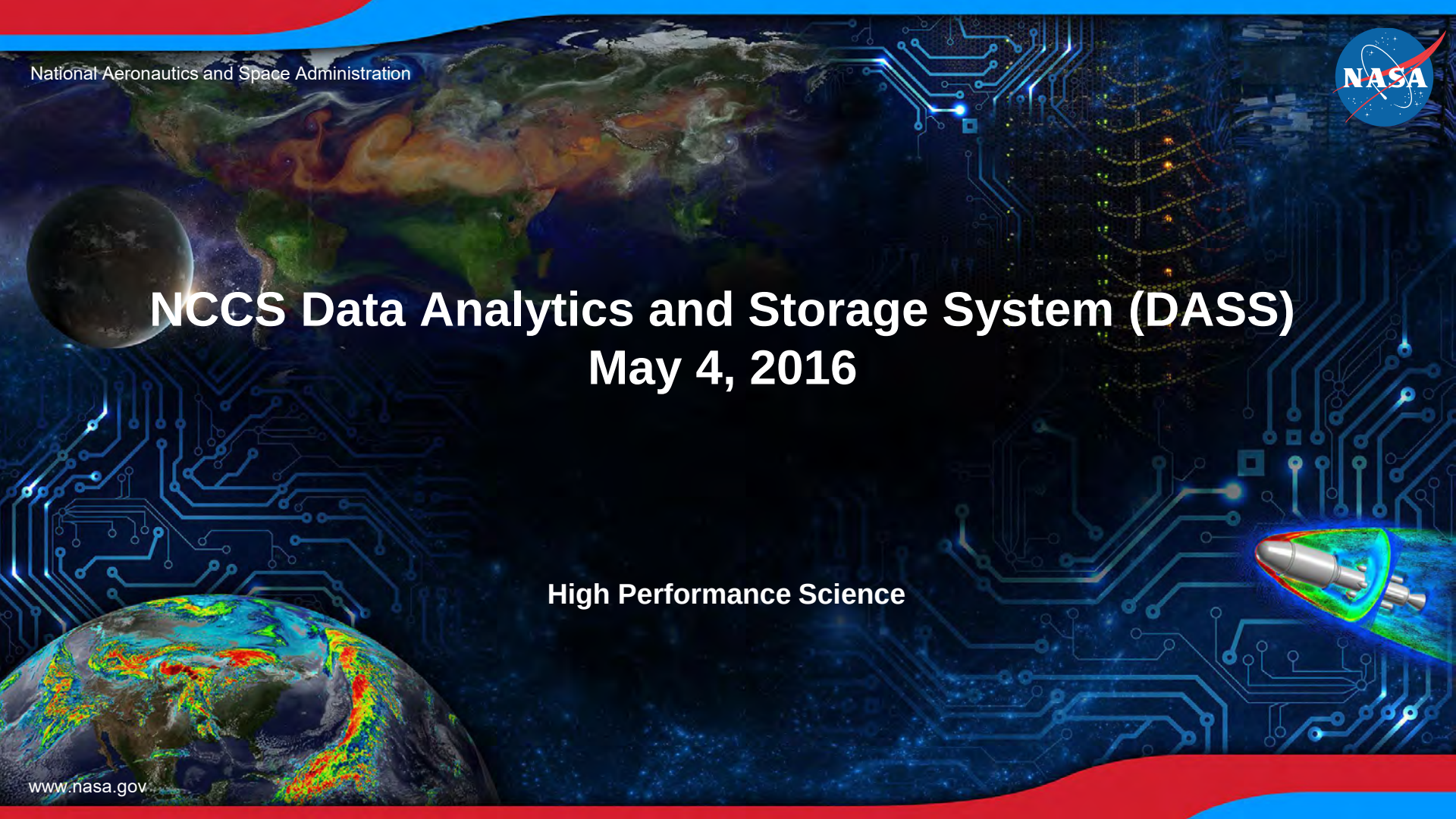
National Aeronautics and Space Administration

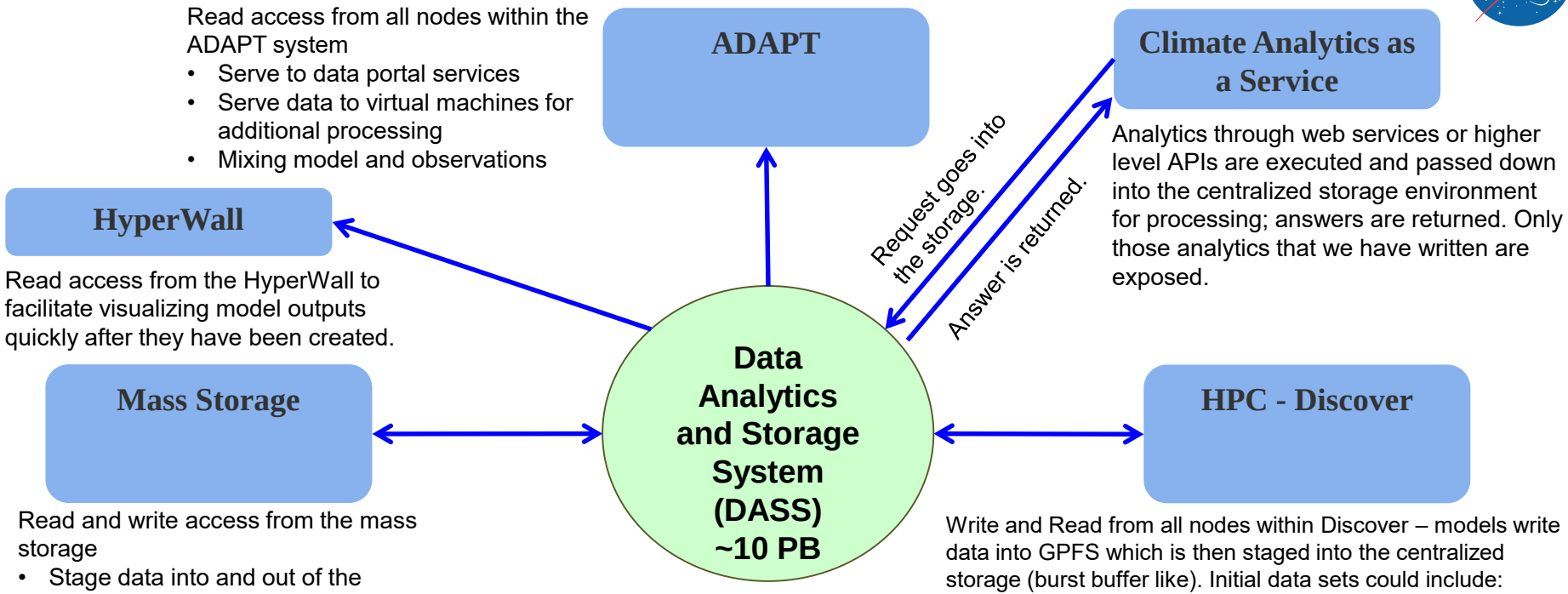


NCCS Data Analytics and Storage System (DASS)
May 4, 2016

High Performance Science

www.nasa.gov

DASS Concept



Read access from all nodes within the ADAPT system
- Serve to data portal services
- Serve data to virtual machines for additional processing
- Mixing model and observations

Analytics through web services or higher level APIs are executed and passed down into the centralized storage environment for processing; answers are returned. Only those analytics that we have written are exposed.

Read access from the HyperWall to facilitate visualizing model outputs quickly after they have been created.

Read and write access from the mass storage
- Stage data into and out of the centralized storage environment as needed

Write and Read from all nodes within Discover – models write data into GPFS which is then staged into the centralized storage (burst buffer like). Initial data sets could include:
- Nature Run
- Downscaling Results
- Reanalysis (MERRA, MERRA2)
- High Resolution Reanalysis

Note that more than likely all the services will still have local file systems to enable local writes within their respective security domain.

What are we doing to get there?

- **The NCCS is interest in POSIX compliant Object Storage so the following options are being evaluated**
  - HDFS to establish a baseline
  - Cloudera with the GPFS HDFS Transparency connector
  - Lustre with the Hadoop Adapter for MapReduce/Yarn (HAM) and Hadoop Adapter for Lustre (HAL)

DASS Software Stack