

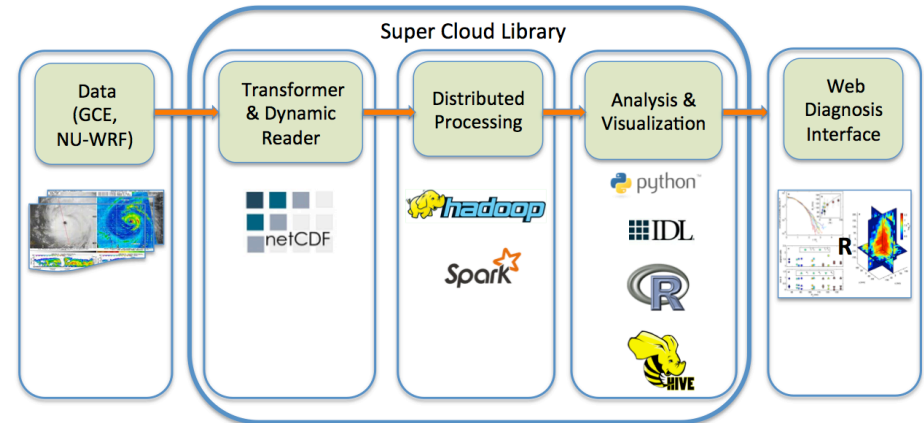


# Super Cloud Library (SCL): Empowering Data Management, Diagnosis, and Visualization of Cloud-Resolving Models (CRM) with Spark and Hadoop

PI: Wei-kuo Tao, NASA GSFC

## Objective

- Improve CRM output usability by science community:
  - Accelerate visualization of CRM output
  - Enable intercomparisons among runs of large volumes of CRM sensitivity experiments with observational measurements
  - Diagnose key parameters for cloud-precipitation process
- Demonstrate the value to distribute, visualize, analyze and intercompare CRM output and data with Goddard Cumulus Ensemble (GCE) and NASA Unified Weather Research and Forecast (NU-WRF) models using National Center for Climate Simulation (NCCS)



System Architecture of Super Cloud Library

## Accomplishments

- Demonstrated 20x speedup over original process, replacing a labor intensive series of manual steps with much file conversion
- Demonstrated system effectiveness and utility sufficient to warrant request by GPM Project Scientist to test drive system in accessing GPM data
- Eliminated one copy of very large data files (TB-scale), reducing clock time to make copies and conserving disk space
- Developed workflow process for populating data system from simulation output
  - Developed efficient and flexible methods for building HIVE tables from various simulation outputs
  - Developed flexible, scalable methods to adaptively subset data with a user specified criterion such as maximum upward wind speed
  - Developed a tool to convert the subset data into a NetCDF file
- Developed a web interface to subset, analyze the subset data, convert the subset data into a NetCDF file, and download it.
- Developed an effective, fast and efficient interface to visualize CRM data for use by research community

**Co-Is/Partners:** X. Li, Morgan State University; X. Sun, Illinois Institute of Technology; T. Matsui, University of Maryland; S. Zhou, Northrop Grumman; D. Duffy, GSFC

TRL<sub>in</sub> = 2      TRL<sub>out</sub> = 5