



# **Data Processing , Product Generation and Distribution at the NWS National Centers for Environmental Prediction**

Dan Starosta  
Production Management Branch Chief  
NCEP Central Operations

13 June 2012

# The NWS National Centers for Environmental Prediction (NCEP)

---

Virtually all the meteorological data collected over the globe arrives at NCEP, where environmental scientists analyze this information and generate a wide variety of environmental guidance information. NCEP delivers national and global weather, water, climate and space weather guidance, forecasts, warnings and analyses to a broad range of users and partners

NCEP Central Operations (NCO) sustains and executes the operational suite of numerical analyses and forecast models and prepares NCEP products for dissemination.

# Current Computing Capability

---

## Location

- Primary
  - Gaithersburg, MD (IBM provided facility)
- Backup
  - Fairmont, WV (GFE NASA IV&V facility)

## Configuration

- Identical Systems (per site)
  - IBM Power 6/P575/AIX
  - 73.9 trillion calculations/sec
  - 5,314 processing cores
  - 800 trillion bytes of storage
- Highly Reliable / Highly Available
  - Minimum 99.0% Operational Use Time
  - Minimum 99.0% On-time Product Generation
  - Minimum 99.0% Development Use Time
  - Minimum 99.0% System Availability
  - Failover tested regularly

## Inputs and Outputs

- Processes 1.7 billion observations/day
- Produces over 15 million products/day

## Significance

- Where our Nation's weather forecast process starts for the protection of lives and livelihood
- Produces model guidance at global, national, and regional scales

### *Examples:*

- Hurricane Forecasts
  - Aviation / Transportation
  - Air Quality
  - Fire Weather
- **Ensures No gap in service**



# Operational Model Data Ingest

- A global suite of environmental data is input into models
  - Observations are ingested continuously | 24x365
- Observations summary
  - Satellite sources - ~1.7 billion observations are received each day
    - Operational geostationary and polar orbiters
    - Research satellites
  - Non-satellite sources — ~580 thousand observations are received each day
    - Surface reports
      - Land
      - Marine
    - Upper-air profiles
      - Land Soundings
      - Aircraft

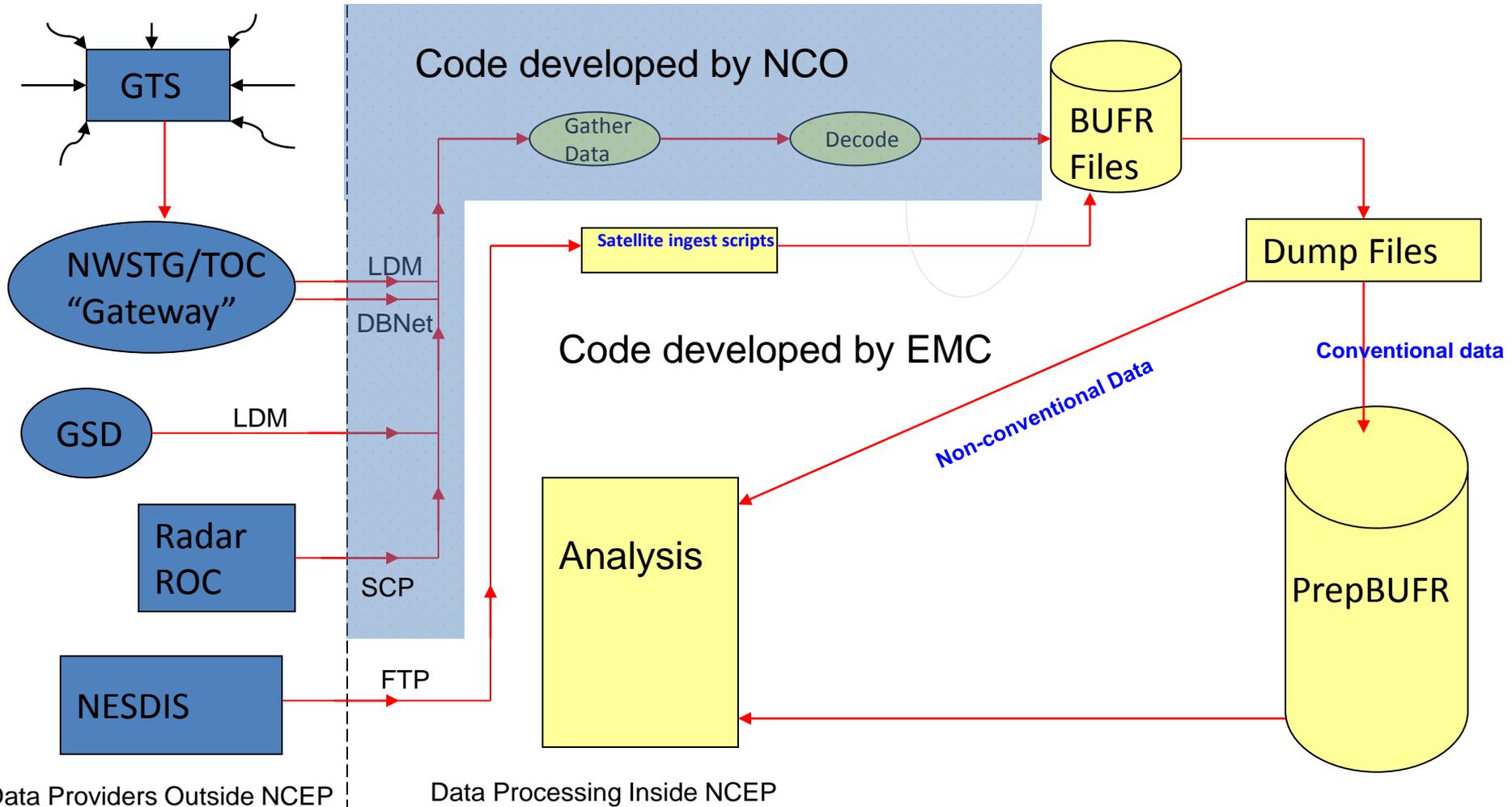


# NCEP Data Overview

---

- Ingest over 1.7 billion observations per day
  - Totaling about 6 TB per day
- Use about 9.25% of total supercomputer processing power to ingest, process and QC incoming data and disseminate output
- Generate over 5TB of raw output daily
  - Includes work files and experimental output
- Generate over 15 million operational products daily
  - Operational data volume is 1.5TB
- NCEP supported servers disseminate over 20 TB daily to millions of customers globally
  - Customers include other Government agencies, Commercial entities, NWS dissemination systems (to get to NWS Field offices), International Partners and U.S. public
- Archive 2 TB of data daily
  - Observations and output

# Data Ingest Flow



# Data Ingest

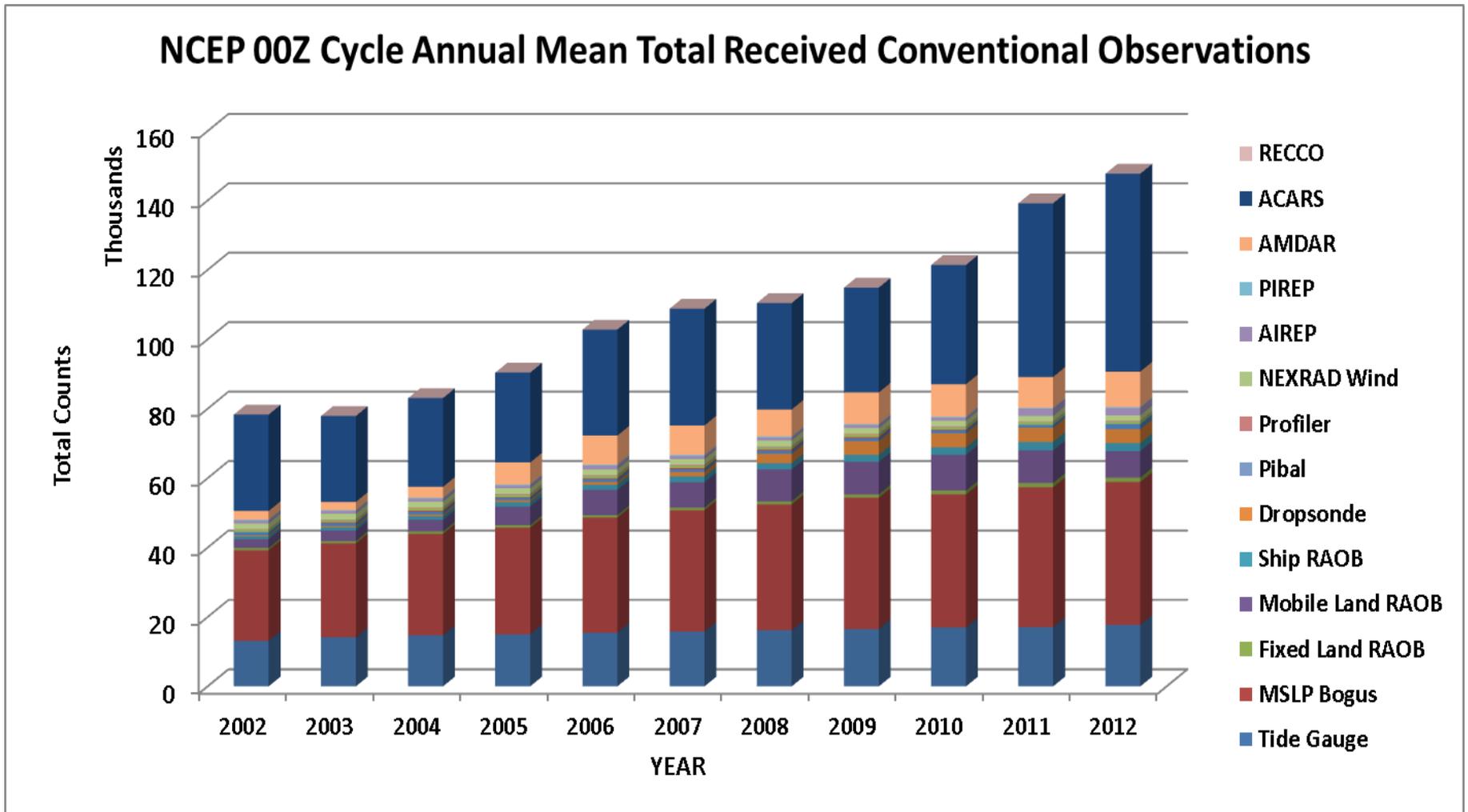
---

- Data is ingested from multiple sources (other State and Federal Government agencies, international partners, commercial providers)
  - Data is ingested onto both production and development systems independently
- Data is decoded, processed and quality controlled and prepared for model use
  - Only about 7.5% of raw ingested data is assimilated into the models

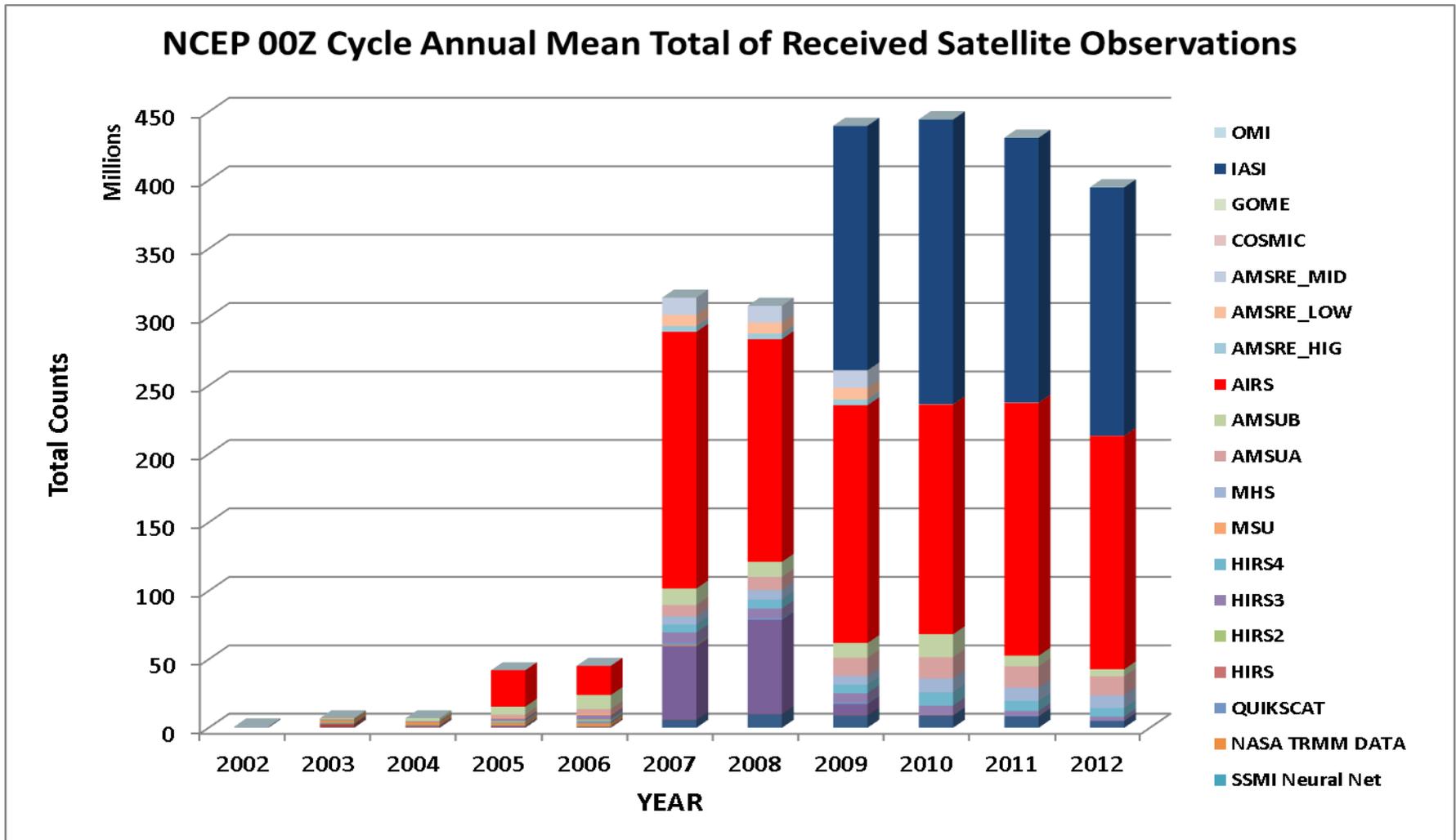
Total Number of Observations

<b>Summary Table</b>			
	<b>Total Received</b>	<b>Total Selected</b>	<b>Total Assim</b>
Non Satellite	576907	576907	576907
GEO	26738378	1767443	1561426
Polar	1709401167	48696568	11015435
Other Satellite	373582	342687	159507
<b>Totals</b>	<b>1737090034</b>	<b>51383605</b>	<b>13313275</b>

# Conventional Data Received



# Non-Conventional Data Received



# Product Generation

---

- NCEP routinely runs over 40 different custom scientific models
  - Typically, models run either hourly or every 6 hours
- Generate over 15 million operational products daily
  - Operational data volume is 1.5TB
- Products are tracked and compared to on-time baseline
  - Products created more than 15 minutes late are considered deficient
  - NCEP has 99.8% on time creation record over last 3 years
- Output is disseminated to :
  - NWS dissemination systems (for additional distribution)
  - NCEP Centers (for use in forecast generation)
  - NCEP dissemination servers (for customer access)
  - Backup Supercomputer (warm backup)
  - NCEP Archive (reanalysis and testing)

# NCEP Major Internet Dissemination Systems

---

- [FTP.ncep.noaa.gov](ftp.ncep.noaa.gov)
  - FTP and HTTP access to model guidance in file form as it is created on the super computer. Only the last 24 hours of output are available on this server.
    - This system disseminated 2.3 PB of data to customers worldwide in the last 12 months
- [MAG.ncep.noaa.gov](mag.ncep.noaa.gov)
  - The Model Analysis and Guidance website displays the graphical output of National Weather Service (NWS) Numerical Weather Prediction guidance
    - This is the newest of our dissemination systems and disseminated about 24 TB in the last 12 months
- [NOMADS.ncep.noaa.gov](nomads.ncep.noaa.gov)
  - Special access to specific model guidance products (FTP and Grads). Customers are able to select areas/parameters of a particular file to meet their needs (save bandwidth and processing at their site). Up to 30 day archive is available.
    - This system disseminated 2.1 PB of data to customers worldwide in the last 12 months

# The NOAA Operational Model Archive and Distribution System (NOMADS)

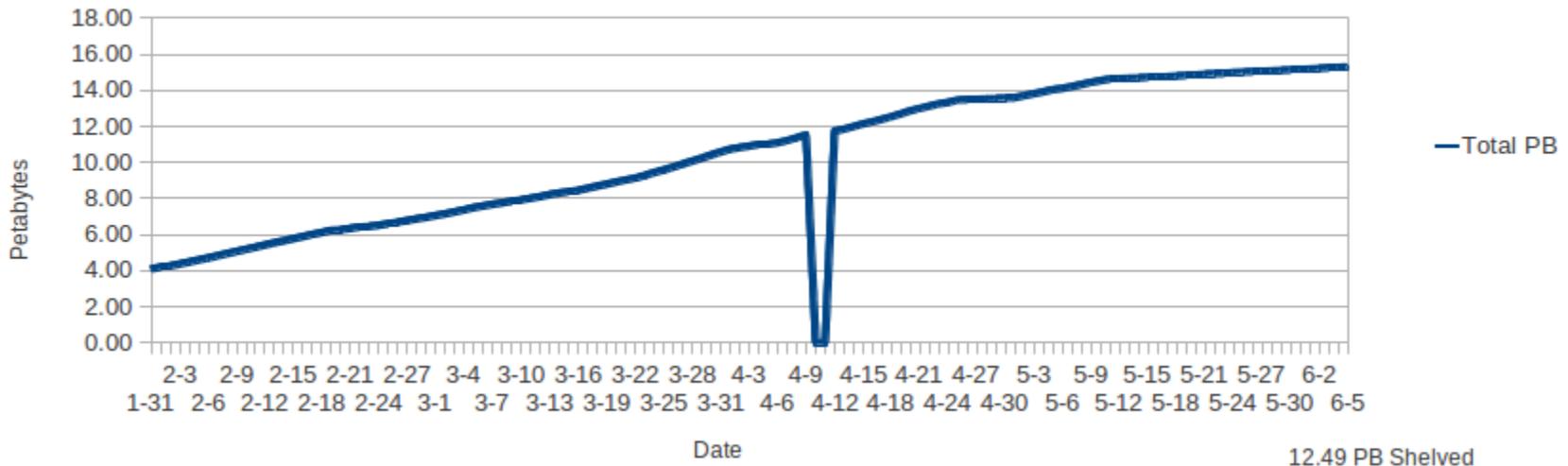
---

*Designed to provide real-time and retrospective access to climate, ocean and weather model data, and advance the integration of real time model data and applications responding to strategic guidance and NOAA's role*

- Pare down large file sizes of high resolution data and products.
  - Customers can get the meteorological data they need without downloading entire files
    - Leverage bandwidth, storage and processing
- Group different data sets to create needed products
  - Users can create their own products accessing only data they need
- Enable customer to subset the data:
  - in parameter space
  - in physical space
  - in temporal space

# NCEP Archive

- NCEP archives to tape about 2 TB of data (both observational and model output) daily.
  - Enable developers to compare/evaluate upgrades with previous guidance
- Archive categories are broken down into 1 year storage, 2 year storage and permanent storage
- We currently have over 15 PB archived



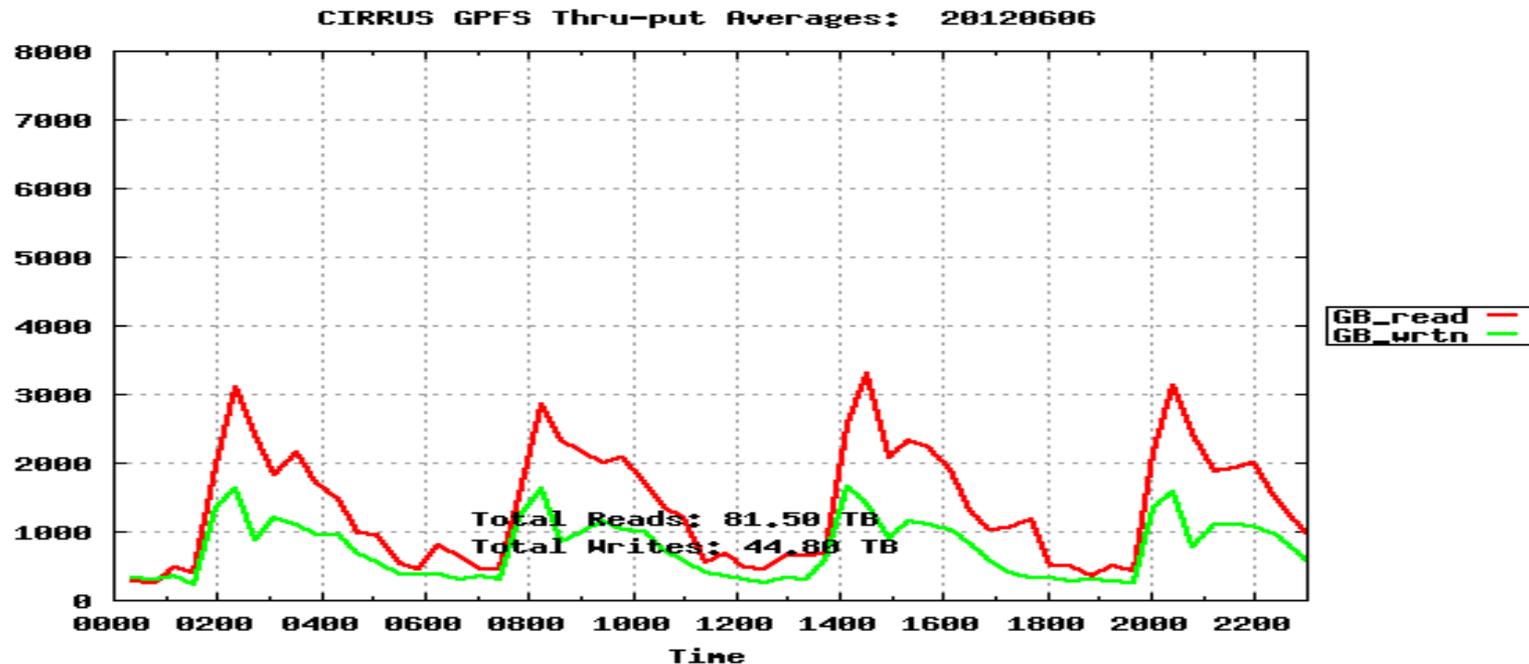
# Network Requirements

---

- 10 Gbps connections to each supercomputer (w/backups)
  - Production system has 6.5 TB of inbound flow and 34 TB outbound daily
  - Backup system has 25 TB of inbound flow and 6 TB of outbound daily
    - Data Ingest (both)
    - Shared storage (backup → production)
    - Data Synchronization (production → backup)
    - Data Dissemination (production)
    - Data Archive (production/write and backup/read)
- Either 10 or 1 Gbps links to partners
  - Data ingest and dissemination
  - Data synchronization to R&D systems
- DS-3 (45 mbps) connections to other NCEP centers
  - Model guidance to centers
  - Forecasts to Supercomputer

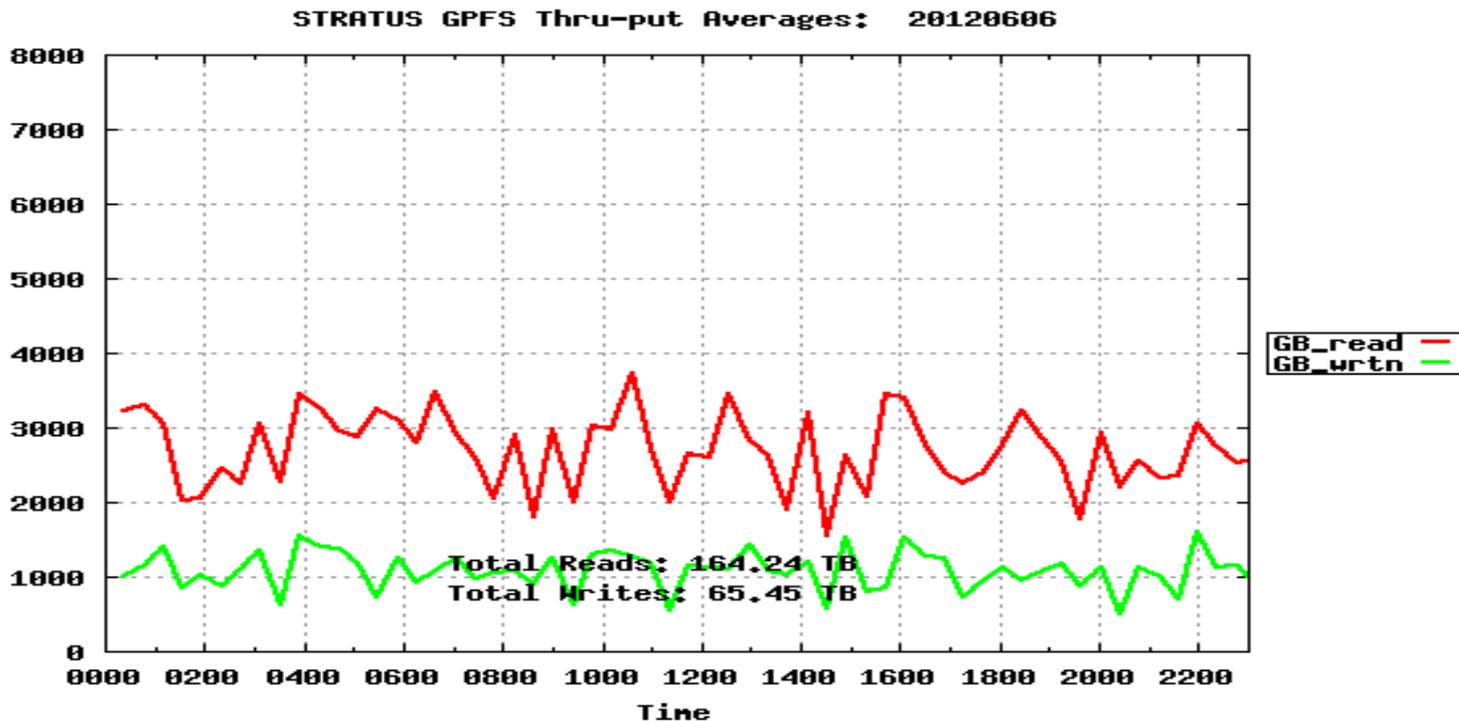
# Disk Usage

- NCEP uses IBM's General Parallel File System (GPFS) for Disk I/O on the supercomputer
- Backup system I/O profile
  - Can vary dramatically since this is a development system



# Disk Usage

- Production system I/O profile
  - Remarkably consistent in routine operations
  - In the event of a model suite problem doubling of thru-put rates is not uncommon



# Questions?



*“From the Sun to the Sea...  
Where America’s Climate, Weather, Ocean and  
Space Weather Services Begin”*

Daniel.Starosta@noaa.gov