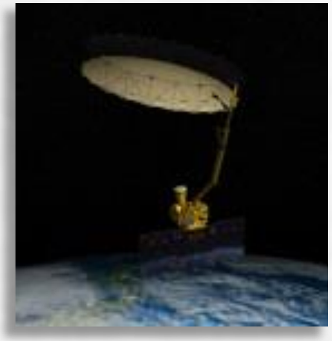
BIG DATA SPATIO-TEMPORAL ANALYSIS ON ASTRONOMICAL IMAGES

Francisco J. de la Mora García, Teruhisa Hochin, Hiroki Nomiya
Kyoto Institute of Technology, Kyoto, Japan.
Contact: fco.delamora@gmail.com

1. Introduction

2015
NASA's Soil Moisture Active Passive (SMAP)



Moist levels on the soil
~485 GB/day

2020
NASA-ISRO Synthetic Aperture Radar (NISAR)

Image Earth through radar
~86 TB/day

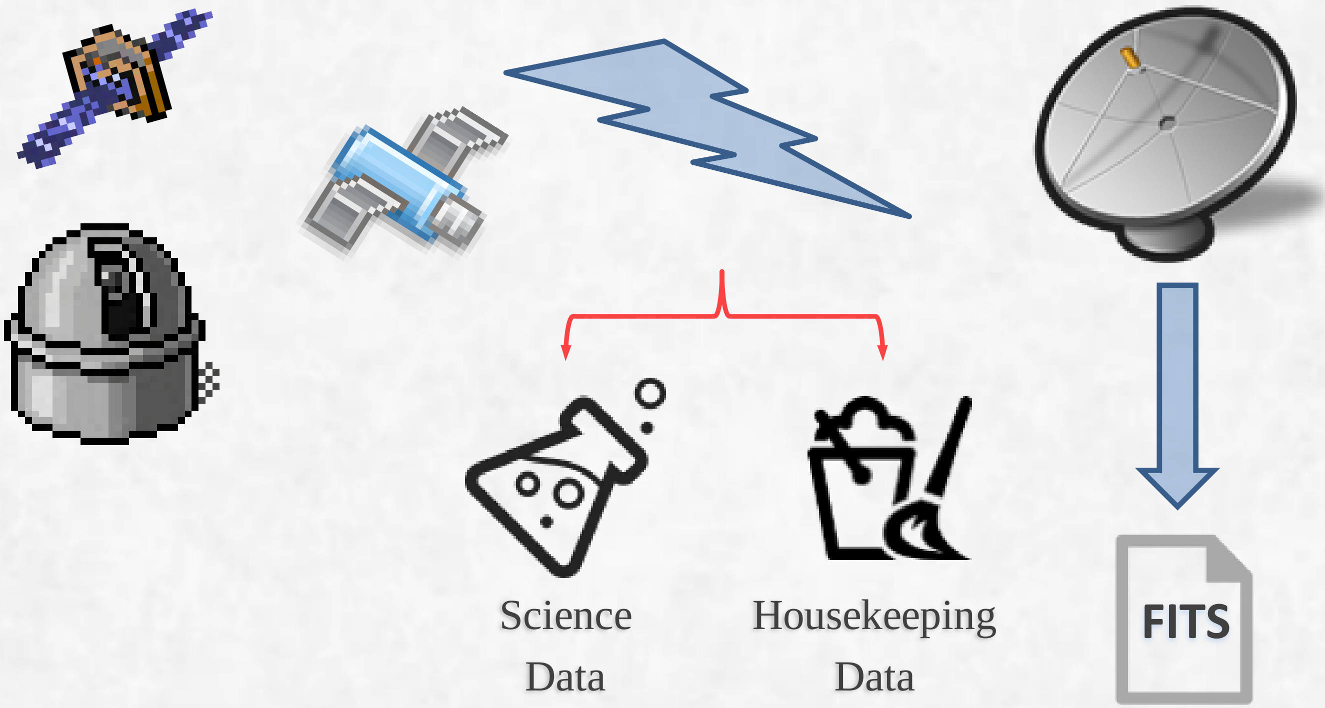
2030
Square Kilometer Array (SKA)

Radio Telescope
~700TB/second!

- Enormous rates of data require appropriate mechanisms to **handle, process, archive** and **search** through it as efficiently as possible.
- Demand for tools to identify celestial bodies hidden among thousands of files collected on different surveys.

2. Background

2.1 **Astronomical Data**



Science Data
- Spectra
- Images
- Count rates
- …

Housekeeping Data
- Instrument Health
- Security
- Pointing direction
- …

3. Research objective

3.1 **Moving faint object detection through image co-addition on spatio-temporal mosaics**

- Take advantage of cloud technologies to process large amounts of files used in Astronomy.
- Incorporate astronomical spatial data into the index.

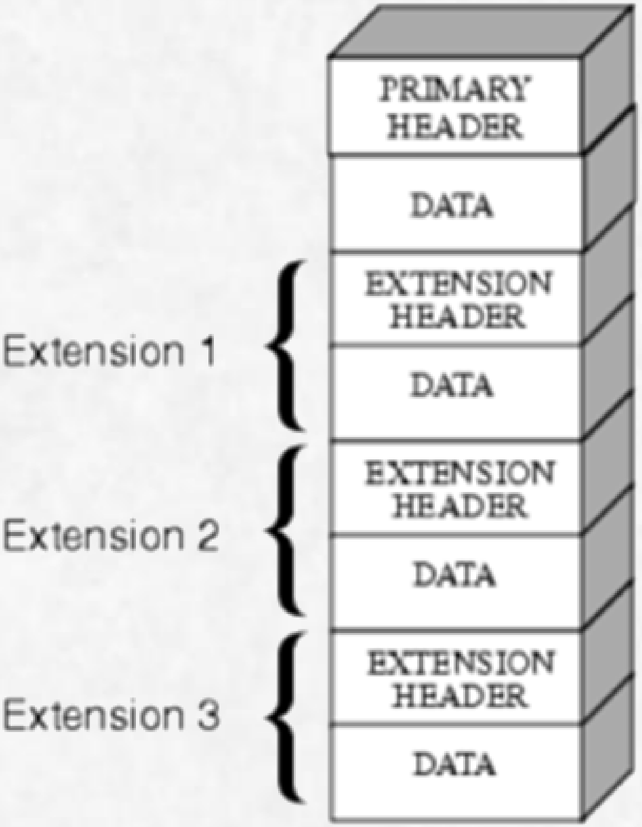
3.2 **Expected improvements**

- Reduce the computational requirements to process Big Data through the use of different nodes in the network.
- Reduction of cost to handle Big Data.

4. Related Works

4.1.1 **Flexible Image Transport System files (FITS Files)**

- Contains data related to the instrument location, pointing direction, etc… along with the image data.
- Instrument and transformation data is stored on pre-defined headers.
- Extensions can be used following the same structure.



4.1 Flexible Image Transport System files (FITS Files)

4.1.2 **FITS Files Header Format**

Any number of repetitions of **80 characters** in the general form:

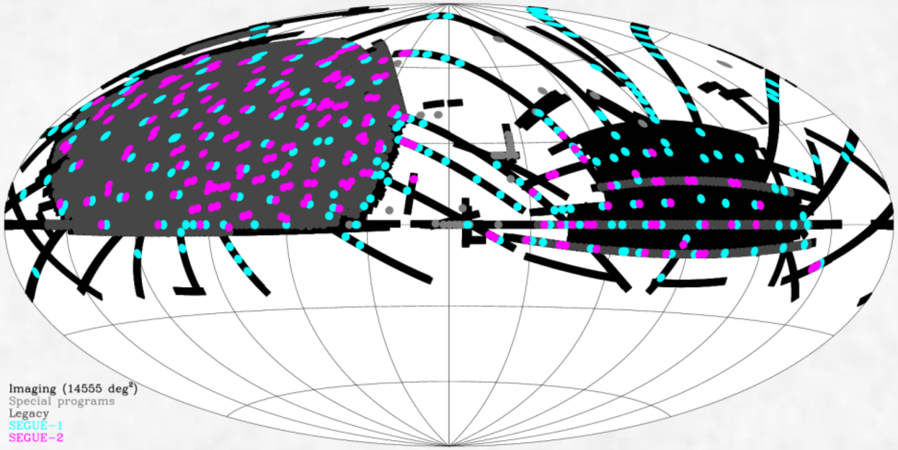
Keyword = Value / Comment

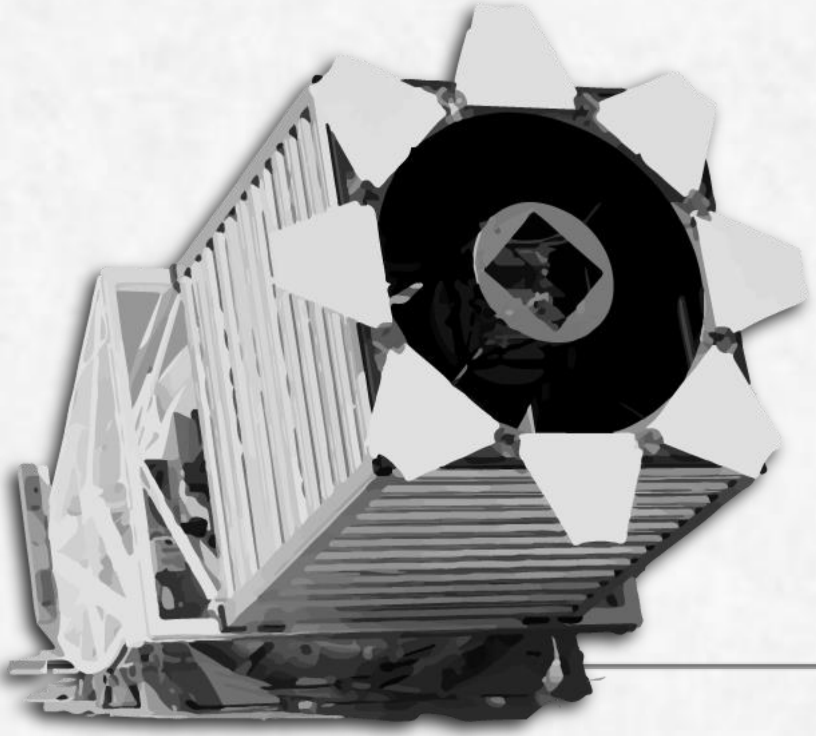
```
TEMP_0  =                 -26.44 / CCD TEMPERATURE IN DEGREES CELSIUS
UTC_0   = '2004-05-18T10:11:49.' /  UTC converted from TI_0
TARGET  = 'EARTH   '             /  Name of image target
```

4. Related Works

4.2.1 **The Sloan Digital Sky Survey (SDSS)**

Multi-filter imaging and spectra sky survey using a 2.5m wide angle optical telescope at New Mexico, USA.





- Sky regions (Stripes) are imaged multiple times.
- Provides imaging information as FITS files.

4. Related Works

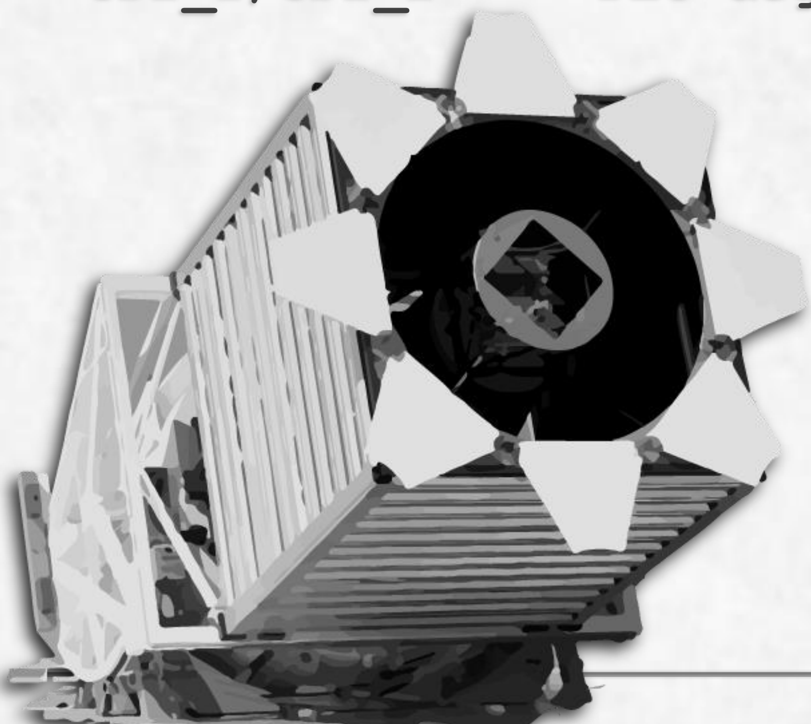
4.2.2 The Sloan Digital Sky Survey (SDSS) FITS Files

CRPIX1/CRPIX2 ← Column/Row Pixel coordinate of ref. pixel.

CRVAL1/CRVAL2 ← RA/DEC at reference pixel

CD1_1/CD1_2 ← RA degrees per Column/Row pixel.

CD2_1/CD2_2 ← DEC degrees per Column/Row pixel.



RADECSYS = 'ICRS   '

↑

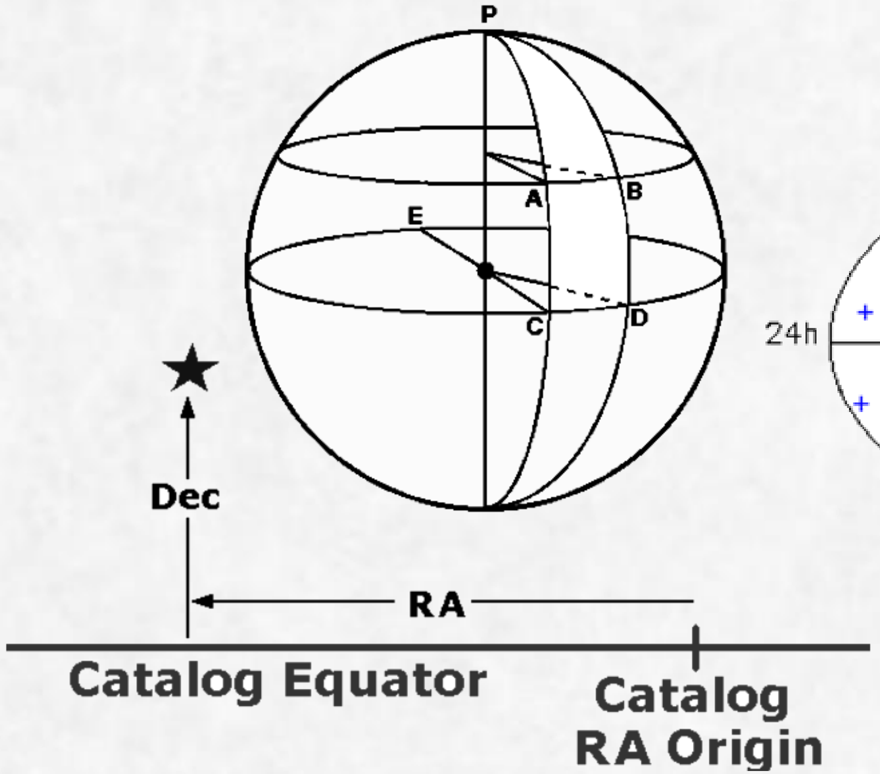
International Celestial Ref. System

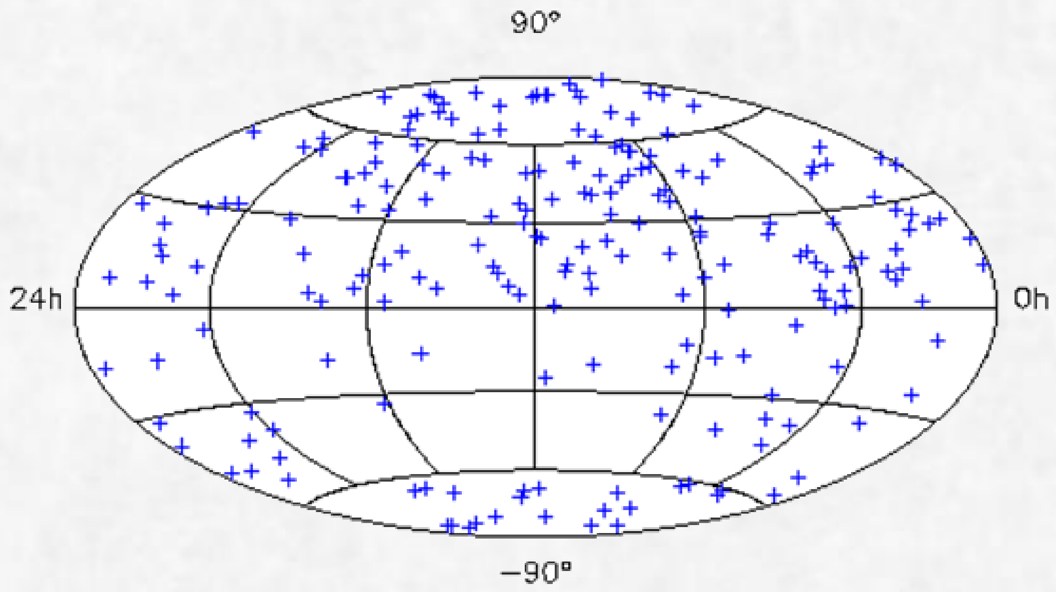
SDSS

4. Related Works

4.3 **Defining regions in the sky using the International Celestial Reference System （ICRS）**

Information on the FITS file headers can locate and define the regions to be processed by the system. This helps us define the mosaics.





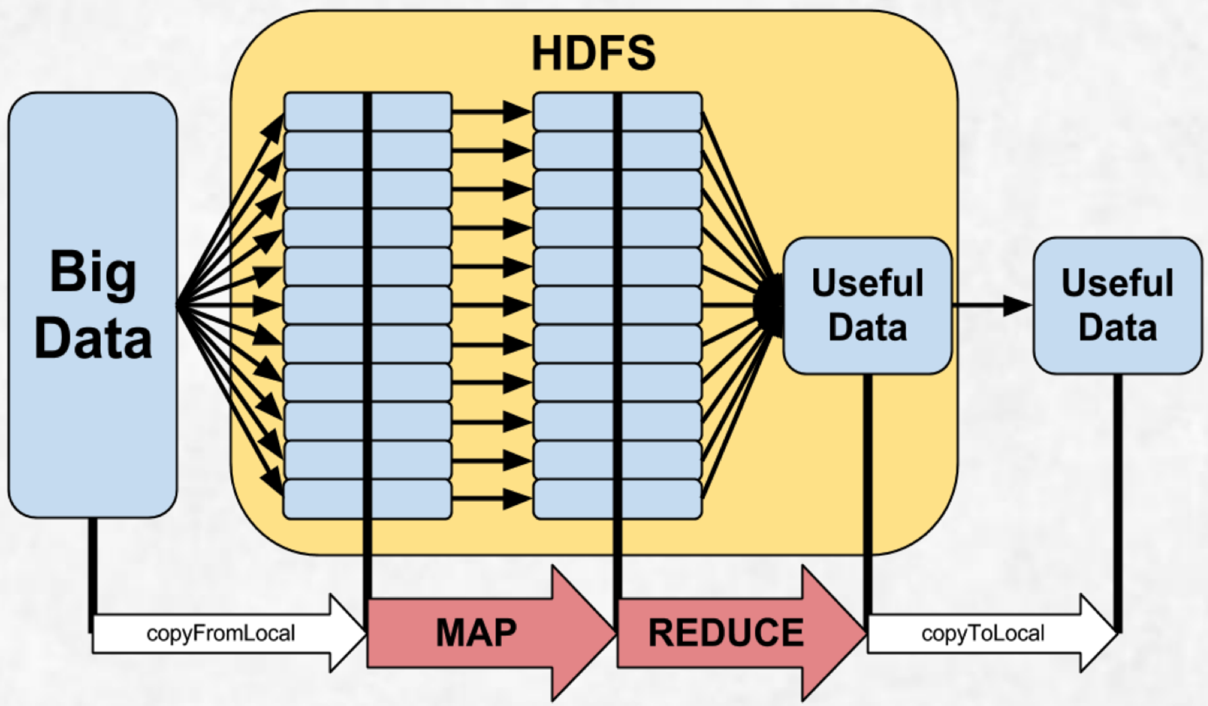
4. Related Works

4.4.1 **MapReduce**

- Two units of computation **mappers** and **reducers** which can be executed in parallel.

- Both units are programmed by the user.

- Output defined on a case by case manner.



4.4.1 MapReduce

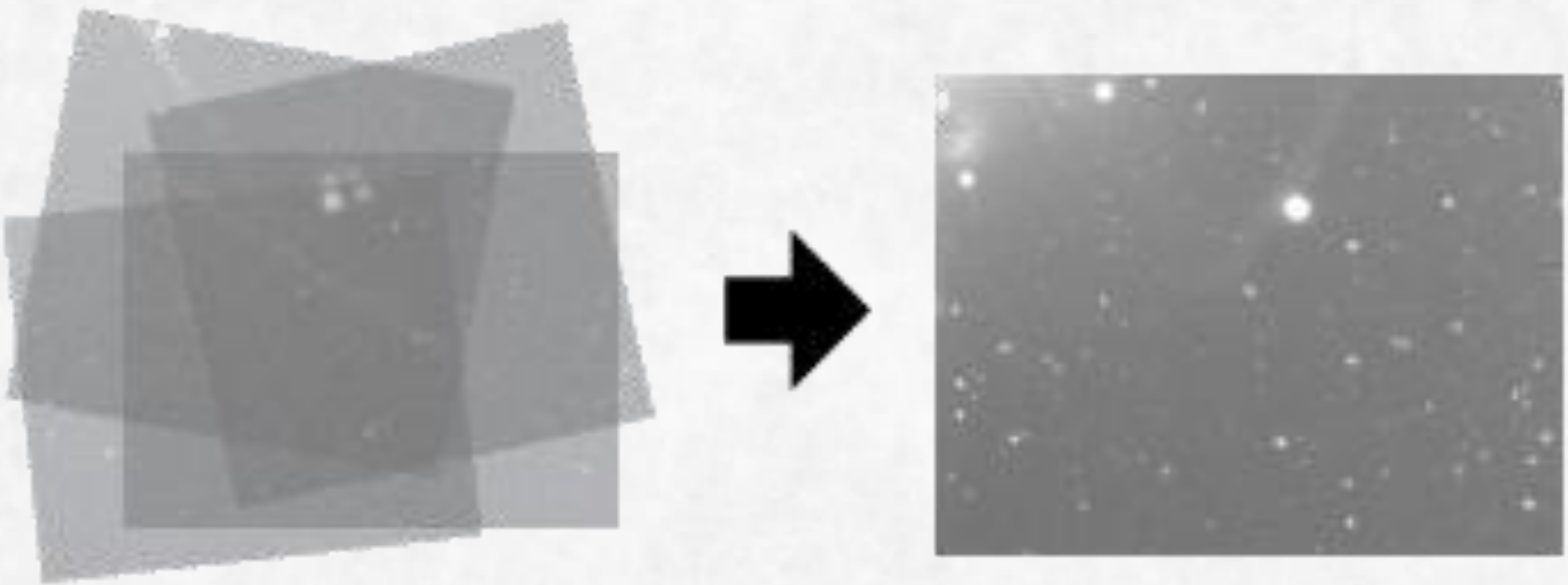
4.4.2 hadoop

- Open-source implementation of the MapReduce Framework.

- Allows distributed processing of large data sets across clusters of computers using a simple programming model.

4. Related Works

4.5 **Astronomy in the Cloud: Using MapReduce for Image Co-Addition (K. Wiley et al)**

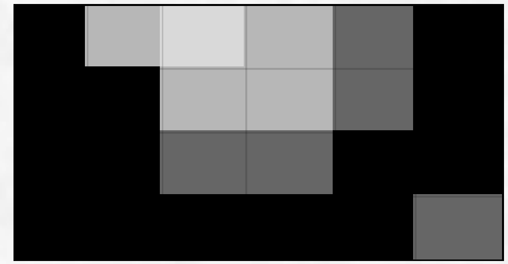


4. Related Works

4.5 **Astronomy in the Cloud: Using MapReduce for Image Co-Addition (K. Wiley et al)**

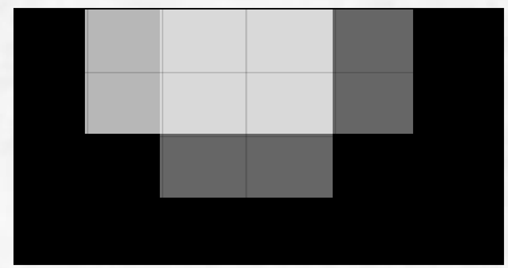
1st Image ⇒

0	2	3	2	1	0
0	0	2	2	1	0
0	0	1	1	0	0
0	0	0	0	0	1



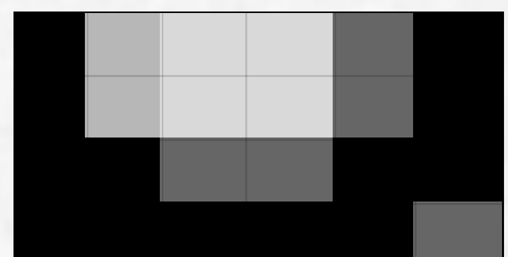
2nd Image ⇒

0	2	3	3	1	0
0	2	3	3	1	0
0	0	1	1	0	0
0	0	0	0	0	0



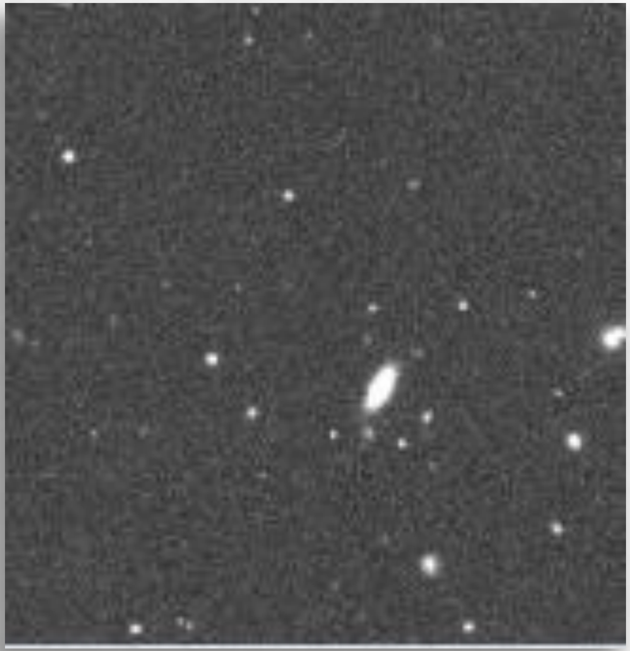
Resulting Image ⇒

0	2	3	3	1	0
0	2	3	3	1	0
0	0	1	1	0	0
0	0	0	0	0	1

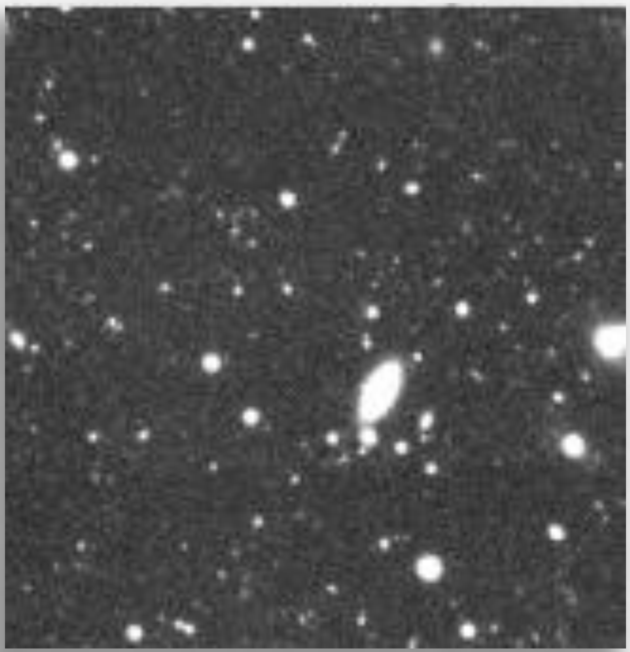


4. Related Works

4.5 Astronomy in the Cloud: Using MapReduce for Image Co-Addition (K. Wiley et al)



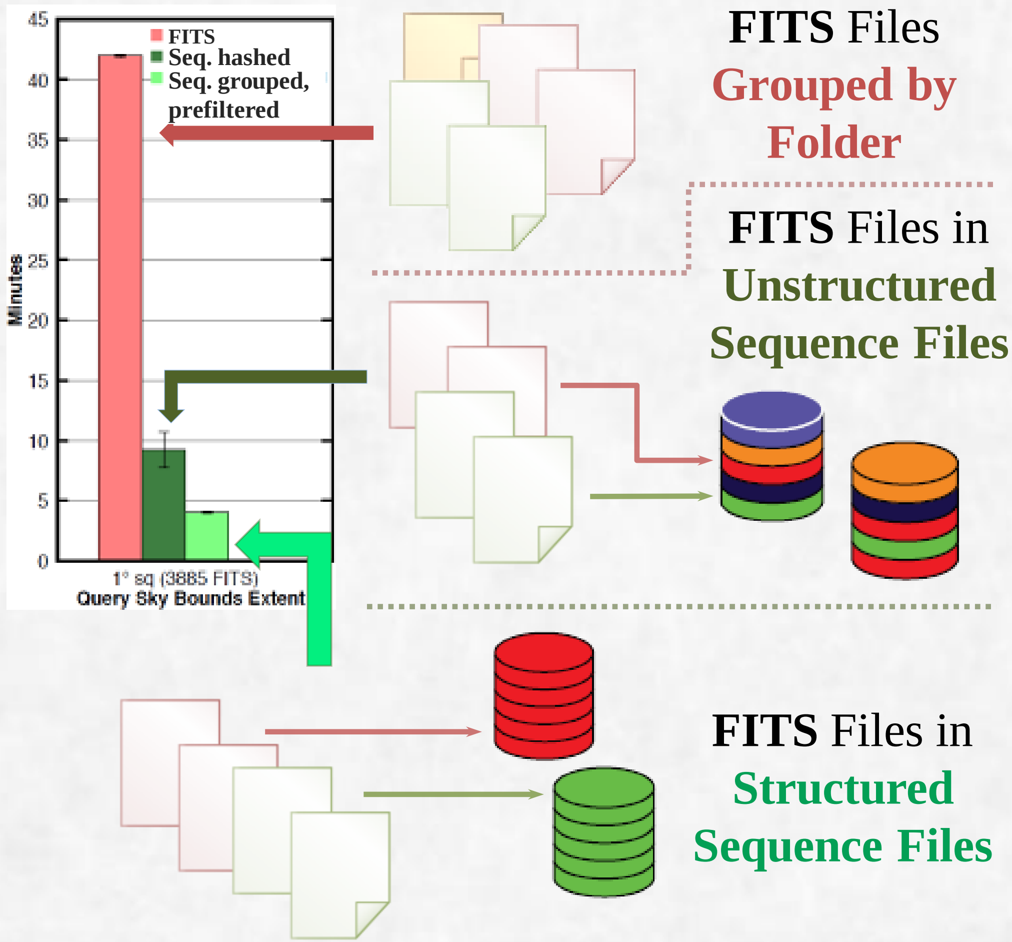
Before



After (96 IMG Co-Add)

4. Related Works

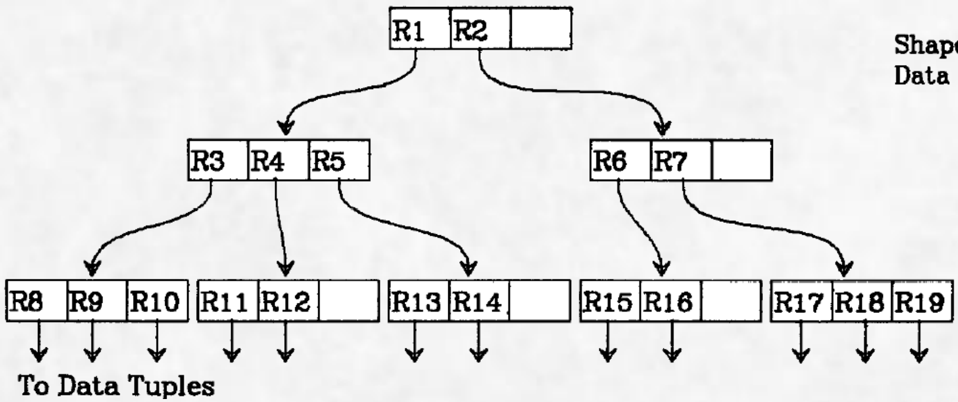
4.5 **Astronomy in the Cloud: Using MapReduce for Image Co-Addition (K. Wiley et al)**

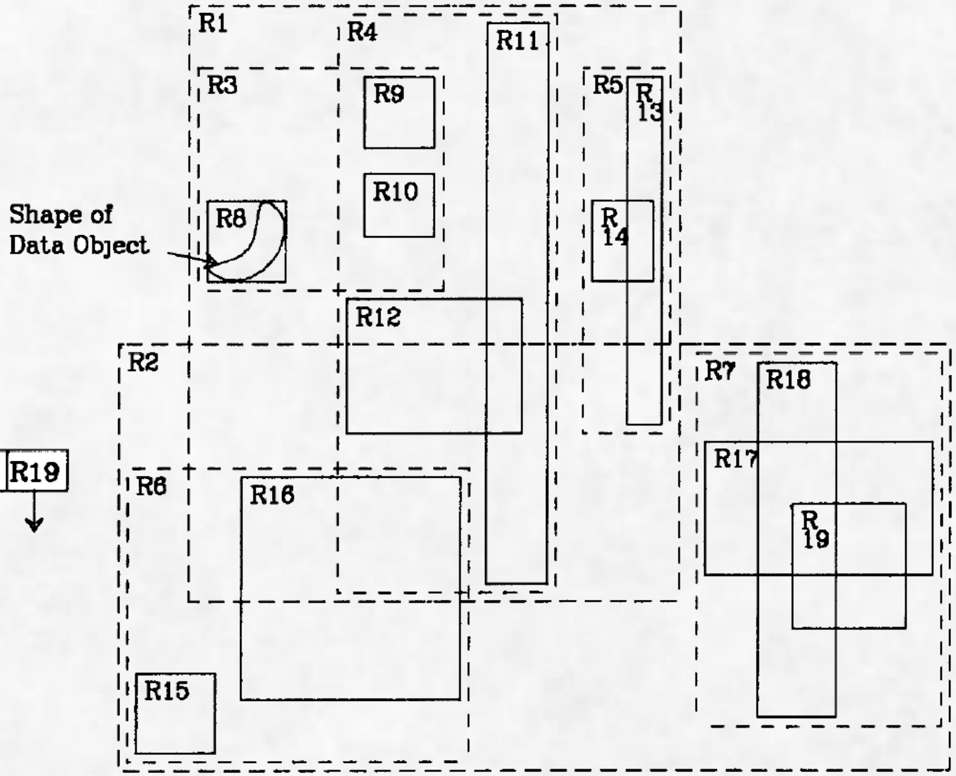


4. Related Works

4.6 **R-Trees**

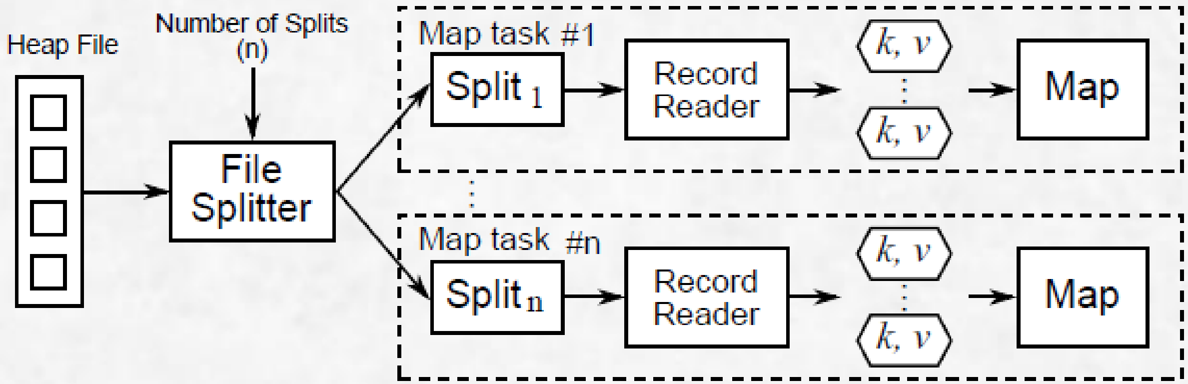
To handle spatial data efficiently a database system needs an index mechanism native to spatial locations.



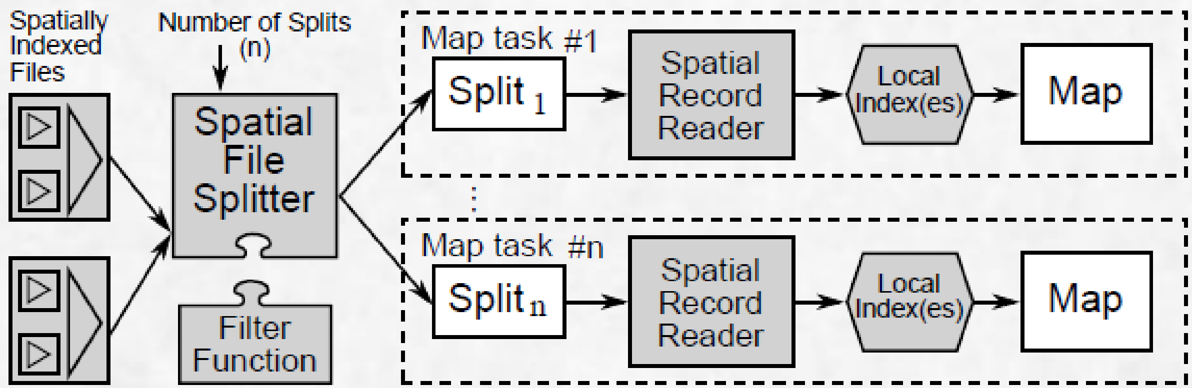


4. Related Works

4.7 GeoJinni (previously Spatial Hadoop)



(a) Hadoop



(b) SpatialHadoop

5. Approach

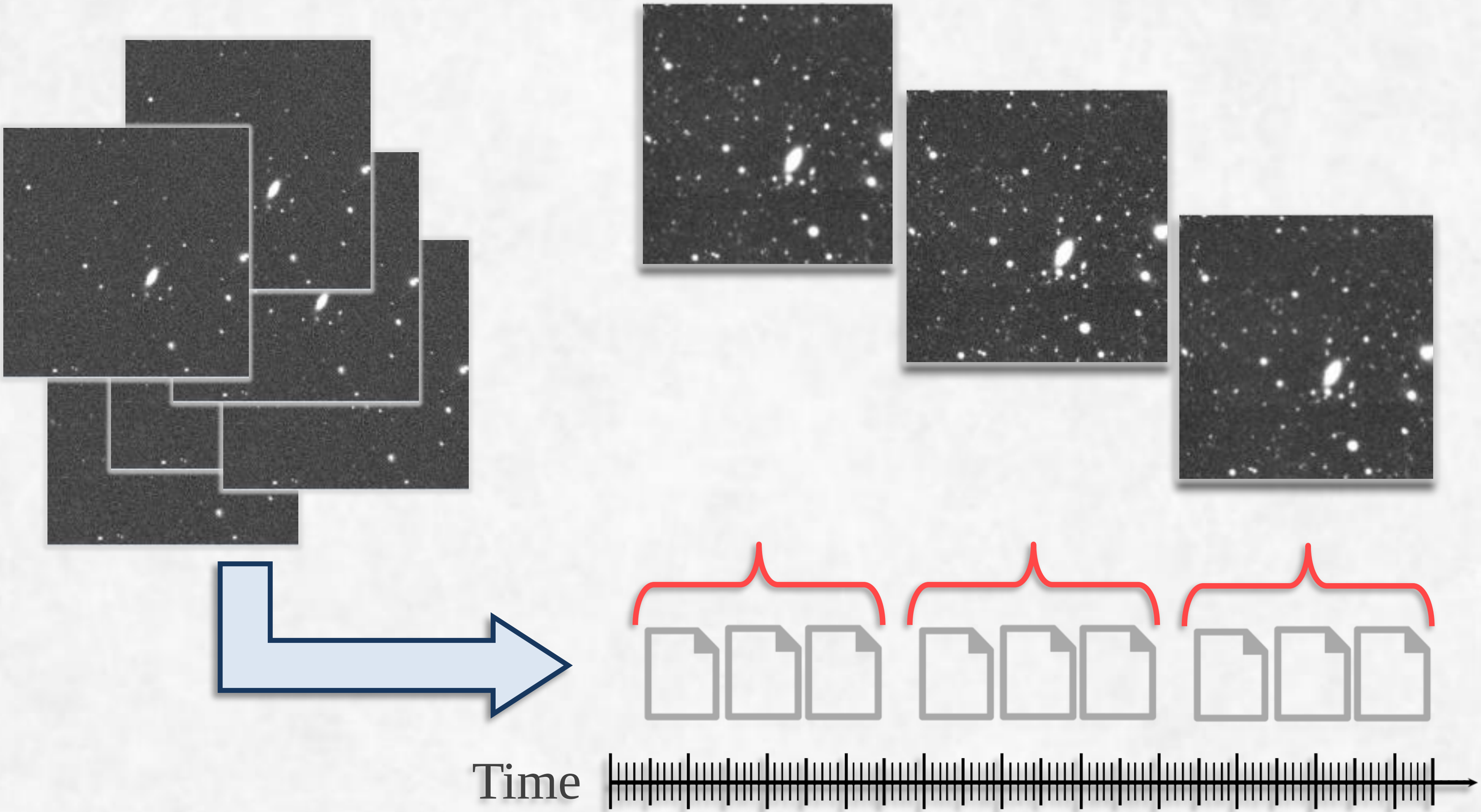
5.1 **Spatial search and indexing of FITS files (GeoJinni)**

- FITS files indexing through the information on their headers can be beneficial to improve query performance.

- Helps us define “Regions” which work as the boundaries for co-added mosaics.

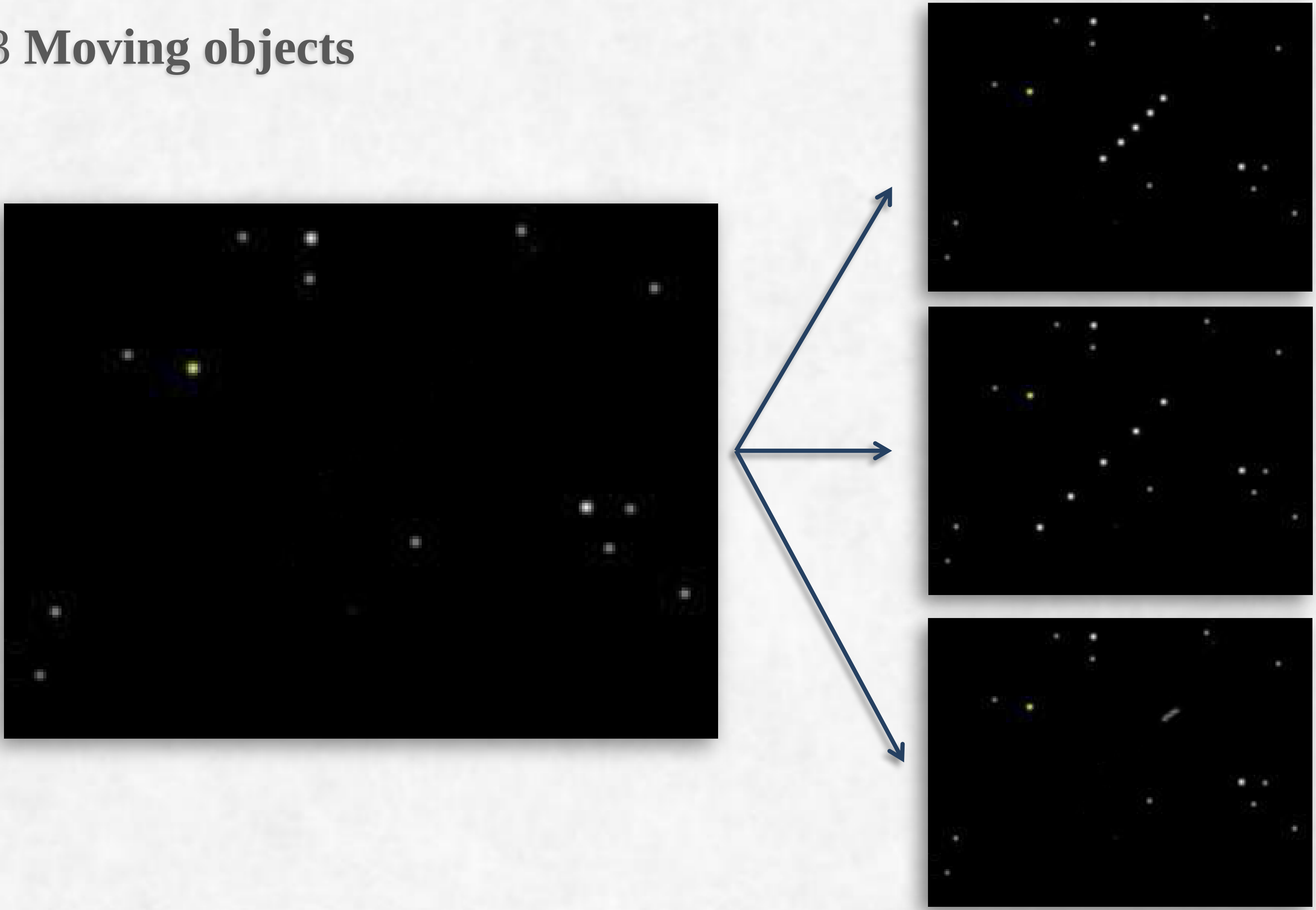
5. Approach

5.2 Image co-addition through cloud computing



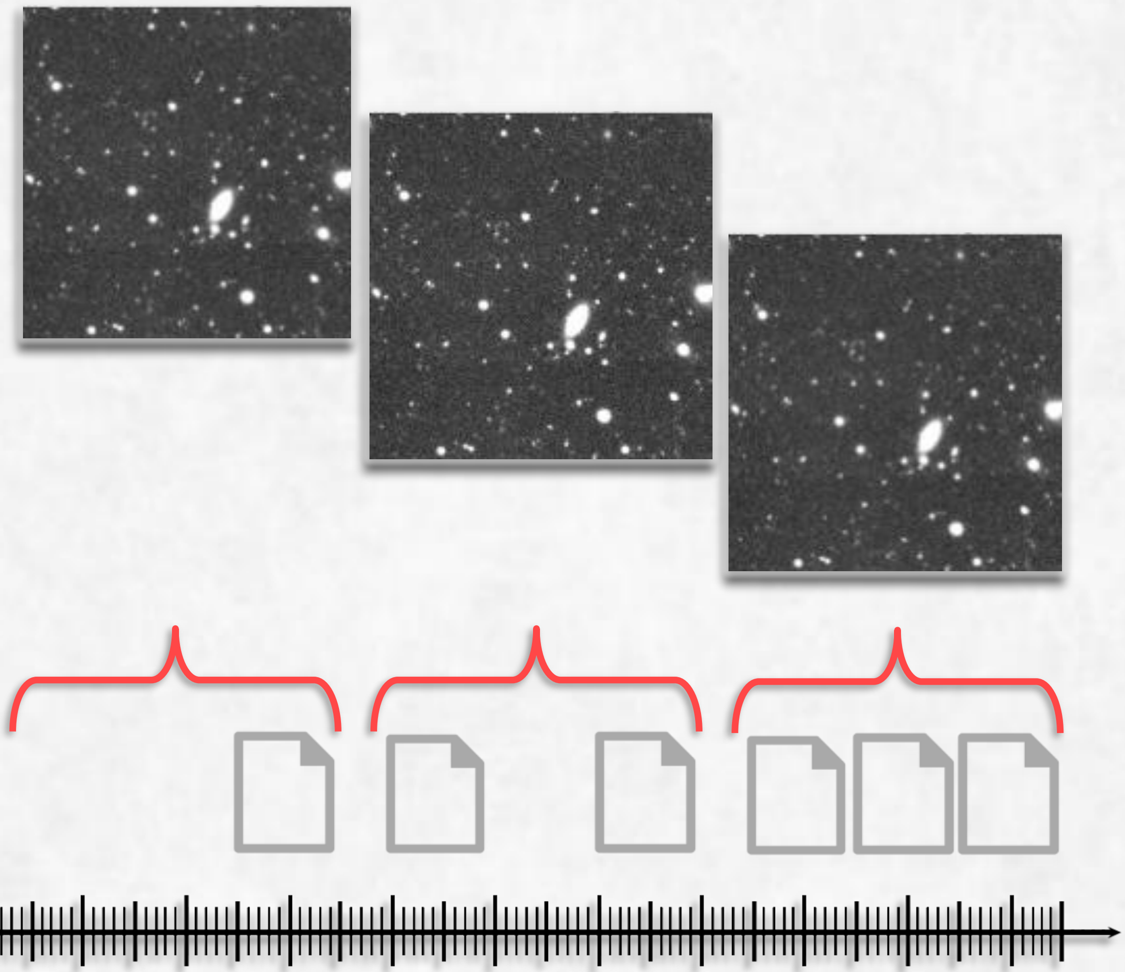
5. Approach

5.3 **Moving objects**



6. Challenges

6.1 **Defined period of time** VS **Equivalent amount of files**



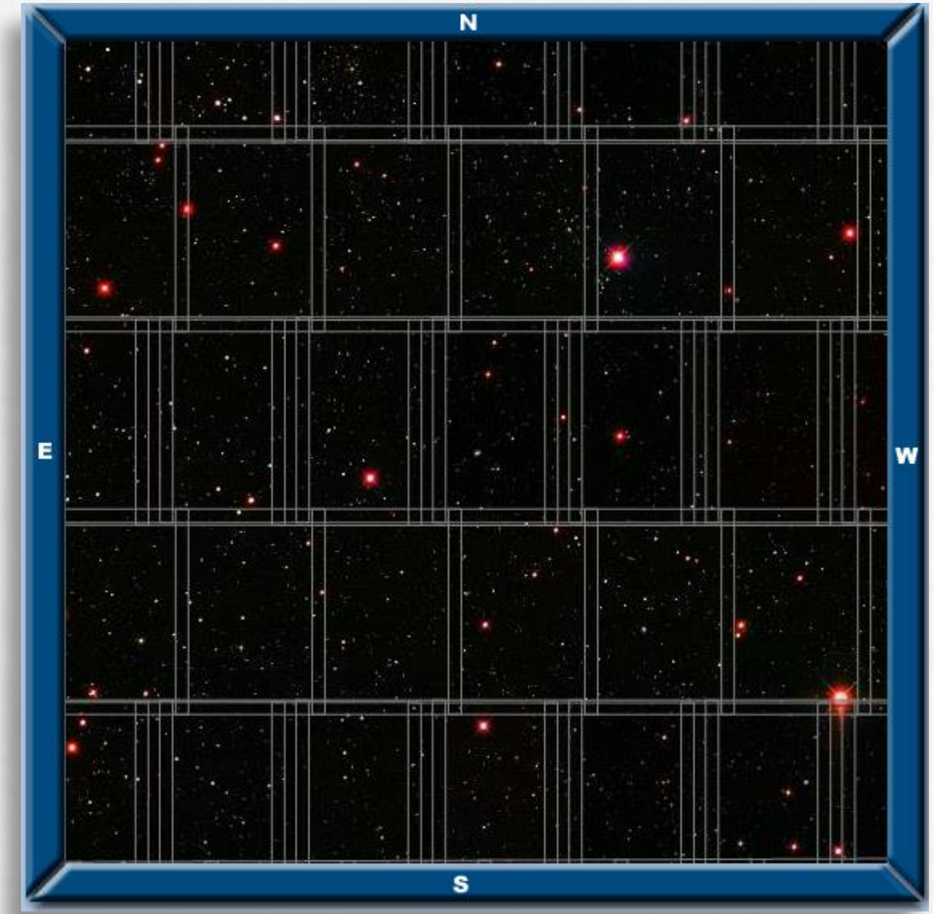
Time

6. Challenges

6.2 **Speed and direction of moving objects**

- Fast moving objects could exit quickly the field of view of the telescope, entering into a zone that wont be imaged until the next re-image of the sky much time later.

- The direction is also a factor on how long this objects would appear through the images.



6. Challenges

6.3 **Locating an appropriate data set**

- Must contain a stripe that has been imaged many times.
- Must contain at least one known moving object.
- Preferably, many light sources.
- Currently looking for a more periodic object, easier to calculate its trajectory to match it to an SDSS image field / stripe.

7. Summary and Next Steps

7.1 **Summary**

- Taking advantage of cloud technologies, process large amounts of files used in Astronomy incorporating astronomical spatial data into the index.
- Spatio-temporal mosaics difference comparison in an attempt to find moving objects.

7.2 **Next Steps**

- Continue developing integration of the ICRS coordinates into the image co-addition prototype system.
- Locate an appropriate data set.

8. References

[1] Tom Cwik, *Towards the Application of Largescale Computation in Space Exploration*. Talk at the NCSA of the University of Illinois (March, 2015). https://youtu.be/tKhS6-zWS-g

[2] Space Telescope Science Institute, http://www.stsci.edu/instruments/wfpc2/Wfpc2_dhb/intro_ch23.html

[3] The Sloan Digital Sky Survey (SDSS). http://www.sdss.org/

[4] Dean Jeffrey, Sanjay Ghemawat. *MapReduce: Simplified Data Processing on Large Clusters*. Communications of the ACM - 50th anniversary issue: 1958 - 2008 CACM,1 (2008), vol 51, pp 107-113

[5] *Hadoop Overview*. www.glennklockwood.com/data-intensive/hadoop/overview.html

[6] *Hadoop* https://hadoop.apache.org/

[7] K. Wiley et al. *Astronomy in the Cloud: Using MapReduce for Image Co-Addition*. Publications of the Astronomical Society of the Pacific, 123 (2011), pp. 366–380

[8] Antonin Guttman*, RTrees. A Dynamic Index Structure for Spatial Searching*. SIGMOD '84 Proceedings of the 1984 ACM SIGMOD international conference on Management of data pp 4757

[9] Ahmed Eldawy and Mohamed F. Mokbel*, SpatialHadoop: A MapReduce Framework for Spatial Data*. Proceedings of the IEEE International Conference on Data Engineering, ICDE 2015, Seoul, South Korea, April, 2015

[10] Arias, E. F. et al. *The extragalactic reference system of the International Earth Rotation Service, ICRS*. Astrophys, 303, pp. 604-608.

[11] U.S. Naval Observatory. http://aa.usno.navy.mil/faq/docs/ICRS_doc.php