

```
/**  
 * CUDA Kernel Device code for combined rotation and cropping of a list of images.  
 */  
template <unsigned int block_size>  
global void  
rotateAndCropTexture kernel(float *rotatedImages, float *image, int neuron_size,  
int neuron_dim, int image_dim, float *cosAlpha, float *sinAlpha)  
{  
    int x2 = blockIdx.x * blockDim.x + threadIdx.x;  
    float cosAlpha_local = cosAlpha[blockIdx.z];  
    float sinAlpha_local = sinAlpha[blockIdx.z];  
    if (inputData.verbose) cout << "\n Size of SOM = " << som.getSize() << endl;  
    float *d_som = cuda alloc float(som.getSize());  
    cuda_copyHostToDevice_float(d_som, som.getDataPointer(), som.getSize());  
    // Memory allocation  
    int rotatedImagesSize = inputData.numberOfChannels * inputData.numberOfRotationsAndFlip * inputData.neuron_size;  
    if (inputData.verbose) cout << "\n Size of rotated images = " << rotatedImagesSize * sizeof(float) << " bytes" << endl;  
    float *d_rotatedImages = cuda_alloc_float(rotatedImagesSize);  
    if (inputData.verbose) cout << "\n Size of euclidean distance matrix = " << inputData.som_size * sizeof(float) << " bytes" << endl;  
    float *d_euclid = cuda_alloc_float(inputData.som_size * inputData.som_size * sizeof(float));  
    // Prepare trigonometric values  
    float *d_cosAlpha = NULL, *d_sinAlpha = NULL;  
    int neuron_dim, int image_dim, float *cosAlpha, float *sinAlpha;  
    int prunedMetricValues(&d_cosAlpha, &d_sinAlpha, inputData.som_size, neuron_dim, image_dim);  
    int x2 = blockIdx.x * blockDim.x + threadIdx.x;  
    int y2 = blockIdx.y * blockDim.y + threadIdx.y;  
    int x0margin = x0 - margin;  
    int y0margin = y0 - margin;  
    float cosAlpha_local = cosAlpha[blockIdx.z];  
    float sinAlpha_local = sinAlpha[blockIdx.z];  
    int x1 = (x2-x0margin)*cosAlpha_local - (y2-y0margin)*sinAlpha_local;  
    int y1 = (y2-y0margin)*cosAlpha_local + (x2-x0margin)*sinAlpha_local;  
    atomicAdd(pCurRot + x2*nCurRot, d_rotatedImages[x1 + y1*neuron_dim]);  
}
```

Wide Field Outlier Finder (WTF)

using cloud infrastructure to cooperate, scale, and democratise e-science

Crowd-sourcing / Citizen Science



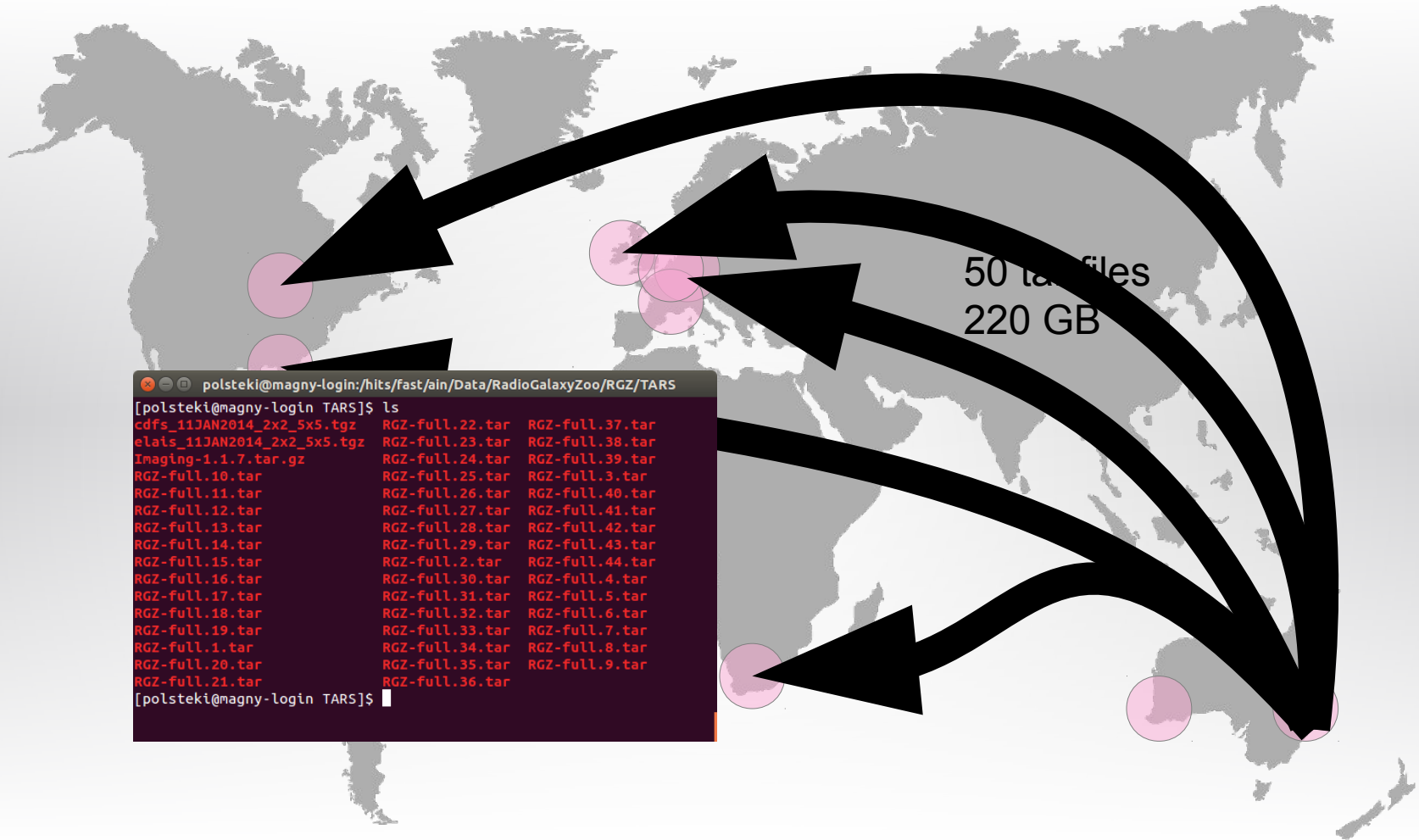
The screenshot shows a web browser displaying the Galaxy Zoo Radio website. The browser's address bar shows the URL www.galaxyzoo.org/#/. The website's navigation menu includes links for CLASSIFY, SCIENCE, TEAM, PROFILE, TALK, and BLOG. The main heading is "In Search of Erupting Black Holes", with a sub-heading: "Help astronomers discover supermassive black holes observed by the KG Jansky Very Large Array (NRAO) and the Australia Telescope Compact Array (CSIRO)". Below the text is a "Search for Black Holes" section with a paragraph of introductory text and a yellow "Begin Hunting" button. A large image of two galaxies with jets is on the right. At the bottom right, there is a small credit line: "NASA, ESA, S. Baum and C. O'Dea (RIT), R. Perley and W. Cotton (NRAO/AUI/NSF), and the Hubble Heritage Team (STScI/AURA)".

Radio Galaxy Zoo



The screenshot displays the Radio Galaxy Zoo web application interface. At the top, a navigation menu includes links for CLASSIFY, SCIENCE, TEAM, PROFILE, TALK, and BLOG. The central area features a large blue-tinted image of a galaxy with white radio emission contours overlaid. To the right of this image are three circular icons: a magnifying glass, a monitor, and a keyboard. Below the main image is a horizontal row of five smaller thumbnail images, each showing a different view or zoom level of the galaxy's radio emission. The top right corner of the interface shows the user's name 'polsterer', a 'SIGN OUT' link, and the language 'English'.

Starting the Project



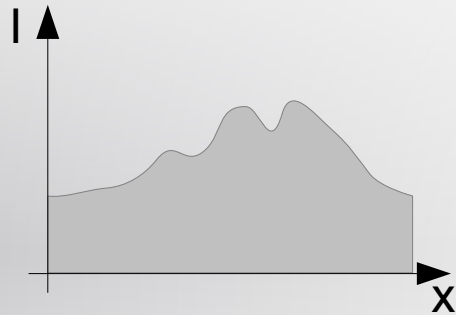
```
polsteki@magny-login:/hits/fast/ain/Data/RadioGalaxyZoo/RGZ/TARS
[polsteki@magny-login TARS]$ ls
cdfs_11JAN2014_2x2_5x5.tgz  RGZ-full.22.tar  RGZ-full.37.tar
elais_11JAN2014_2x2_5x5.tgz  RGZ-full.23.tar  RGZ-full.38.tar
Imaging-1.1.7.tar.gz        RGZ-full.24.tar  RGZ-full.39.tar
RGZ-full.10.tar             RGZ-full.25.tar  RGZ-full.3.tar
RGZ-full.11.tar             RGZ-full.26.tar  RGZ-full.40.tar
RGZ-full.12.tar             RGZ-full.27.tar  RGZ-full.41.tar
RGZ-full.13.tar             RGZ-full.28.tar  RGZ-full.42.tar
RGZ-full.14.tar             RGZ-full.29.tar  RGZ-full.43.tar
RGZ-full.15.tar             RGZ-full.2.tar   RGZ-full.44.tar
RGZ-full.16.tar             RGZ-full.30.tar  RGZ-full.4.tar
RGZ-full.17.tar             RGZ-full.31.tar  RGZ-full.5.tar
RGZ-full.18.tar             RGZ-full.32.tar  RGZ-full.6.tar
RGZ-full.19.tar             RGZ-full.33.tar  RGZ-full.7.tar
RGZ-full.1.tar              RGZ-full.34.tar  RGZ-full.8.tar
RGZ-full.20.tar             RGZ-full.35.tar  RGZ-full.9.tar
RGZ-full.21.tar             RGZ-full.36.tar
[polsteki@magny-login TARS]$
```

Preprocessing

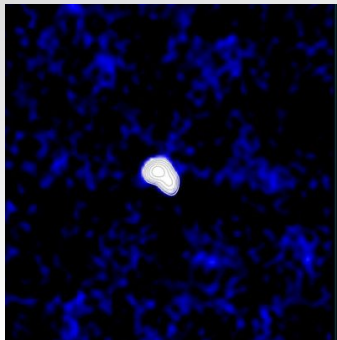
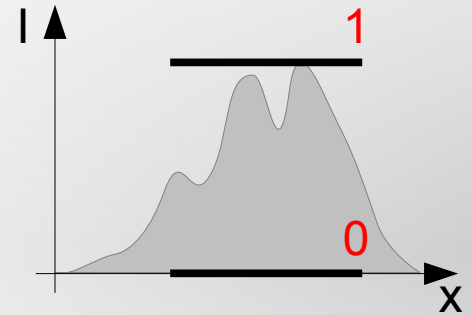


extract
matrix from fits

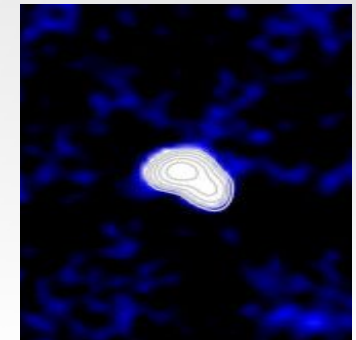
$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & & & A_{2n} \\ \vdots & & & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix}$$



normalize
flux relative to the maximum



cutout
interesting region



Speeding up Preprocessing



single core python = **48**
hours



with  **hadoop**

on **4x16** cores = **4**
hours



file **access** is still the bottleneck!

New Images Extracted



```
polsteki@magny-login:/hits/fast/ain/Data/RadioGalaxyZoo/RGZ/TARS
[polsteki@magny-login TARS]$ ls
cdfs_11JAN2014_2x2_5x5.tgz  RGZ-full.22.tar  RGZ-full.37.tar
elais_11JAN2014_2x2_5x5.tgz  RGZ-full.23.tar  RGZ-full.38.tar
Imaging-1.1.7.tar.gz        RGZ-full.24.tar  RGZ-full.39.tar
RGZ-full.10.tar             RGZ-full.25.tar  RGZ-full.3.tar
RGZ-full.11.tar             RGZ-full.26.tar  RGZ-full.40.tar
RGZ-full.12.tar             RGZ-full.27.tar  RGZ-full.41.tar
RGZ-full.13.tar             RGZ-full.28.tar  RGZ-full.42.tar
RGZ-full.14.tar             RGZ-full.29.tar  RGZ-full.43.tar
RGZ-full.15.tar             RGZ-full.2.tar   RGZ-full.44.tar
RGZ-full.16.tar             RGZ-full.30.tar  RGZ-full.4.tar
RGZ-full.17.tar             RGZ-full.31.tar  RGZ-full.5.tar
RGZ-full.18.tar             RGZ-full.32.tar  RGZ-full.6.tar
RGZ-full.19.tar             RGZ-full.33.tar  RGZ-full.7.tar
RGZ-full.1.tar              RGZ-full.34.tar  RGZ-full.8.tar
RGZ-full.20.tar             RGZ-full.35.tar  RGZ-full.9.tar
RGZ-full.21.tar             RGZ-full.36.tar
[polsteki@magny-login TARS]$
```

Similarity Measure



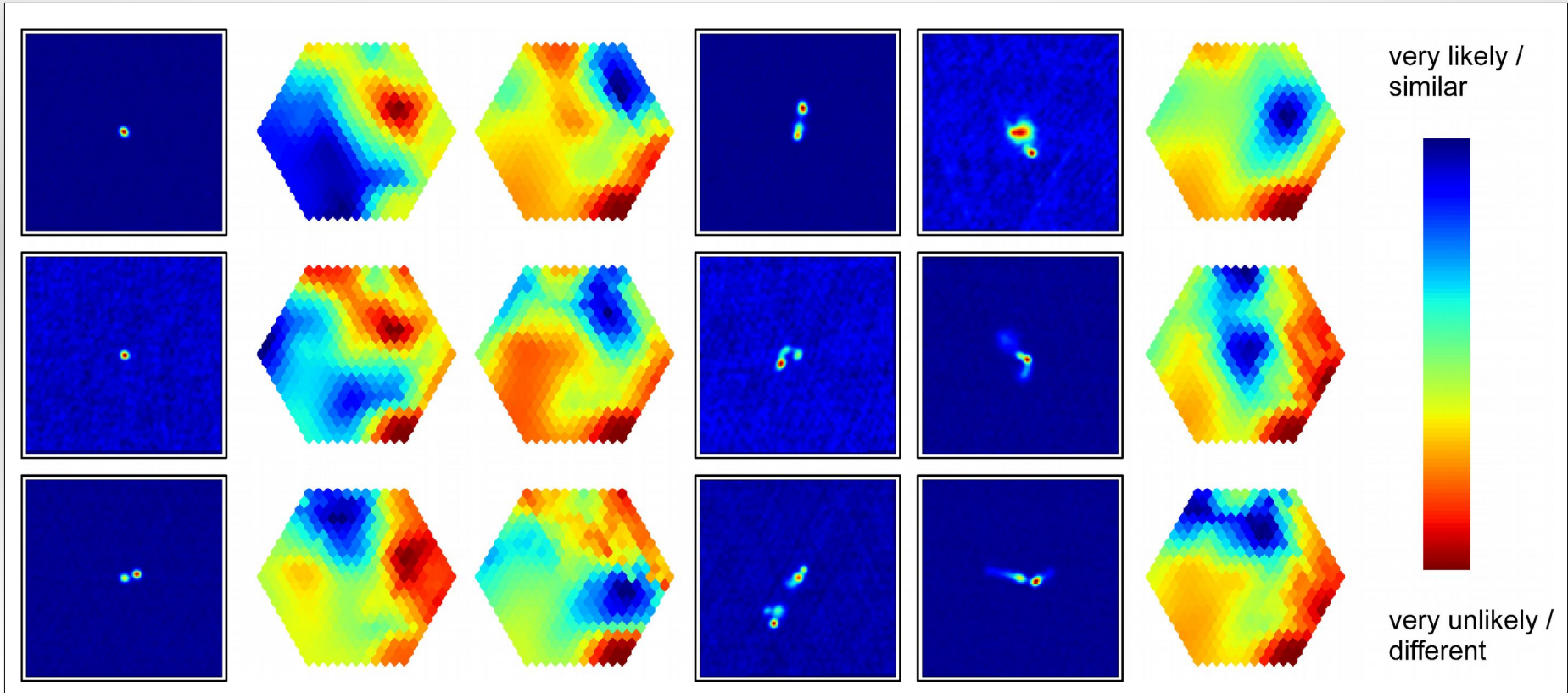
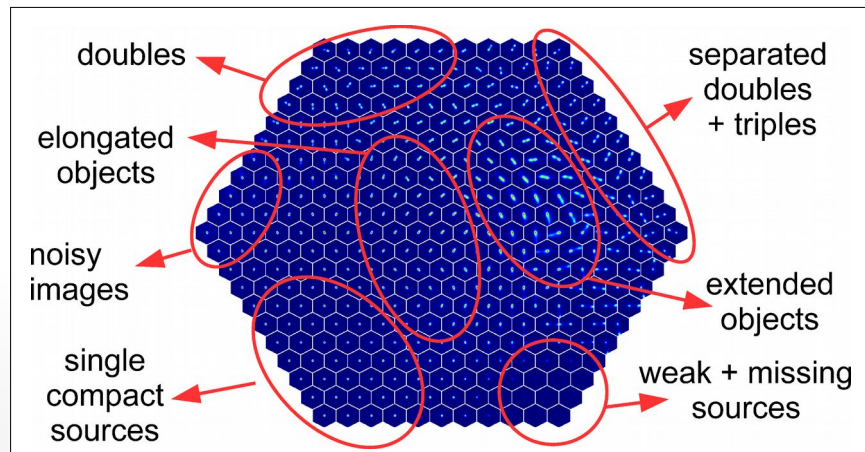
calculate the pixel based
Euclidean distance

for all possible
rotations
and find the best matching angle

A series of green, 3D-style text 'rotations' arranged in a semi-circular arc, illustrating the concept of testing multiple angles.

via
minimization → **GPU**

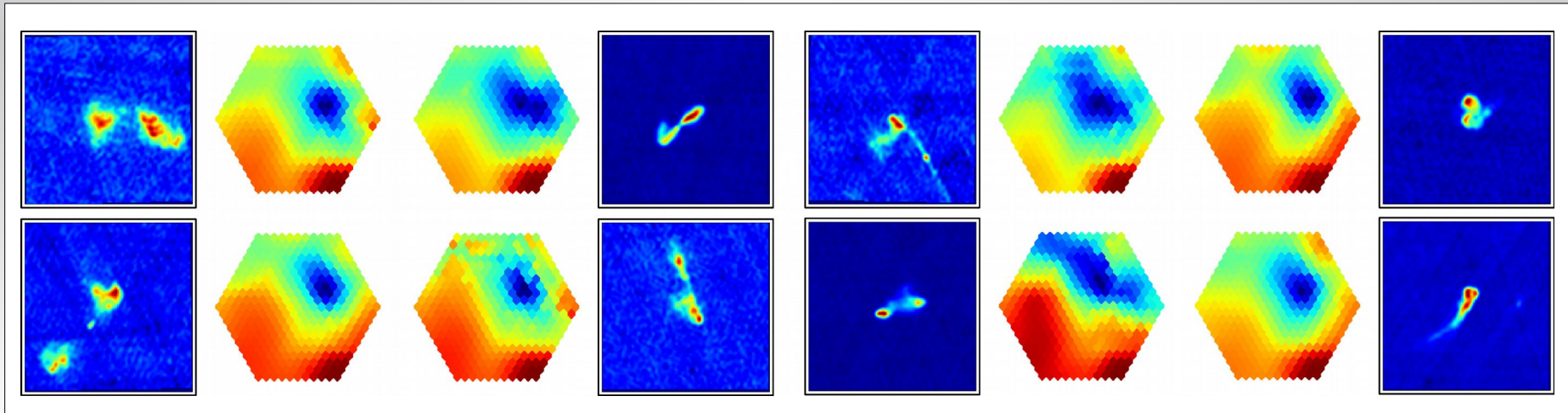
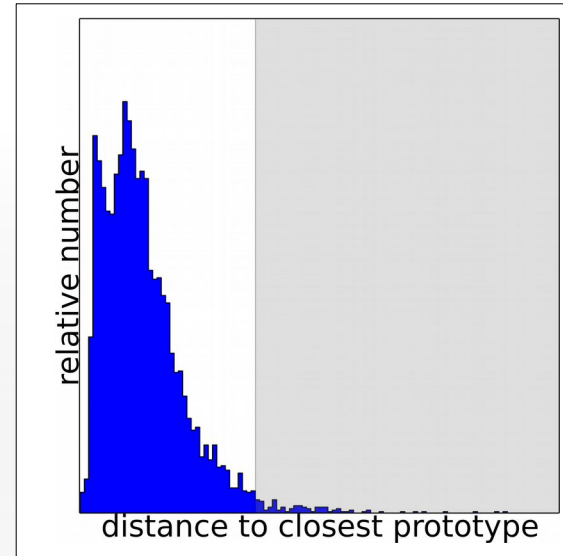
Processing



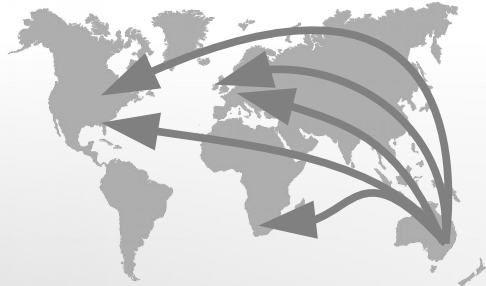
Interesting Objects



select
outliers
based
on distribution of distances



The Downsides of this Approach



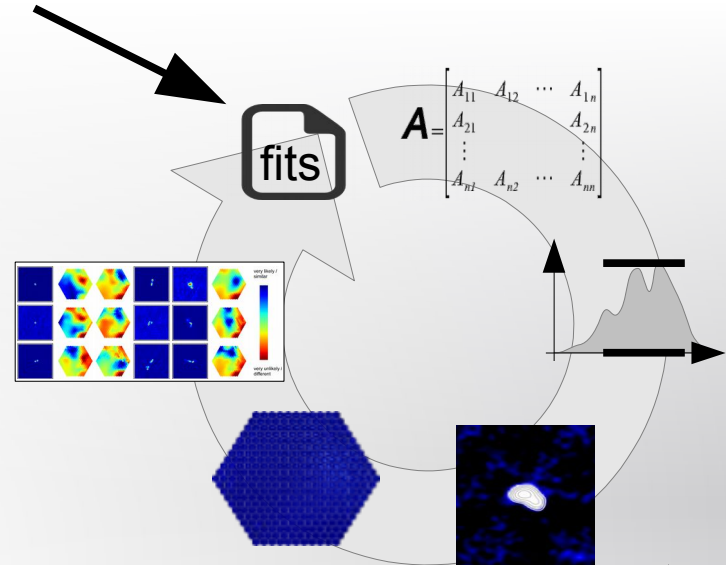
a lot of local copies

no orchestration of work-flow



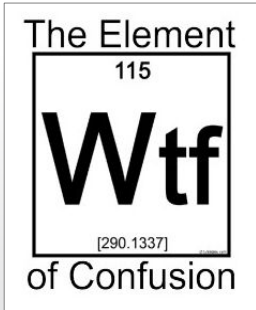
bad exchange of intermediate results

very exclusive concerning hardware requirements



NVIDIA Tesla K40

A new Project



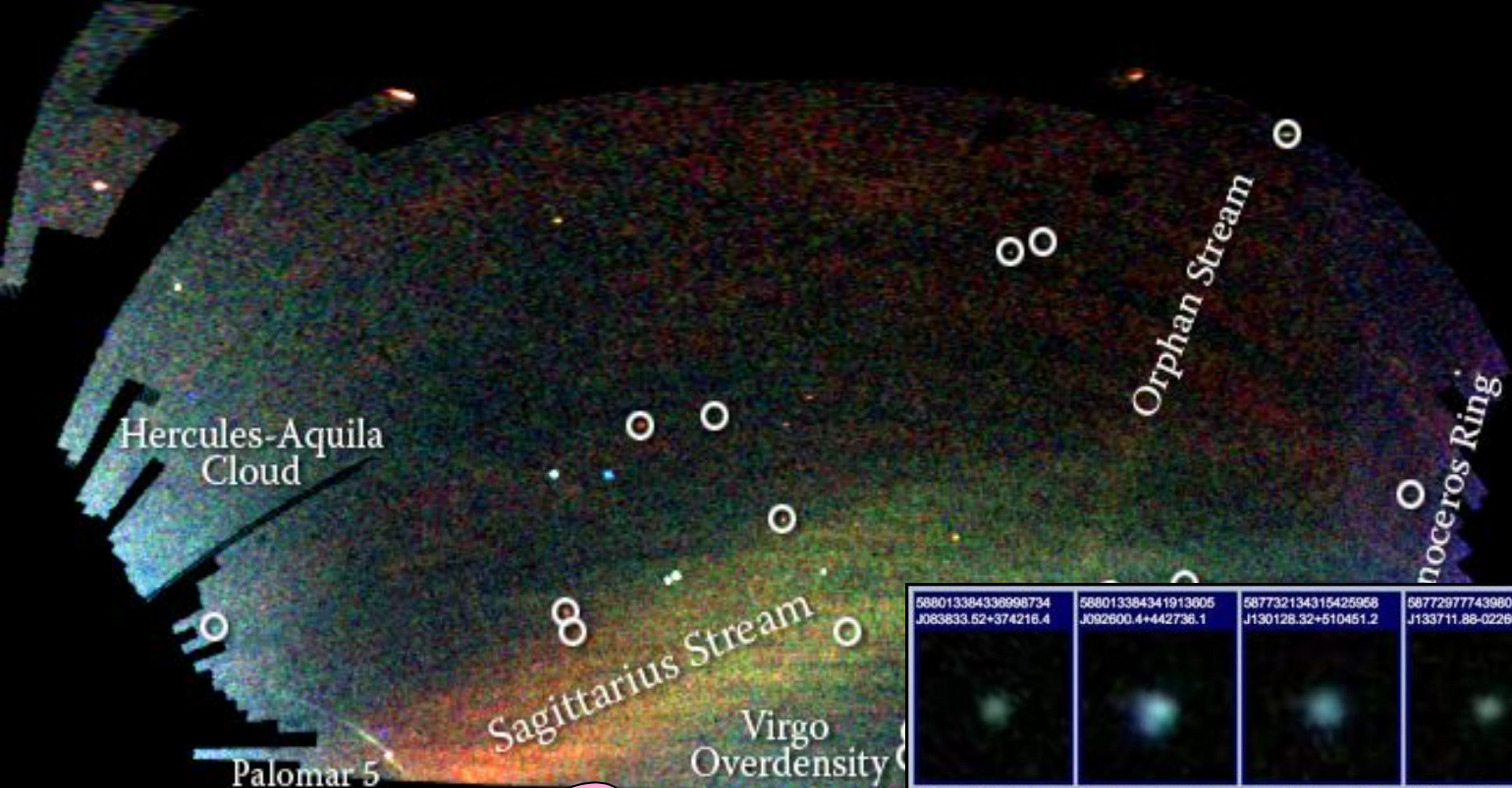
Widefield ou**T**lier **F**inder
(Ray Norris)

ASKAP-EMU project
from **2.5 mill.** → **70 mill.** known radio sources

develop methodology to discover
unexpected

science in large data sets

What are Outliers?



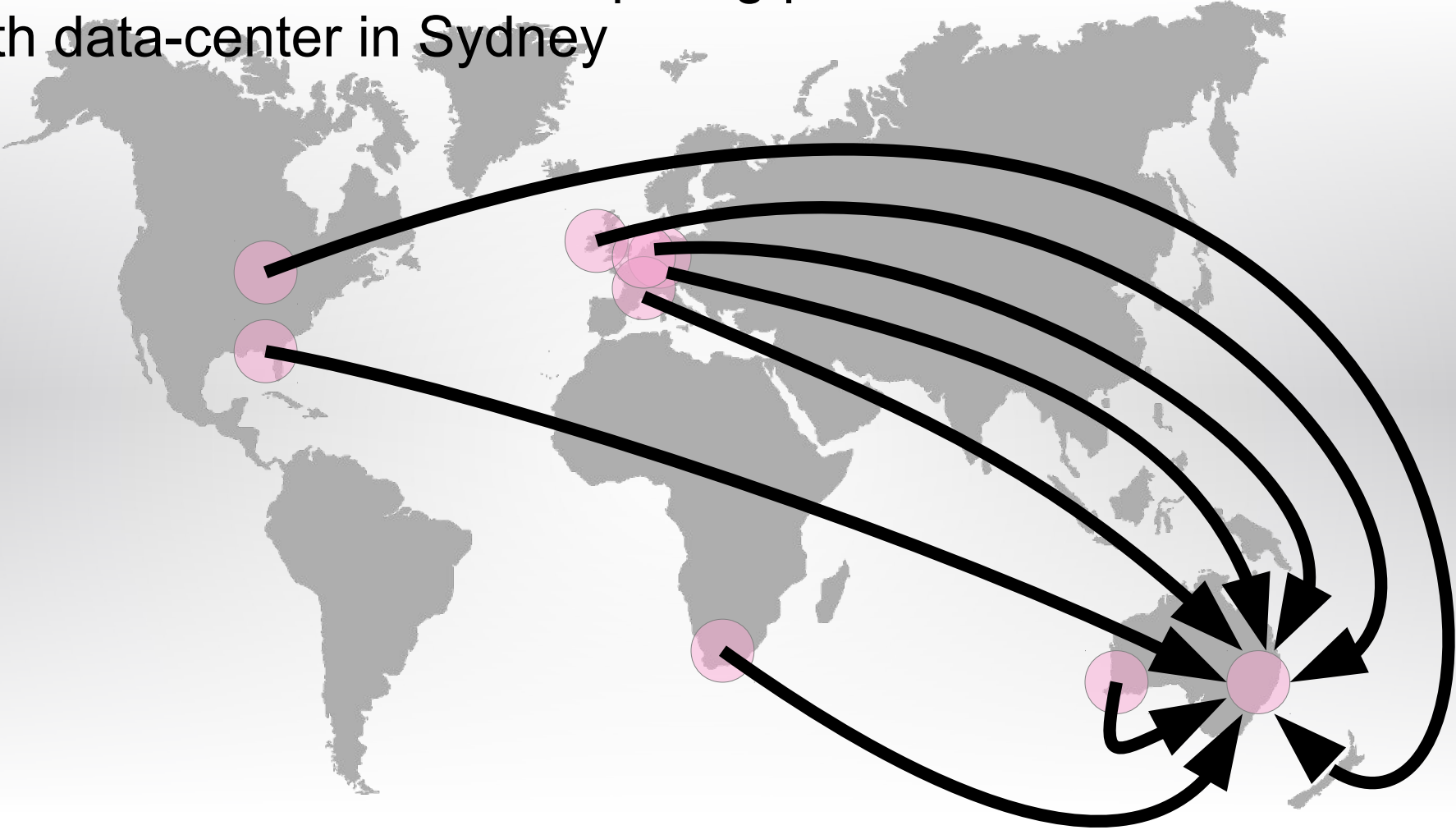
IFRS

588013384338998734 J083833.62+374216.4	588013384341913605 J092600.4+442736.1	587732134315425958 J130128.32+510451.2	587729777439801619 J133711.88-022605.4	587729777446945029 J144231.37-020952
587729970180522426 J150728.4-023351.2	587732152555864324 J074758+233632.8	587732156853846378 J080816.9+281431.1	587732153639829774 J090918.36+392924.7	587732152033345685 J095618.32+430727.8

A new Concept



use commercial cloud computing provider
with data-center in Sydney



Why use a Company?

international project:

- who is providing resources?
- who has the infrastructure?
- who provides the services?

why not run our own science cloud?

- using open stack



Commercial Cloud Provider

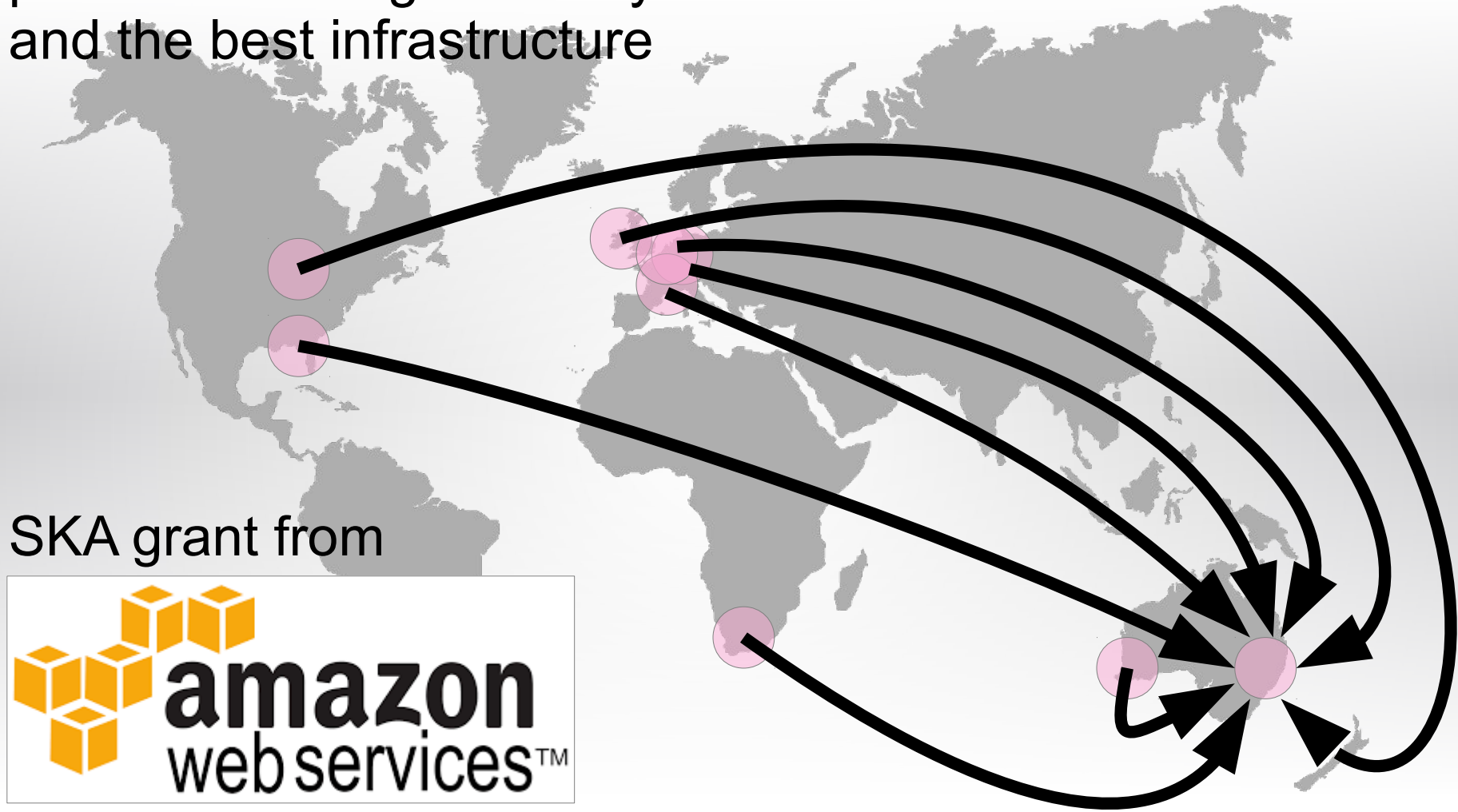
There are many Cloud Providers



Use Amazon Web Services



provides the largest variety of services and the best infrastructure



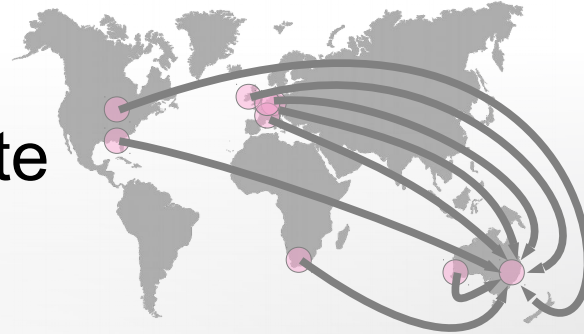
SKA grant from



The Concept

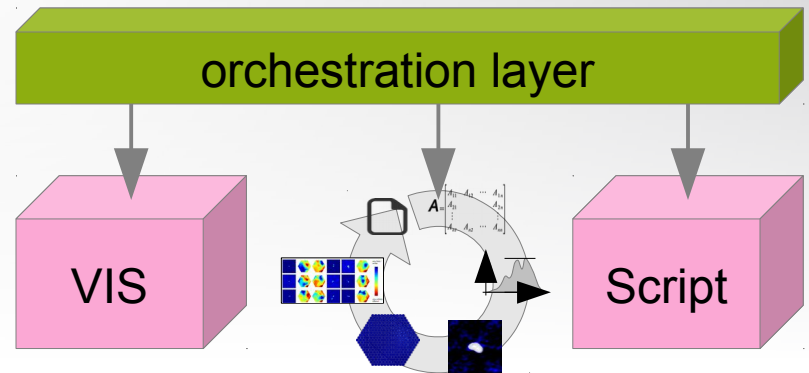


centralized storage and compute

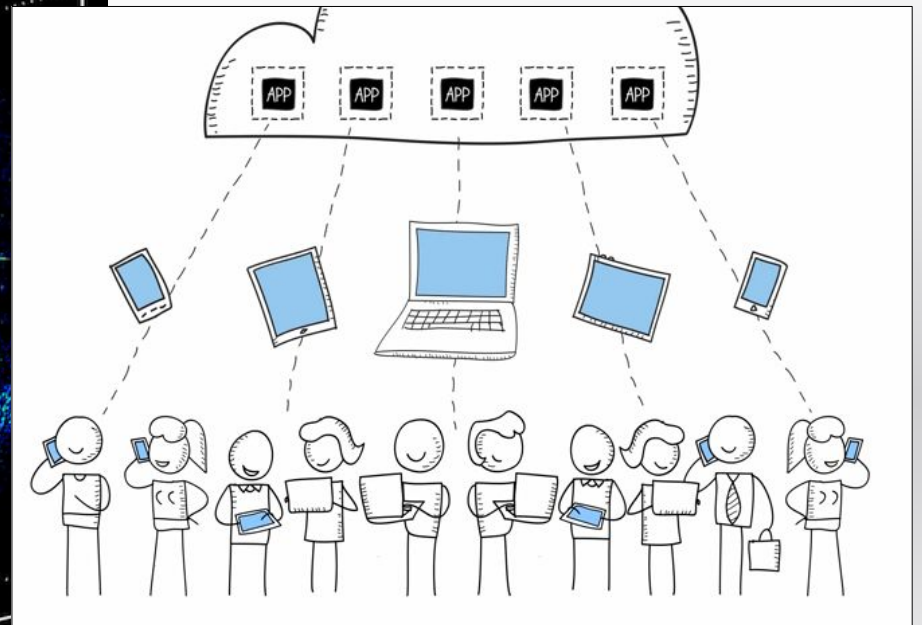
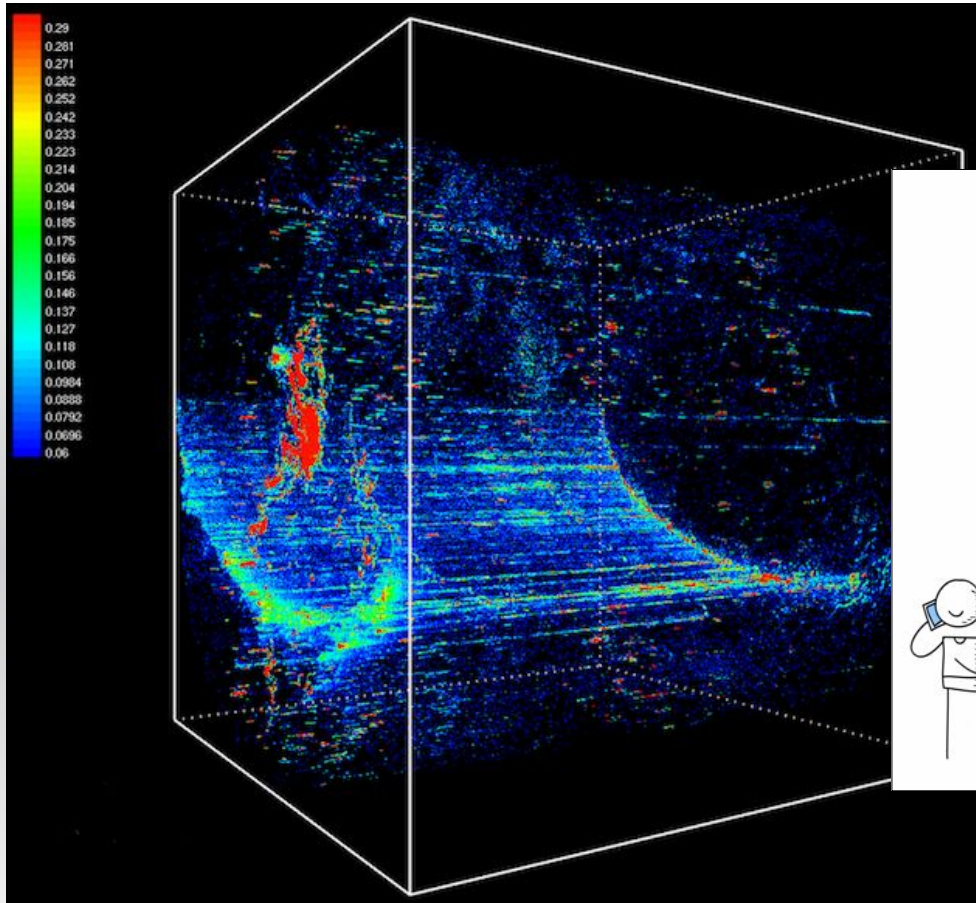


embedded repository and project management tool

script all preprocessing and provide orchestration



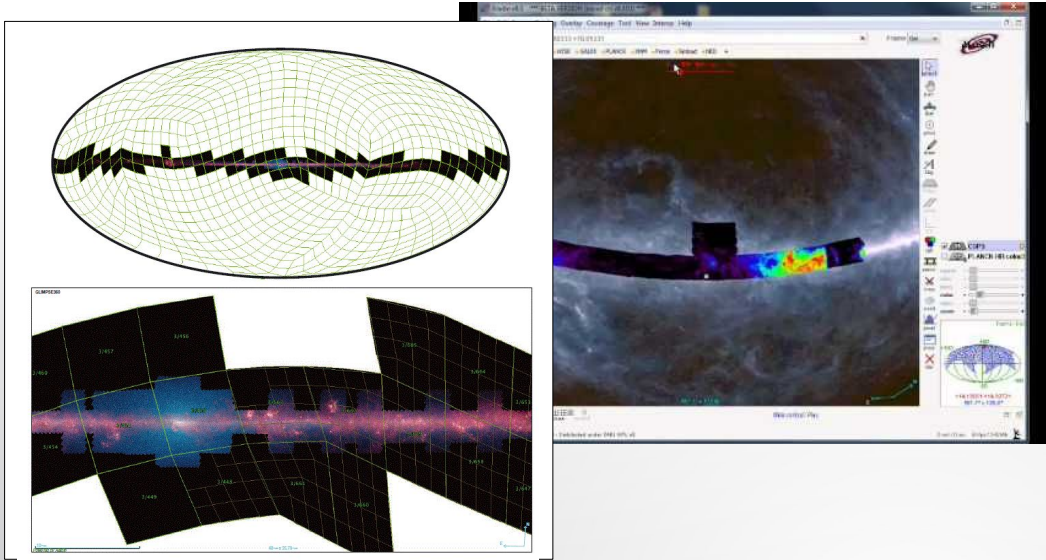
Remote Visualization



AWS appStream

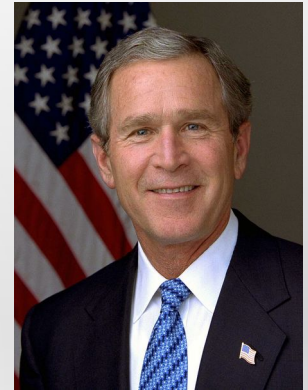
Amr Hassan

HIPS is Underestimated



Fernique et al. 2015 A&A

don't miss underestimate me



more than just
visualization

it's a
hierarchical data-structure

What is really new?

scaling **UP** and **DOWN** immediately

no longer depending on
exclu\$ive
hardware requirements

no **difference** between small university institutes and
large research centers

data via credits
challenges open to all

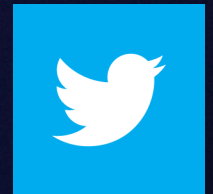
Conclusion



thanks to **Virtual
Observatory** we know
how to access data

we still need to learn how to
uniformly
process and
analyze data

Thank you for your attention!



@AstroInformatix