has been put into practice for the creation of the z band now available in the HiPS network and will be applied to the generation of the remaining r, i and y bands, but also to other large image surveys we will process in the future. It also provides us with additional flexibility regardingthe creation of JPEG tiles. Creation of HiPS tiles on the server-side leads the way to some creative science-driven usage of the HiPS standard with minimal changes to the client applications.

References

Astropy Collaboration 2018, AJ, 156, 123. 1801.02634

Boch, T., & Fernique, P. 2014, in Astronomical Data Anaylsis Software and Systems XXIII, edited by N. Manset, & P. Forshay, vol. 485 of Astronomical Society of the Pacific Conference Series, 277

Bonnarel, F., Fernique, P., Bienaymé, O., Egret, D., Genova, F., Louys, M., Ochsenbein, F., Wenger, M., & Bartlett, J. G. 2000, Astronomy and Astrophysics Supplement Series, 143, 33