Enabling Large-Scale Dataset Analytics: Hierarchical Adaptive Tiling Scheme (HATS)

Mario Juric
DiRAC Institute Director | LINCC FW
Professor of Astronomy, University of Washington

with Wilson Bebe, Doug Branton, Sandro Campos, Neven Caplar, Melissa DeLucchi, Derek Jones, Jeremy Kubica, Kostya Malanchev, Sean McGuire, Colin Slater, Steven Stetzler, Max West, Sam Wyatt

and the LINCC Frameworks Analytics Group

DiRAC
Data Intensive Research in Astrophysics and Cosmology

LINCC

NSF
AST-2003196

The LINCC Frameworks Project

LSST Interdisciplinary Network For Collaboration And Computing

To collaboratively develop open computing systems and algorithms needed for large-survey analyses.

LSST Discovery Alliance

SCHMIDT FUTURES

Two LINCC-FW hubs:

- Carnegie Mellon University
- University of Washington

The Legacy Survey of Space and Time

Deep synoptic optical survey, coming in 2025.

Repeated imaging of the visible sky to ~24th mag.
10 years of operation.
60 PB of raw data.
40 billion stars, galaxies, asteroids.
30 trillion observations.

Objective

Develop a system (format, tools) to deliver and enable end-user analyses on 10TB+ (catalog) datasets

Large-dataset Analytics: Partitioned Files

- Relational databases are not ideal for this type of work. Poor UX, too many bottlenecks.
- Industry state of-the-art is to use distributed analytics tools operating on files.
- Distributed computation achieved through partitioning.

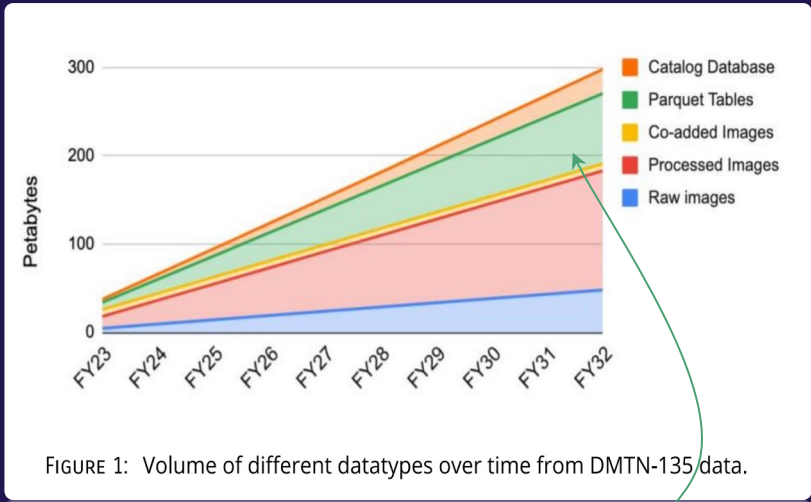
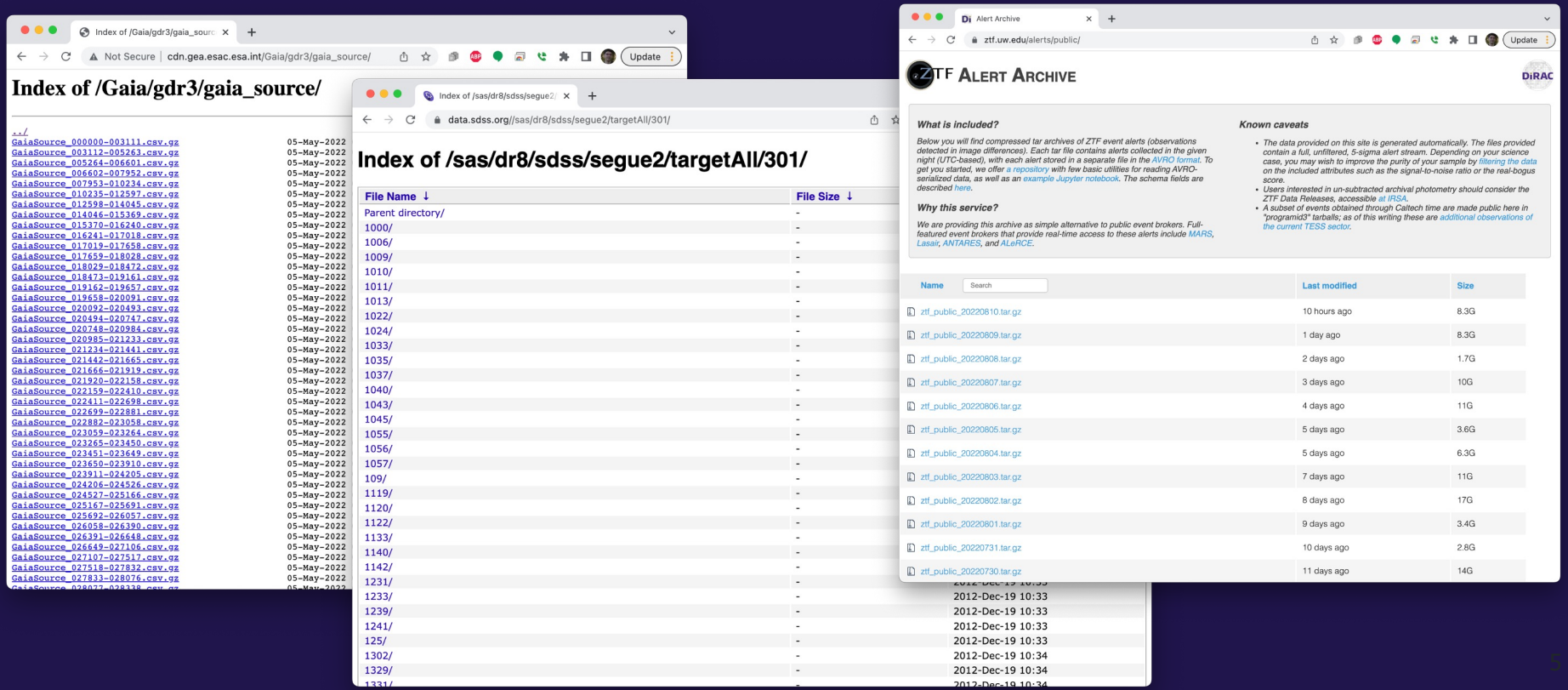


FIGURE 1: Volume of different datatypes over time from DMTN-135 data.

```
import pandas as pd
pd.read_parquet('example_pa.parquet', engine='pyarrow')
```

How to partition? Historically, we haven't generally given this much thought…



There’s value in thinking this through, and standardizing

- Users know what to expect and how to handle the dataset
- High quality, shared, analytics tools can be written
- Pre-staging/ETL may be avoided
- Multi-dataset analytics can supported

- Bulk export files == bulk analytics files
- Easier to generate and support for providers
- Can share code and infrastructure (e.g. mirroring, caching)

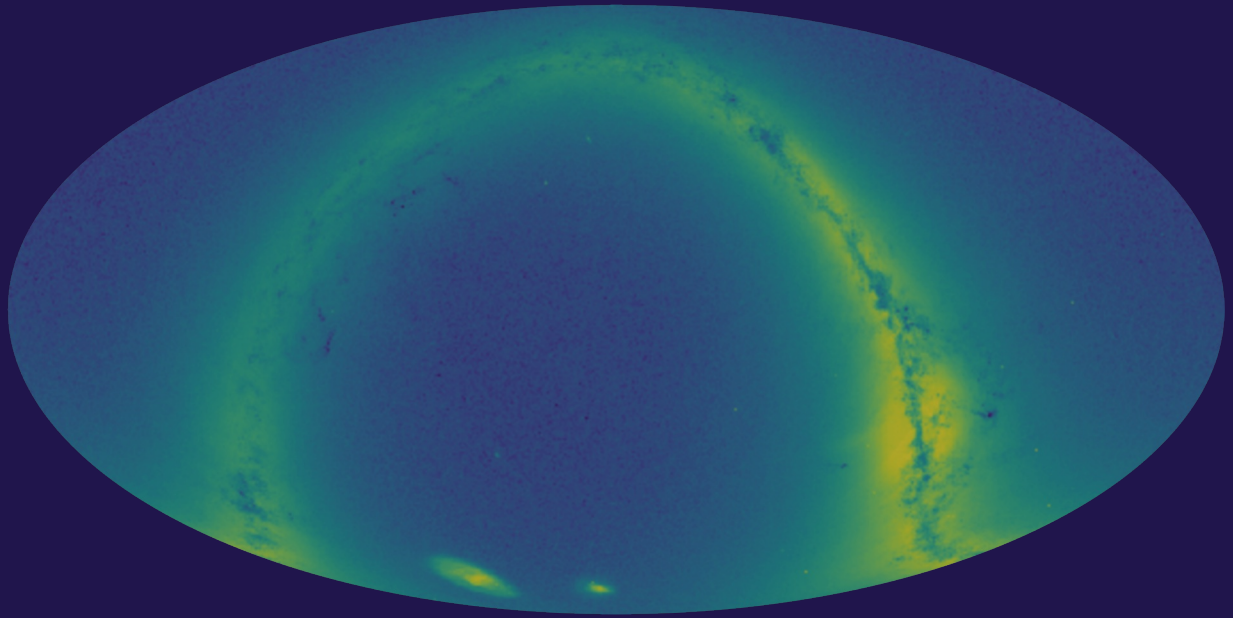
Think of all the wonderful tools and ecosystems that sprung up around HiPS!

Introducing HATS:
Hierarchical Adaptive Tiling Scheme

Gaia DR3

The 1.8Bn sources released in Gaia DR3

A single ASCII file would be about ~680GB in size, (gzip compressed).



1. Partitioning: HEALPix

Partition the sky into NSIDE=1 (order=0) HEALPix tiles, map tiles to files.

Example:

```
Norder0-Npix0.csv.gz
Norder0-Npix1.csv.gz
Norder0-Npix2.csv.gz
Norder0-Npix3.csv.gz
Norder0-Npix4.csv.gz
Norder0-Npix5.csv.gz
Norder0-Npix6.csv.gz
Norder0-Npix7.csv.gz
Norder0-Npix8.csv.gz
Norder0-Npix9.csv.gz
Norder0-Npix10.csv.gz
Norder0-Npix11.csv.gz
```

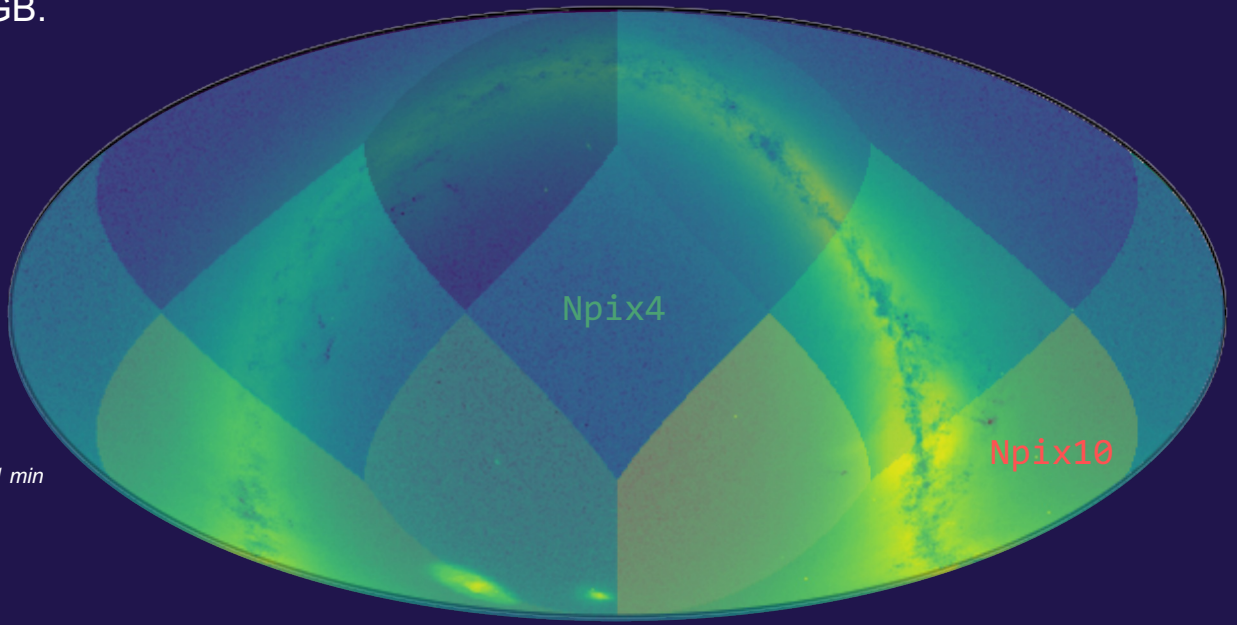
Problem: Severely unbalanced file sizes

Pixel 4 (Galactic pole) ~ 20GB
Pixel 10 (Galactic center) ~ 400GB.

Simple file-based parallelization fails.

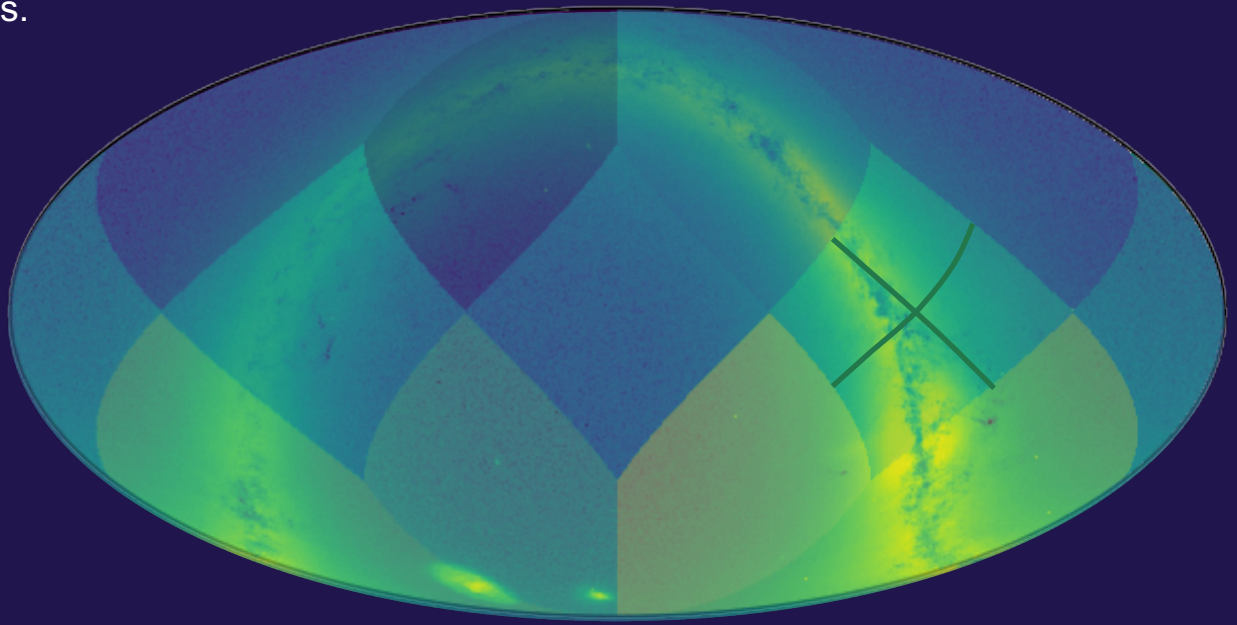
Example

```
Norder0-Npix0.csv.gz
Norder0-Npix1.csv.gz
Norder0-Npix2.csv.gz
Norder0-Npix3.csv.gz
Norder0-Npix4.csv.gz   → 1 min
Norder0-Npix5.csv.gz
Norder0-Npix6.csv.gz
Norder0-Npix7.csv.gz
Norder0-Npix8.csv.gz
Norder0-Npix9.csv.gz
Norder0-Npix10.csv.gz  → 20 min
Norder0-Npix11.csv.gz
```



Solution: Partition Hierarchically

If too many sources fall into a pixel,
split it into four higher order pixels.

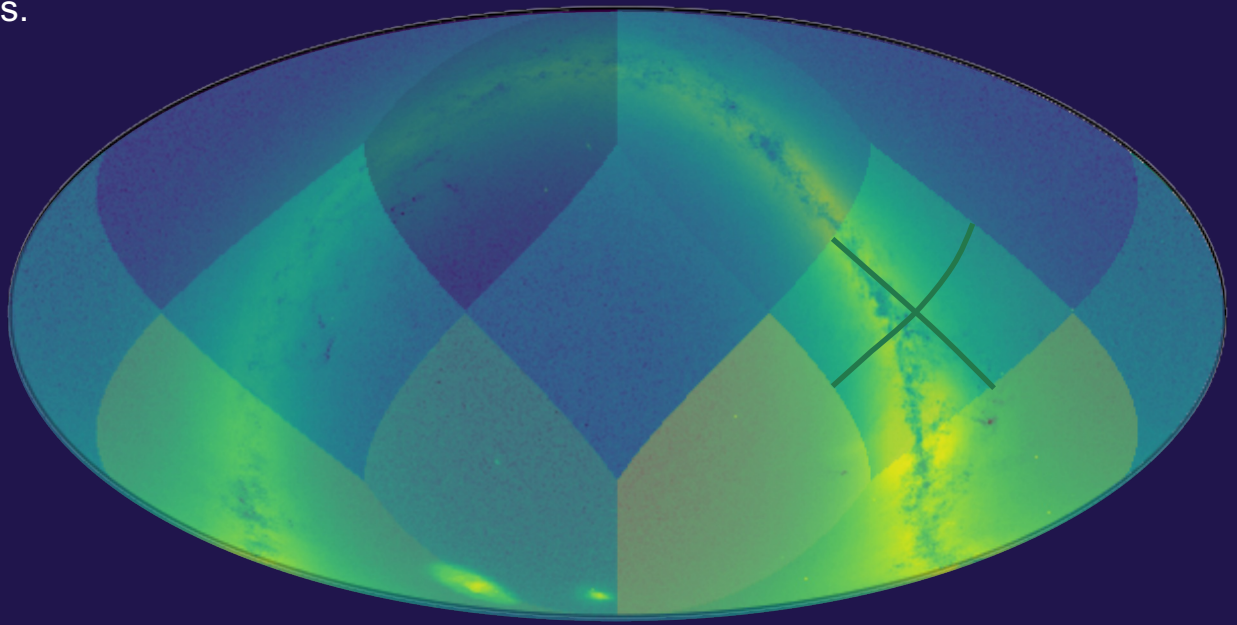


Example

```
Norder0-Npix0.csv.gz
…
Norder0-Npix7.csv.gz
…
Norder0-Npix11.csv.gz
```

1. Partition Hierarchically

If too many sources fall into a pixel,
split it into four higher order pixels.



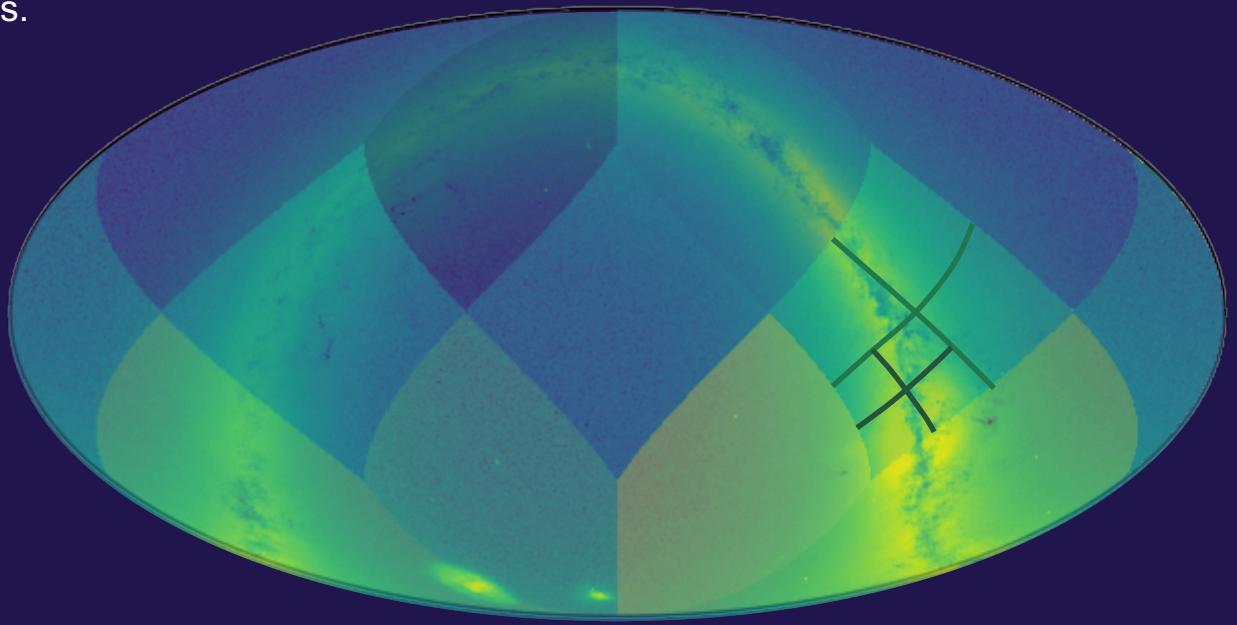
Example

```
Norder0-Npix0.csv.gz
…
Norder1-Npix28.csv.gz
Norder1-Npix29.csv.gz
Norder1-Npix30.csv.gz
Norder1-Npix31.csv.gz
…
Norder0-Npix11.csv.gz
```

1. Partition Hierarchically

If too many sources fall into a pixel,
split it into four higher order pixels.

Repeat.



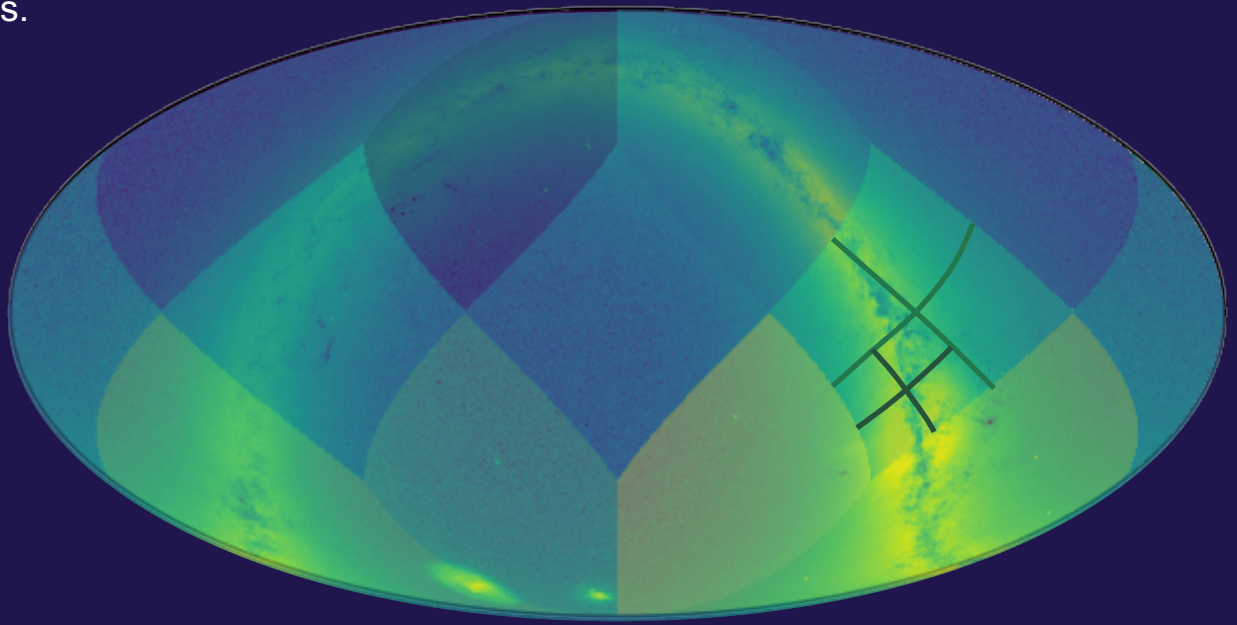
Example

```
Norder0-Npix0.csv.gz
…
Norder1-Npix28.csv.gz
Norder1-Npix29.csv.gz
Norder1-Npix30.csv.gz
Norder1-Npix31.csv.gz
…
Norder0-Npix11.csv.gz
```

1. Partition Hierarchically

If too many sources fall into a pixel, split it into four higher order pixels.

Repeat.



Example

```
Norder0-Npix0.csv.gz
…
Norder1-Npix28.csv.gz
Norder1-Npix29.csv.gz
Norder1-Npix30.csv.gz
Norder2-Npix112.csv.gz
Norder2-Npix113.csv.gz
Norder2-Npix114.csv.gz
Norder2-Npix115.csv.gz
…
Norder0-Npix11.csv.gz
```

1. Partition Hierarchically

If too many sources fall into a pixel, split it into four higher order pixels.

Repeat until each file size is beneath some pre-defined threshold.

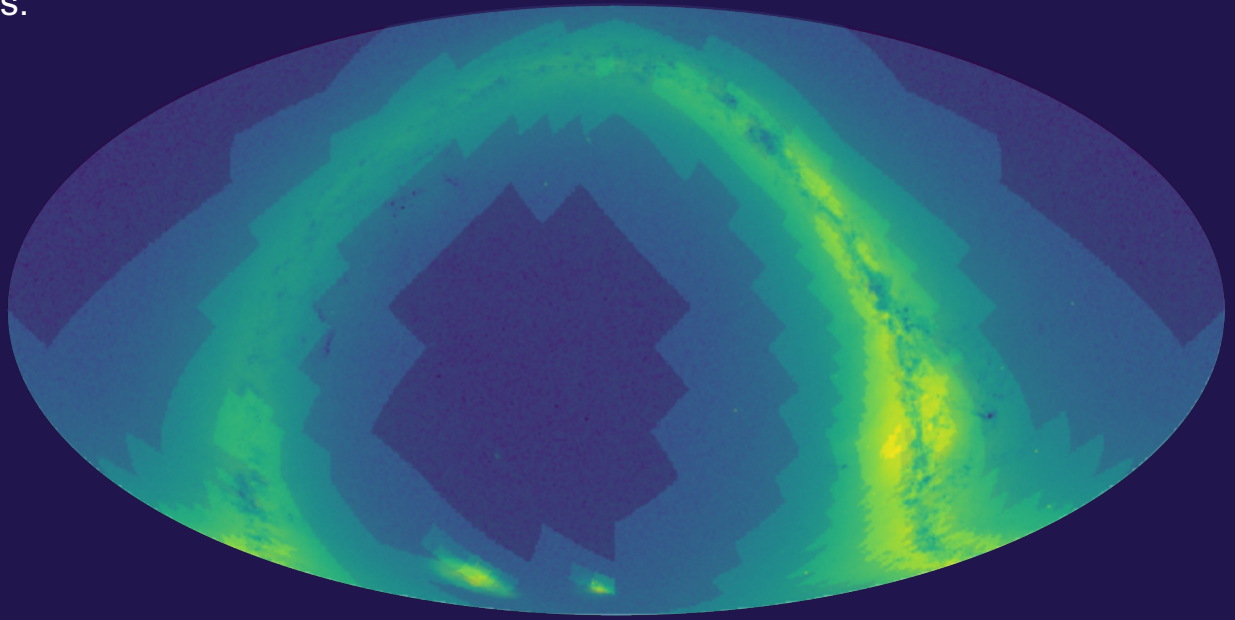


Figure: an overlay of Gaia counts and the partitioning map, taking MAXOBJECTS=1e6

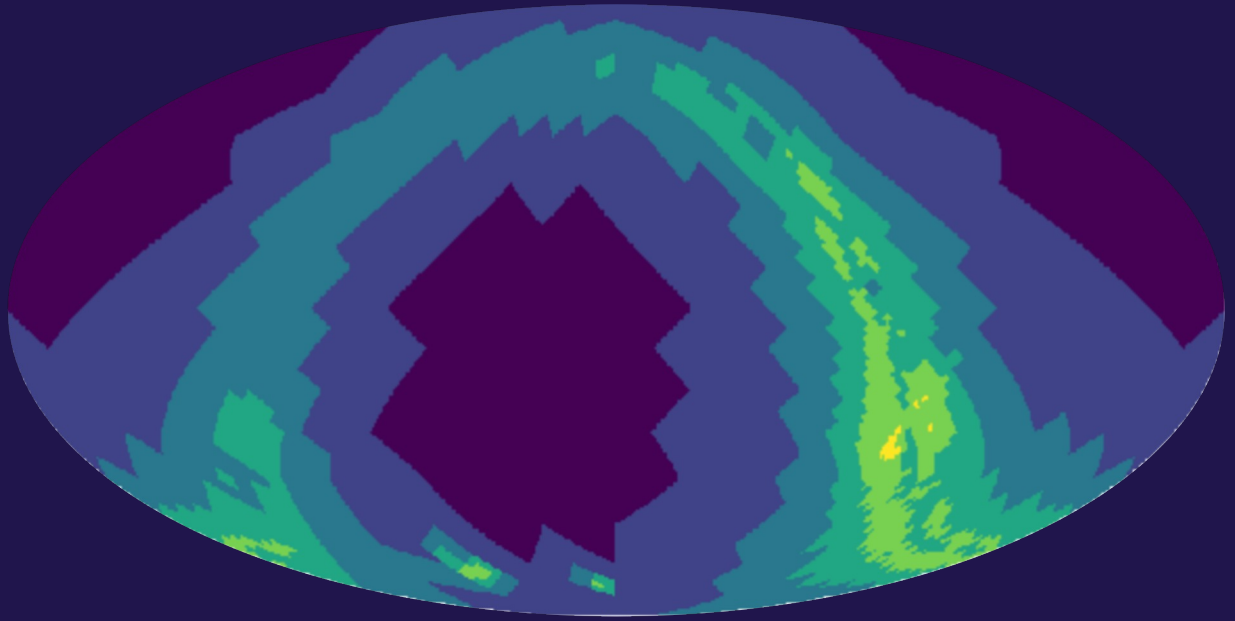
order:

3993 partitions (tiles) for Gaia DR3, with 1M object/partition threshold

2. Serialization: Parquet in HiPS-like Directory Structure

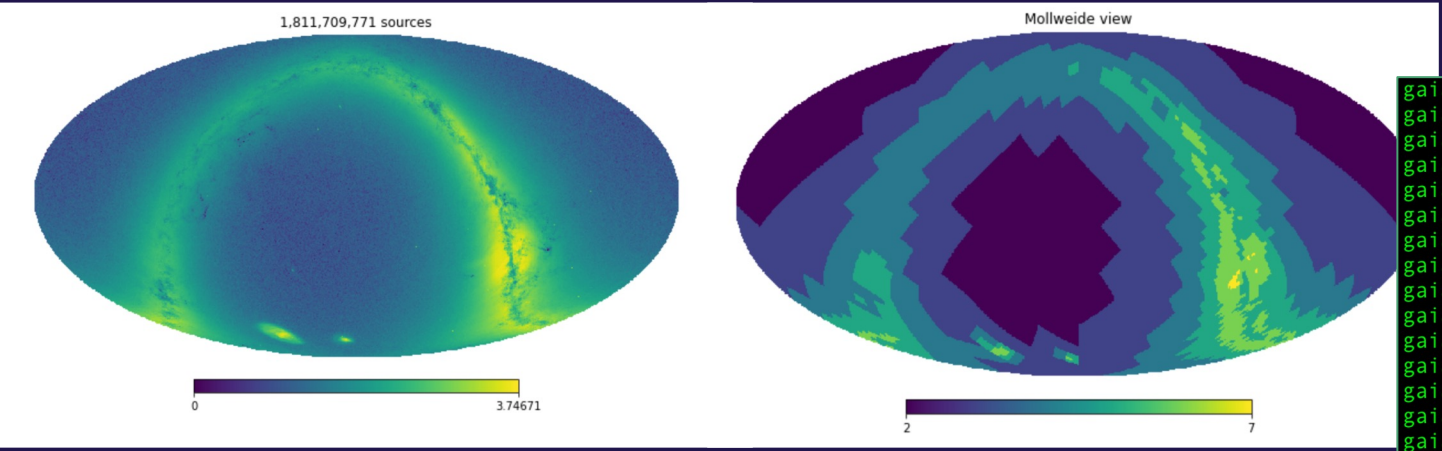
Store the partition files (tiles) into a directory tree.

Parquet files in the leaf nodes.



```
Norder=0/Dir=0/Npix=0.parquet
…
Norder=1/Dir=0/Npix=28.parquet
Norder=1/Dir=0/Npix=29.parquet
Norder=1/Dir=0/Npix=30.parquet
Norder=2/Dir=0/Npix=112.parquet
Norder=2/Dir=0/Npix=113.parquet
Norder=2/Dir=0/Npix=114.parquet
Norder=2/Dir=0/Npix=115.parquet
…
Norder=0/Dir=0/Npix=11.parquet
```

All together: Hierarchical Adaptive Tiling Scheme (HATS)



Gaia DR2 Catalog Counts (log scale)

Visualization of file storage (color = healpix level)
3933 partitions of similar size (128-256 MB)

Layout on "disk": 👇

```
gaia/dataset/_metadata
gaia/dataset/_common_metadata
gaia/properties
gaia/partition_info.csv
gaia/point_map.fits
gaia/dataset
gaia/dataset/Norder=2
gaia/dataset/Norder=2/Dir=0
gaia/dataset/Norder=2/Dir=0/Npix=20.parqu
gaia/dataset/Norder=2/Dir=0/Npix=22.parqu
gaia/dataset/Norder=2/Dir=0/Npix=21.parqu
gaia/dataset/Norder=2/Dir=0/Npix=23.parqu
gaia/dataset/Norder=2/Dir=0/Npix=17.parqu
gaia/dataset/Norder=2/Dir=0/Npix=0.parque
gaia/dataset/Norder=2/Dir=0/Npix=2.parque
gaia/dataset/Norder=2/Dir=0/Npix=19.parqu
gaia/dataset/Norder=2/Dir=0/Npix=28.parqu
gaia/dataset/Norder=2/Dir=0/Npix=29.parqu
gaia/dataset/Norder=2/Dir=0/Npix=34.parqu
gaia/dataset/Norder=2/Dir=0/Npix=40.parqu
gaia/dataset/Norder=2/Dir=0/Npix=35.parqu
gaia/dataset/Norder=2/Dir=0/Npix=42.parqu
gaia/dataset/Norder=2/Dir=0/Npix=41.parqu
gaia/dataset/Norder=2/Dir=0/Npix=43.parqu
gaia/dataset/Norder=2/Dir=0/Npix=44.parqu
gaia/dataset/Norder=2/Dir=0/Npix=46.parqu
gaia/dataset/Norder=2/Dir=0/Npix=64.parqu
gaia/dataset/Norder=2/Dir=0/Npix=65.parqu
```

Updates since Sydney

Format Updates

- Renaming: HiPSCat -> HATS
- Harmonize metadata file naming/contents conventions with HiPS
- Healpix index made non-unique
- Experimental append support
- VOTable metadata in Parquet

```
gaia/dataset/_metadata
gaia/dataset/_common_metadata
gaia/properties
gaia/partition_info.csv
gaia/point_map.fits
gaia/dataset
gaia/dataset/Norder=2
gaia/dataset/Norder=2/Dir=0
gaia/dataset/Norder=2/Dir=0/Npix=20.parquet
gaia/dataset/Norder=2/Dir=0/Npix=22.parquet
gaia/dataset/Norder=2/Dir=0/Npix=21.parquet
gaia/dataset/Norder=2/Dir=0/Npix=23.parquet
gaia/dataset/Norder=2/Dir=0/Npix=17.parquet
gaia/dataset/Norder=2/Dir=0/Npix=0.parquet
gaia/dataset/Norder=2/Dir=0/Npix=2.parquet
gaia/dataset/Norder=2/Dir=0/Npix=19.parquet
gaia/dataset/Norder=2/Dir=0/Npix=28.parquet
gaia/dataset/Norder=2/Dir=0/Npix=29.parquet
gaia/dataset/Norder=2/Dir=0/Npix=34.parquet
gaia/dataset/Norder=2/Dir=0/Npix=40.parquet
gaia/dataset/Norder=2/Dir=0/Npix=35.parquet
gaia/dataset/Norder=2/Dir=0/Npix=42.parquet
gaia/dataset/Norder=2/Dir=0/Npix=41.parquet
gaia/dataset/Norder=2/Dir=0/Npix=43.parquet
gaia/dataset/Norder=2/Dir=0/Npix=44.parquet
gaia/dataset/Norder=2/Dir=0/Npix=46.parquet
gaia/dataset/Norder=2/Dir=0/Npix=64.parquet
gaia/dataset/Norder=2/Dir=0/Npix=65.parquet
gaia/dataset/Norder=2/Dir=0/Npix=70.parquet
gaia/dataset/Norder=2/Dir=0/Npix=67.parquet
```

Code on GitHub/PyPI/conda-forge

Astronomy Data Commons

Software Infrastructure for Science Platforms and Scalable Astronomy on Cloud Resources

Pinned

Customize pins

axs Public

Astronomy eXtensions for Spark: Fast, Scalable, Analytics of Billion+ row catalogs

Python ☆ 23 12

hats Public

Hierarchical Adaptive Tiling Scheme

Python ☆ 17 5

Low level format routines

lsdb Public

Large Survey DataBase

Python ☆ 19 5

End-user analytics tool

hats-import Public

HATS import - generate HATS-partitioned catalogs

Python ☆ 6 5

Robust importer

https://github.com/astronomy-commons

All of these are also available on PyPI and conda-forge.

LSDB: Python Analytics for HATS

- LSDB: Large Survey Database
- Enable Pandas-like analysis on trillions of observations with thousands of cores
- Build on existing tools: Dask (looking at Ray).
- Full HATS awareness: spatial queries, cross-matching, timeseries, multi-dataset joining.
- Stabilizing the API

```
img = gaia
      .query(“pm > 10”)
      .crossmatch(ztf)
      .join(ztf_sources)
      .for_each(varstar_classify)
      .query(“pRRLy > 0.95”)
      .skymap()

hp.mollview(img)
```

LSDB target APIs: The API center science. Multi-processing, autoscaling, fail-over, etc. are all implicit. Good user experience.

Wyatt et al. (2023)
https://github.com/astronomy-commons/lsdb

Nested Pandas: Compact/lazy JOIN evaluation

- Typical catalogs are structured as Object tables, to be joined to Source/Detection tables (time series).

- SQL JOIN-ing Object to Source tends to duplicate fields from Object, and generates a flat table that's inconvenient for the end-user to work with

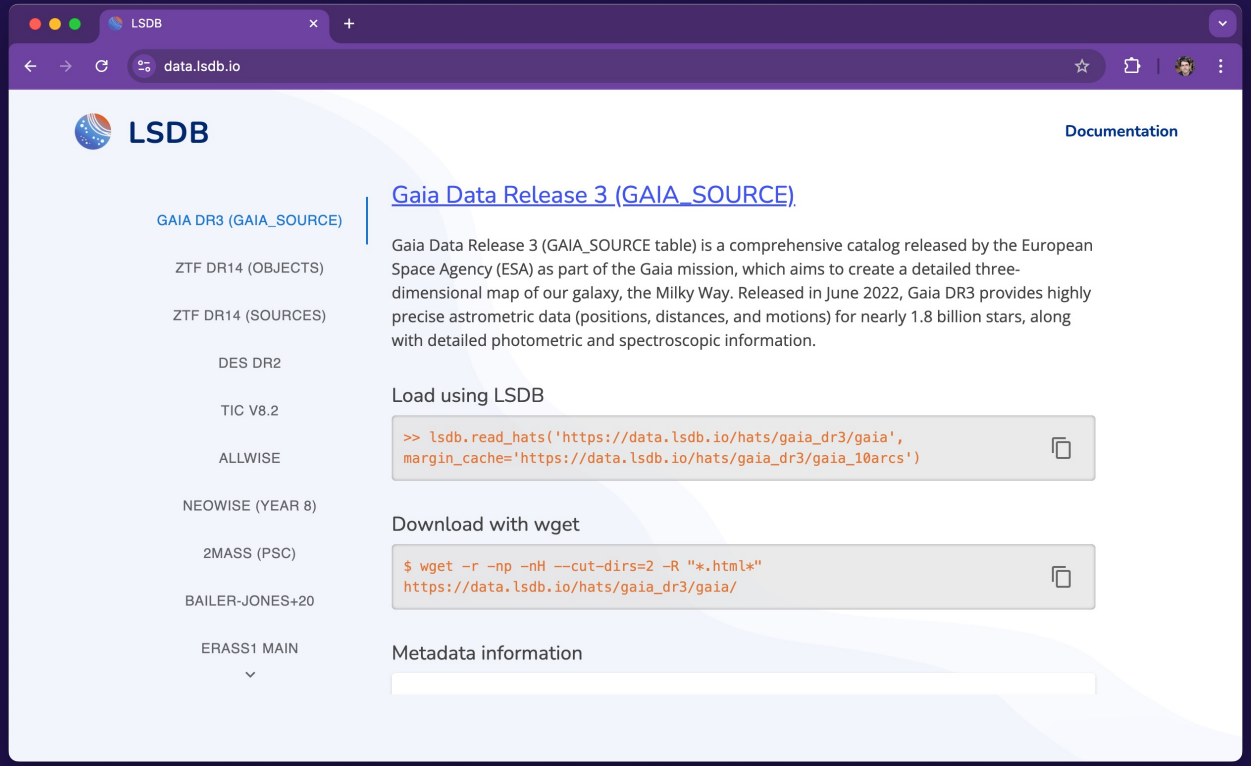
	a	b	nested
0	0.417022	0.184677	t flux band 0 8.383890 1...
1	0.720324	0.372520	t flux band 0 13.704390 1...
2	0.000114	0.691121	t flux band 0 4.089045 9...
3	0.302333	0.793535	t flux band 0 17.562349 6...
4	0.146756	1.077633	t flux band 0 0.547752 ...

- nested_pandas is a memory-efficient pandas extension that presents a "time-series" view to the user, while retaining the speed of evaluation equivalent to that of a flat table.

https://github.com/lincc-frameworks/nested-pandas

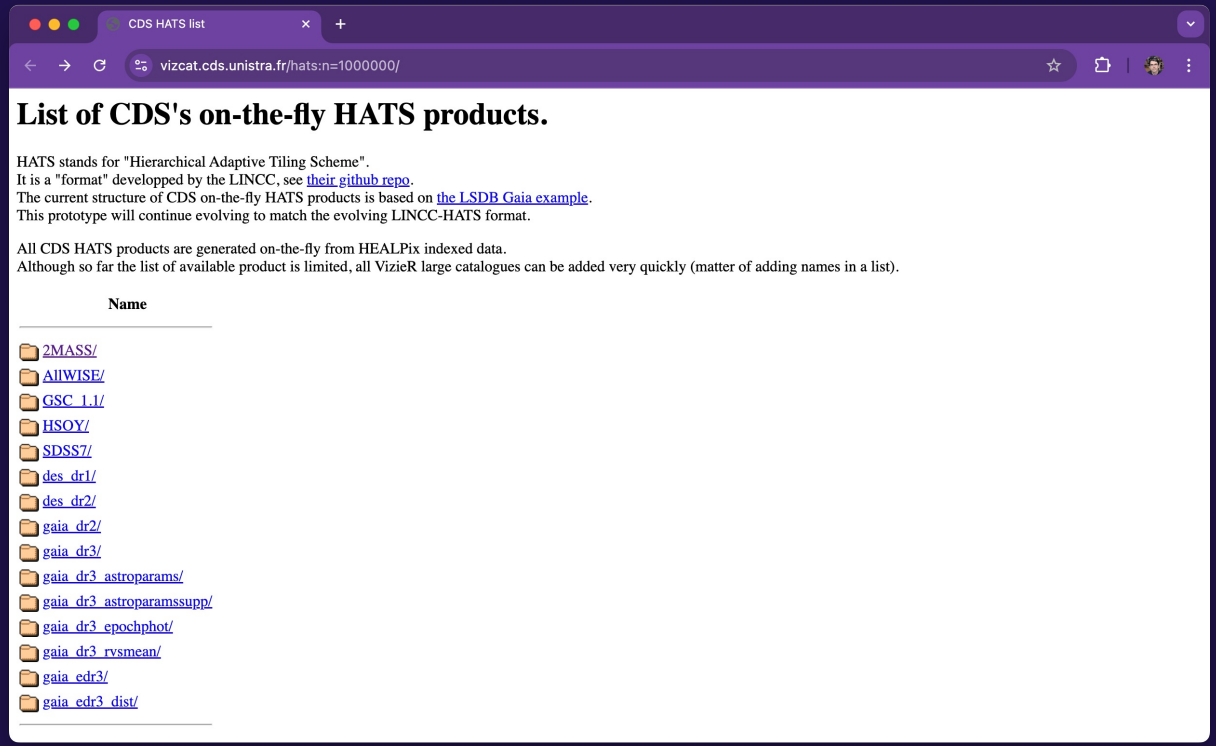
HATS Datasets @ https://data.lsdb.io

- A library of HATS-partitioned datasets to support testing and experimentation with the format
- Download LSDB (the main HATS analytics tool), point it to this website, and immediately start running queries.



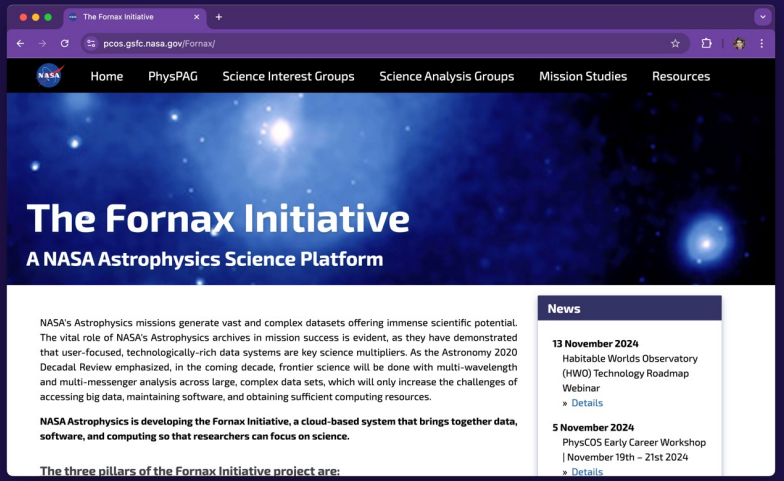
HATS Facade @ CDS

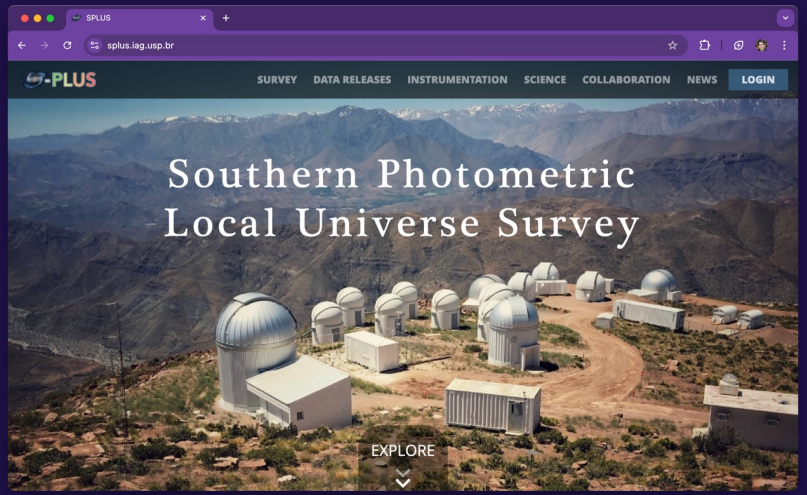
- An experiment by FX Pineau et al. to test if CDS catalog holdings can be on-the-fly represented as HATS.
- Converts internal CDS database representation of the catalogs into HATS-partitioned files, served over HTTP.



Fornax, SPLUS, …

- Continuing the integration and testing of HATS/LSDB within Fornax (IPAC & STScI Fornax teams)
- A “HATS/Parquet” server for SPLUS with server-size predicate evaluation (Gustavo Schwarz, Mackenzie U.)





Status and Looking Ahead

- Format is in a reasonably solid draft form now
  - Remains to be done: optimized storage of distributed cross-match margins
- Working or in touch with most major archives/projects.
- Preparing an IVOA Note.

- Application support is maturing for Python (LSDB), but we're looking to expand support in other ecosystems (Rust, Java)
  - E.g. a Spark DataSource for reading HATS would enable advanced support for HATS within Spark

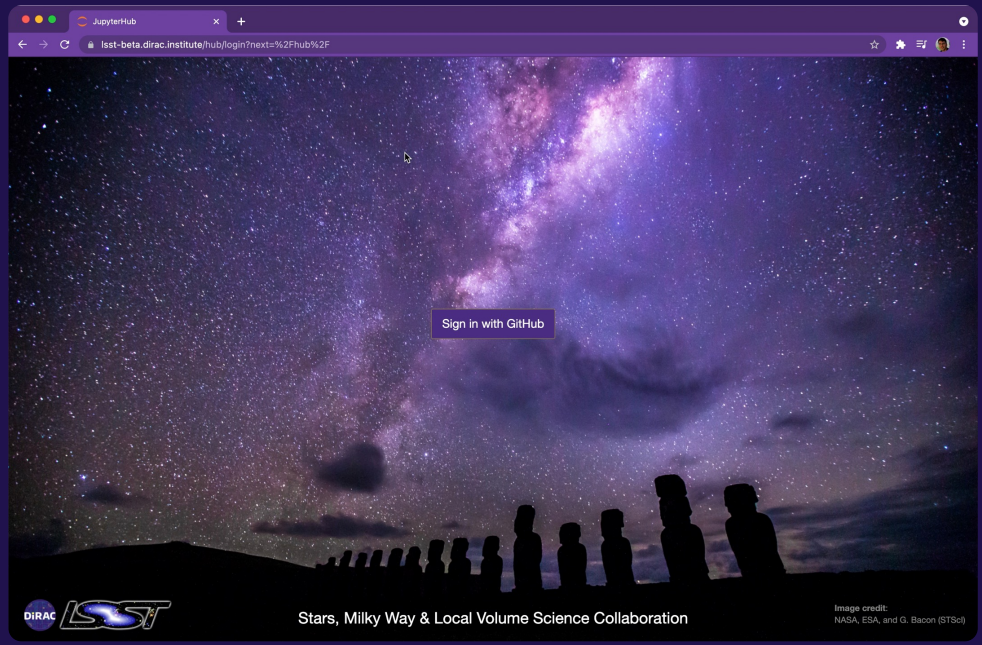
- Field testing within Rubin Commissioning
- Thinking about Iceberg serialization

Where to Try

LINCC Hub: https://lsst.dirac.dev

- LINCC Hub: *The LINCC Science Platform for the LSST Science Collaborations*
  (email ncaplar@uw.edu for access)

- NASA Astrophysics Science Platform: Working with the Fornax team at IRSA/MAST/HEASARC to deploy and test.



- Or run `conda install -c conda-forge lsdb` on your fav. science platform.

Summary and next steps

Mailing list: https://groups.google.com/g/hats-wg
Repositories: https://github.com/astronomy-commons
Meetings: *10am PT, every third Friday of the month*

- Aiming to enable end-user analyses on 1-100T+ catalog datasets. Ad-hoc collaboration of scientists/engineers from LINCC, Rubin, MAST, IRSA, HESARC, LINeA, CDS. Developing formats and tools.

- Lots of progress since last year. Data format and Python tools solidifying. We've gathered enough initial real-world usage to start drafting the formal format spec. Preparing an IVOA Note.

- Next year:
  - Perfect the tools (UX and performance are a priority)
  - Larger scale user deployment
  - Would love to get addt'l feedback and implementations (e.g. Java or Rust)


LINCC
Carnegie Mellon University
UNIVERSITY OF WASHINGTON
LVX SIT
1861
Collaboratively advancing data-intensive astronomy.
LSST Discovery Alliance
NSF
SCHMIDT FUTURES
AST-2003196
Thank You !
Contact: mjuric@uw.edu
UNIVERSITY of WASHINGTON