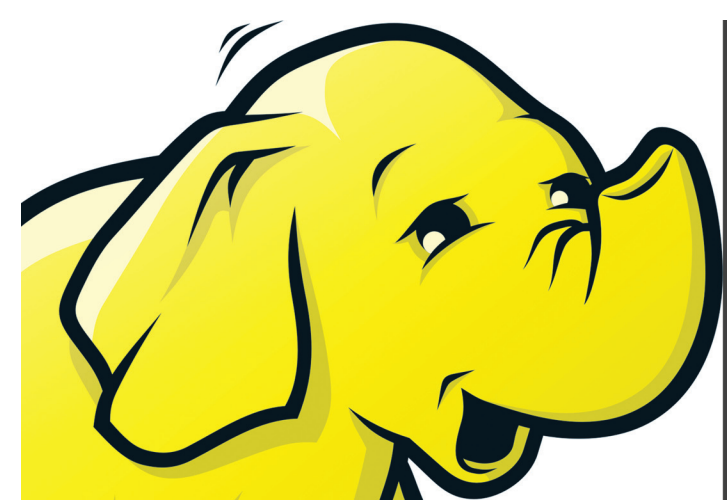


Data Intensive Supercomputing HPC Application Exploration

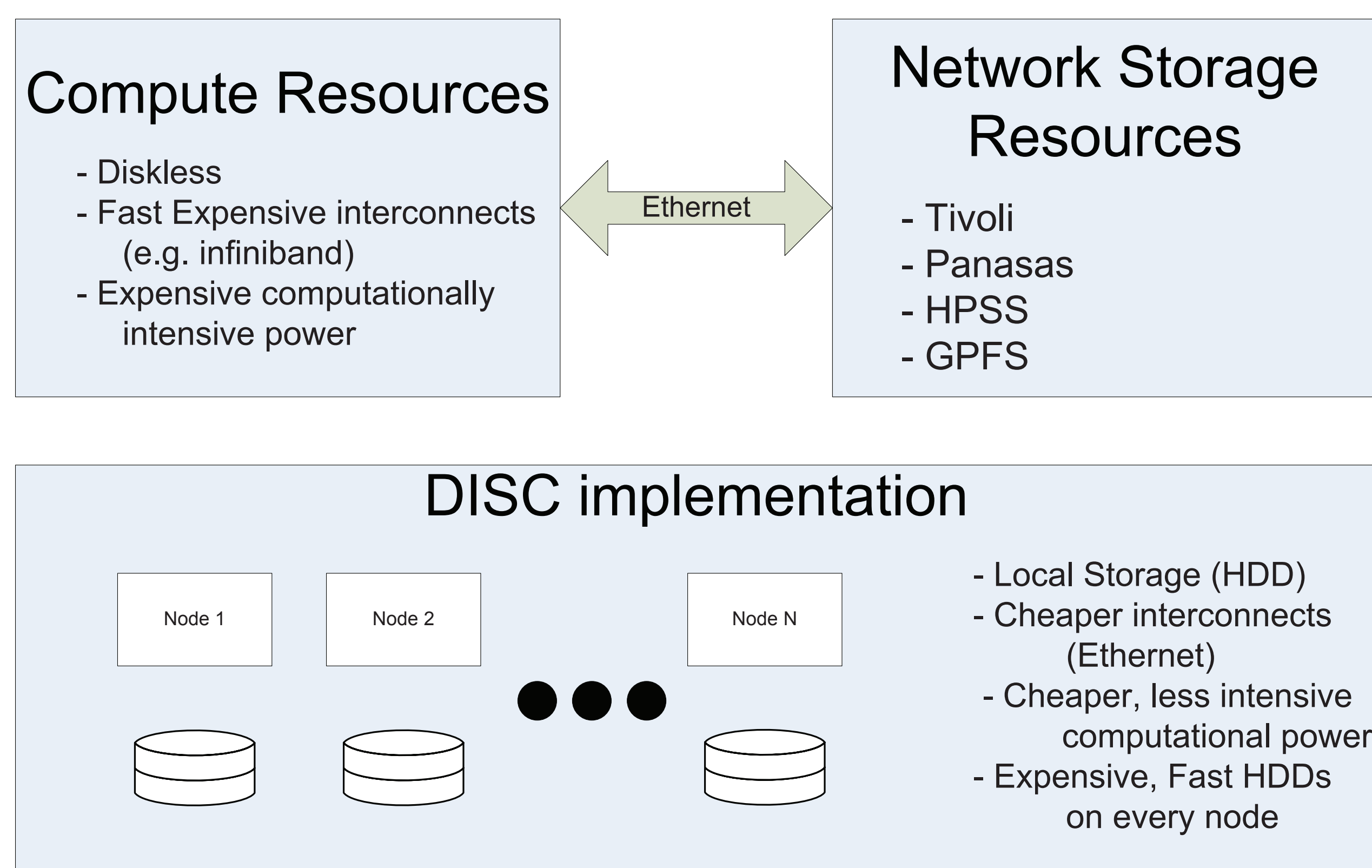
Grant Mackey (HPC-5) and Christopher Mitchell (HPC-5)

Mentors: John Bent (HPC-5) and James Nunez (HPC-5) / Project Sponsor: Gary Grider (HPC-DO)



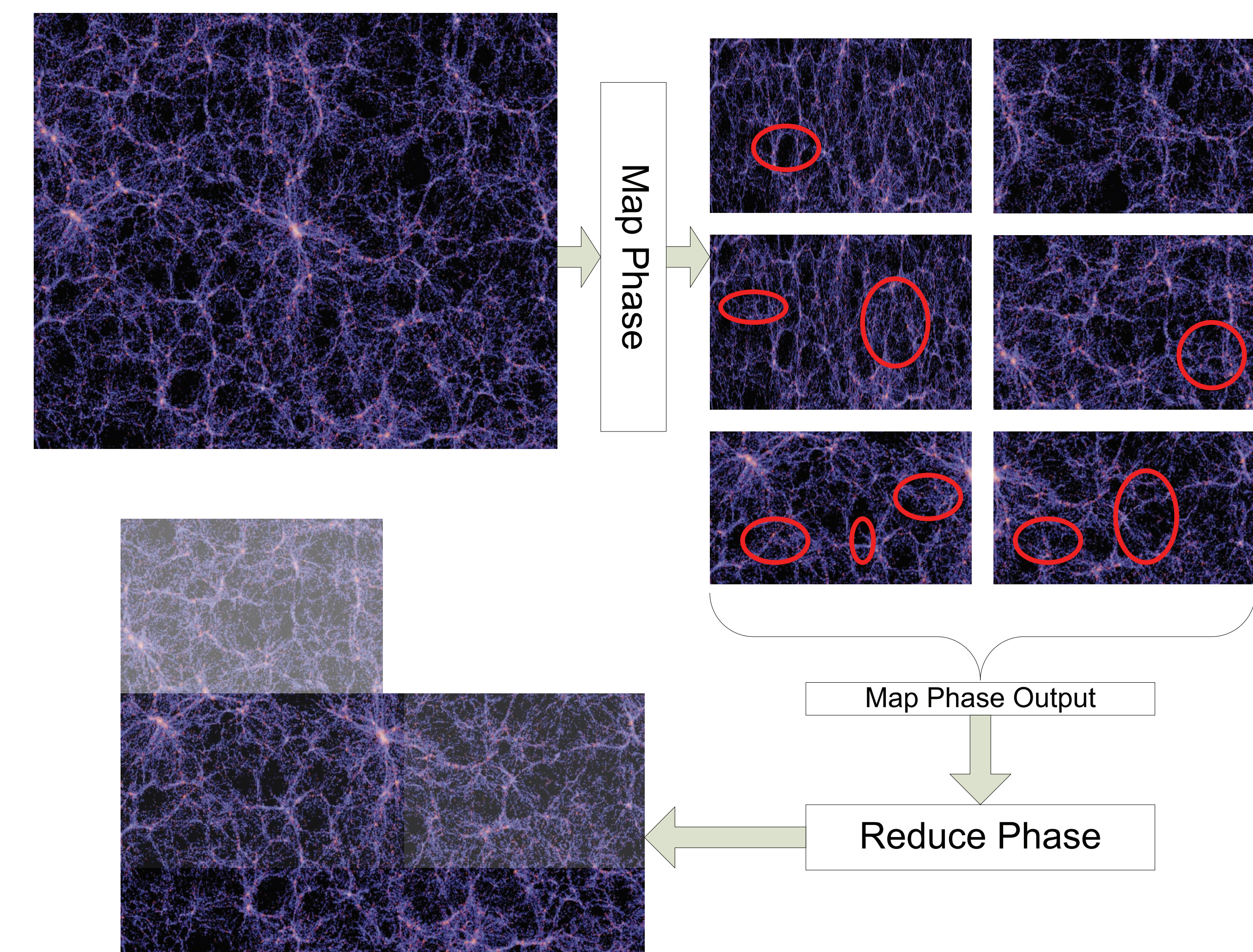
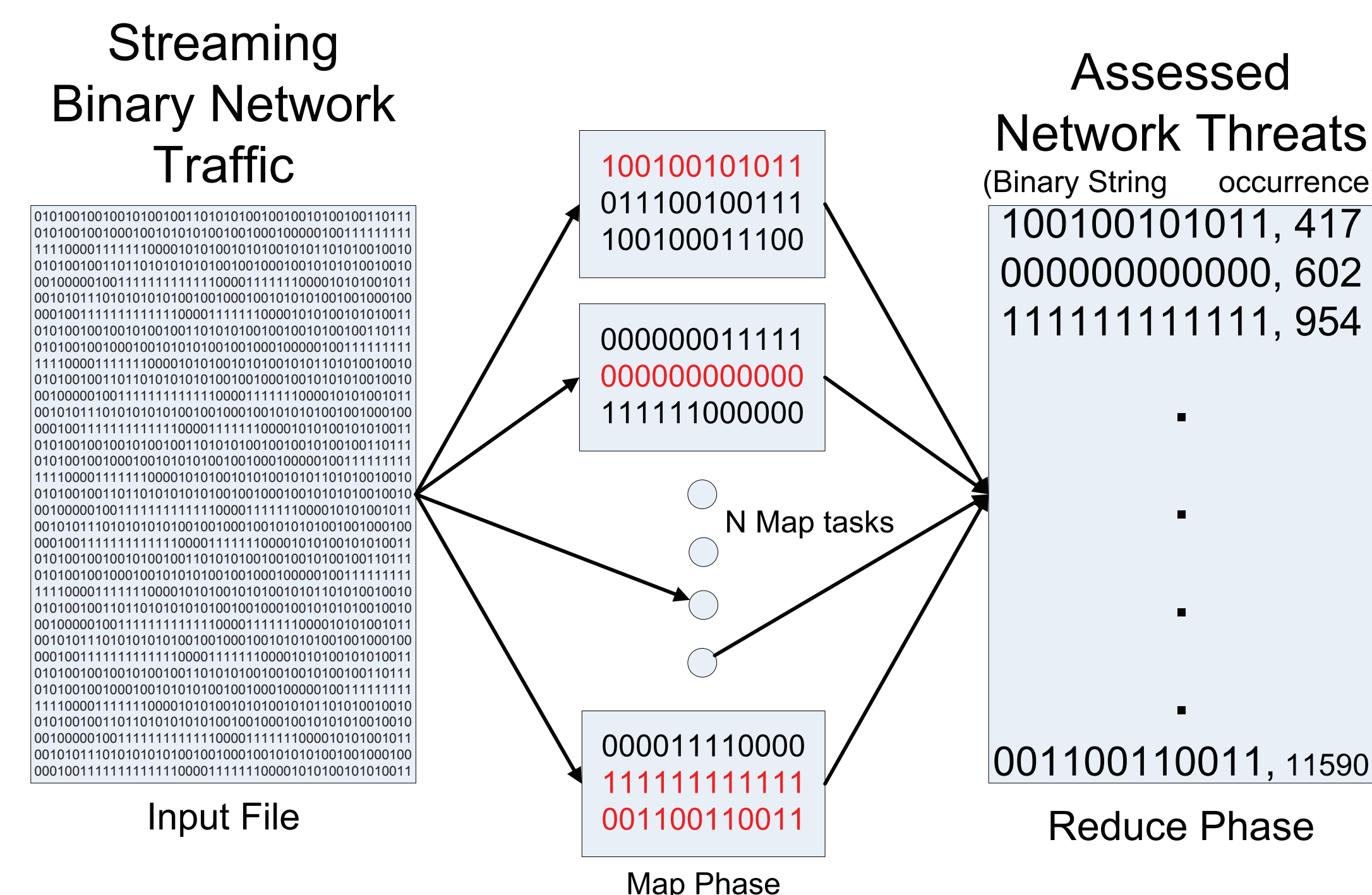
DISC Overview

Data Intensive Supercomputing (DISC) offers an alternative solution to scientific applications that are massively data bound rather than computationally bound.



split among multiple TaskTrackers. Threats are identified and recorded in each map task.

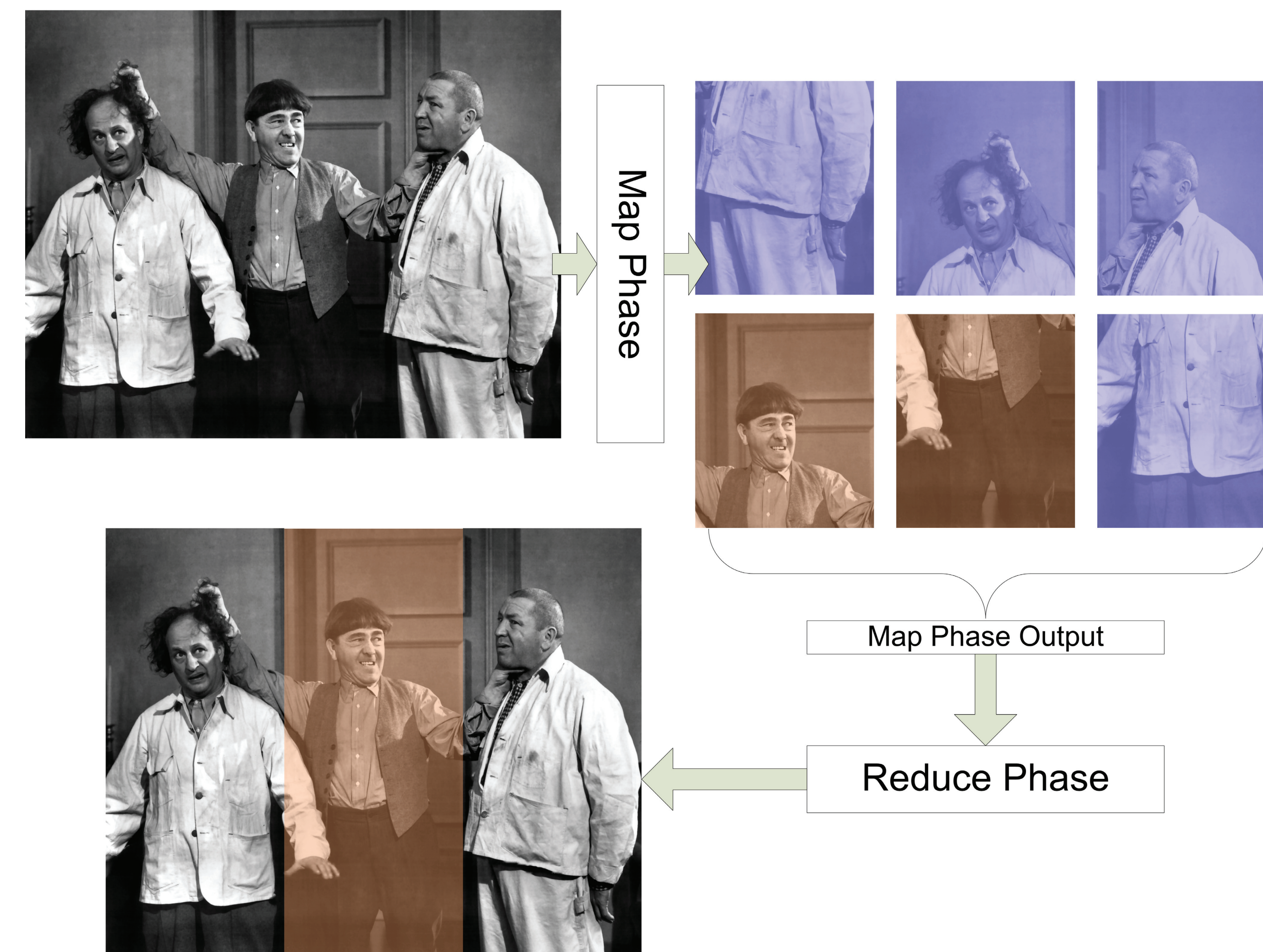
Reduce Phase: The malicious activity identified by the map tasks are collected into one file and delivered to cyber-security personnel in real time



DISC For Cosmology

Cosmology simulations have the ability to generate petabytes, even exabytes of data. Once the data is generated it must be processed in order to provide useful knowledge; which is a heavily I/O bound operation. One such cosmology application, uses a clustering technique to highlight interesting regions of the simulation space. Shown in the figure:

- Map phase: Simulation area is divided into smaller datasets and regions of interest are identified
- Reduce phase: Simulation space is reassembled into one file with areas of interest highlighted for the user.



- Reduce Phase: The image is reassembled with the identified threat areas easily identifiable to the user.

While this example is simplistic, this idea can easily be expanded to something more complex like threat analysis on GIS data or even real-time video, which becomes much more data-intensive.

DISC For Cyber Security

A present and mounting concern at the lab now is Cybersecurity. One way to insure lab security is to analyze all lab network traffic. With so much I/O traffic generated on a daily basis, analysis becomes very data-intensive.

Map Phase: Network traffic is streamed into the application and

DISC For Image Processing

This example of image processing takes a photo and identifies potential threats.

- Map Phase: Similar to Cosmology, the image is chunked into smaller regions. These regions are then processed to identify potential threats.

We would like to acknowledge the contributions of the following people who assisted with the success of this project:
Frank Alexander (INST-OFF), Parks Fields (HPC-5),
Salman Habib (T-2), Laura Monroe (HPC-5).

LANL LA-UR: 09-04881