

# Distributed Computing

Mark Govett

Global Systems Division

# Modeling Activities

- Prediction & Research
  - Weather forecasts, climate prediction, earth system science
- Observing Systems
  - Denial experiments
  - Observing System Simulation Experiments (OSSEs)
    - UAVs, GOES Sounder

# Modeling Development

- Increasingly complex, higher resolution
  - Local, Regional, Global, Climate
    - FIM Model – being developed at ESRL
      - 15KM global – 7 day forecast / run every day
      - 1000 PEs on wjet, 6-8 hours per run
      - Produces 1.3 TB total output per run
- Ensemble Modeling Systems
  - Probabilistic forecasts
- Data Assimilation
  - Higher spatial and temporal scales
- Systematic Model Testing & Verification
  - WRF NMM & ARW require hundreds of runs per experiment

# NOAA's High Performance Computing (HPC) Facilities



Boulder

West Virginia

Princeton

Silver Springs

Internet

Background topography courtesy Ray Stemer, Johns Hopkins Uni

# NOAA's Primary Responsibility: to manage and provide data or information



## Archive Data

- Model Development
- Model Verification
- Observing System Experiments
- Model Comparisons

## Model Data

- Local, Regional, Global
- Ocean, Atmospheric, etc

## Observations

- Expect a 100 fold increase in 7 years
- GSD will manage more global data than ever

# Data Access & Discovery

- Data Management Stovepipes
  - Difficult to locate and obtain data
    - Manually pull files from NCEP (eg. FTP)
    - Difficulty in running long simulations
      - Locate, store and stage data
- We cannot continue to store all the data
  - Requires data integration solutions
  - Standard solutions for defining, locating and accessing the data we need.

# Future Data Growth

- Global Earth Observation System of Systems
  - Needs technical solutions for system integration
- New observing systems
  - GOES-R, NPOESS, National Composite Radar
- High resolution models, ensembles

How can NOAA handle and effectively utilize these data?

- Support operational requirements
- Enable model research and development

*“End goal - A fully wired, networked and integrated system that provides for data processing, distribution, & archiving”*

- C. Lautenbacher, NOAA Administrator



# **The Grid: An IT Infrastructure for NOAA in the 21<sup>st</sup> Century**

*Mark Govett, Mike Doney, Paul Hyder*

*May 2004*

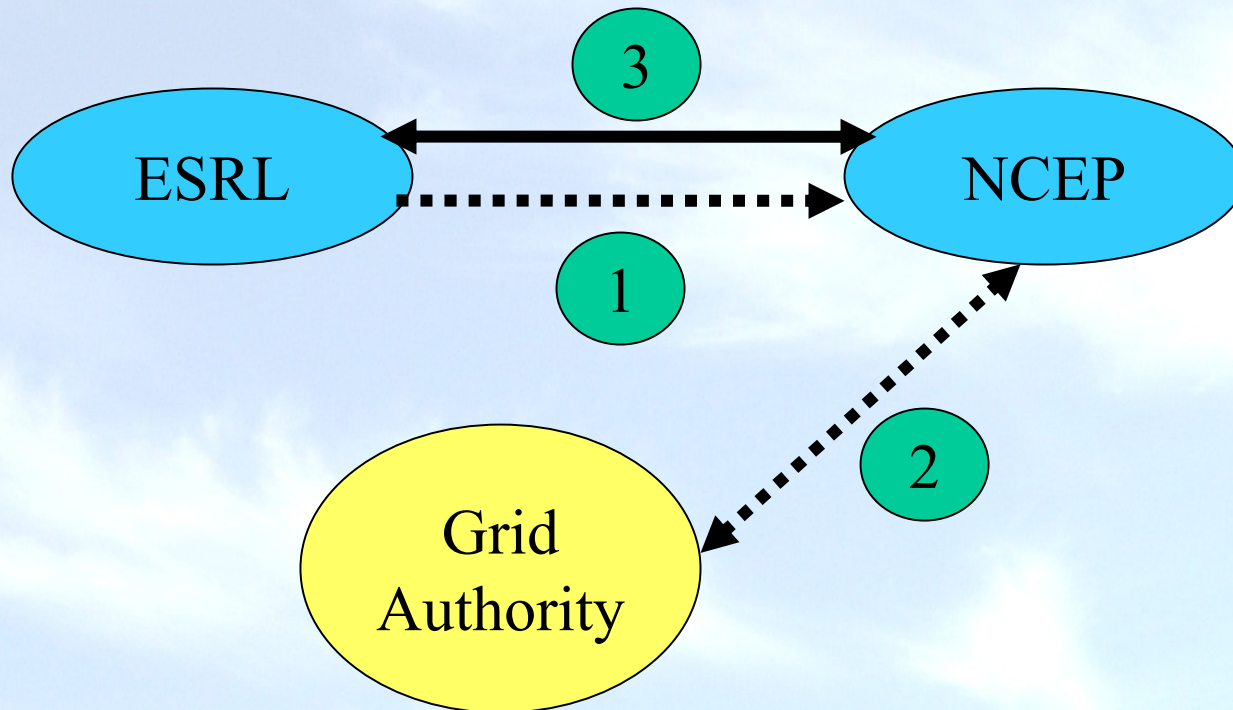
- **To meet challenges facing NOAA in the next decade and beyond**
  - **100 Fold Increase in Data Volume in 10 years**
    - **GOES-R, NPOESS, IEOS, Radar, GPS**
  - **More Complex Modeling Systems**
    - **higher resolutions, ensembles, data assimilation, more data**
- **Proposed an Integrated IT Infrastructure based on Grids**
  - **Build Compute Grids to link HPC resources**
  - **Build on existing data systems (e.g. CLASS, NOMADS)**
  - **Develop Dynamic Data Discovery, Access, Integration**
  - **Utilize & Develop Web Services, Grid Portals**

# Grid Computing

- Infrastructure
  - Links compute, data, network resources
- Compute Grids
  - HPC resources (eg. TeraGrid)
  - dedicated computers within an organization
  - desktop systems together (cycle harvester)
  - Vendor supported grid computing
- Data Grids
  - Link only data resources together
  - Easier to implement due to security issues

# Compute Grid Challenges

- Software was not sufficiently mature
- Limitations
- Security – trust relationships



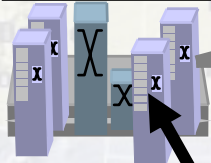
# Distributed Computing

- Computing
  - Push for single authentication mechanism at NOAA
  - Cross NOAA HPC job scheduling
- Data Access, Discovery & Integration
  - Web Services
  - Integration via Standards

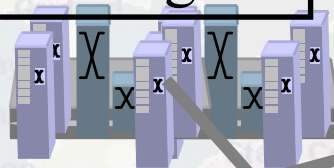
# NOAA's High Performance Computing (HPC) Facilities



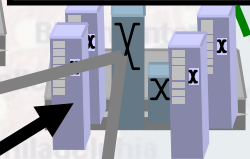
Boulder



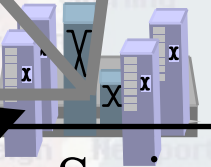
West Virginia



Princeton



Silver Springs



Internet

# Web Services

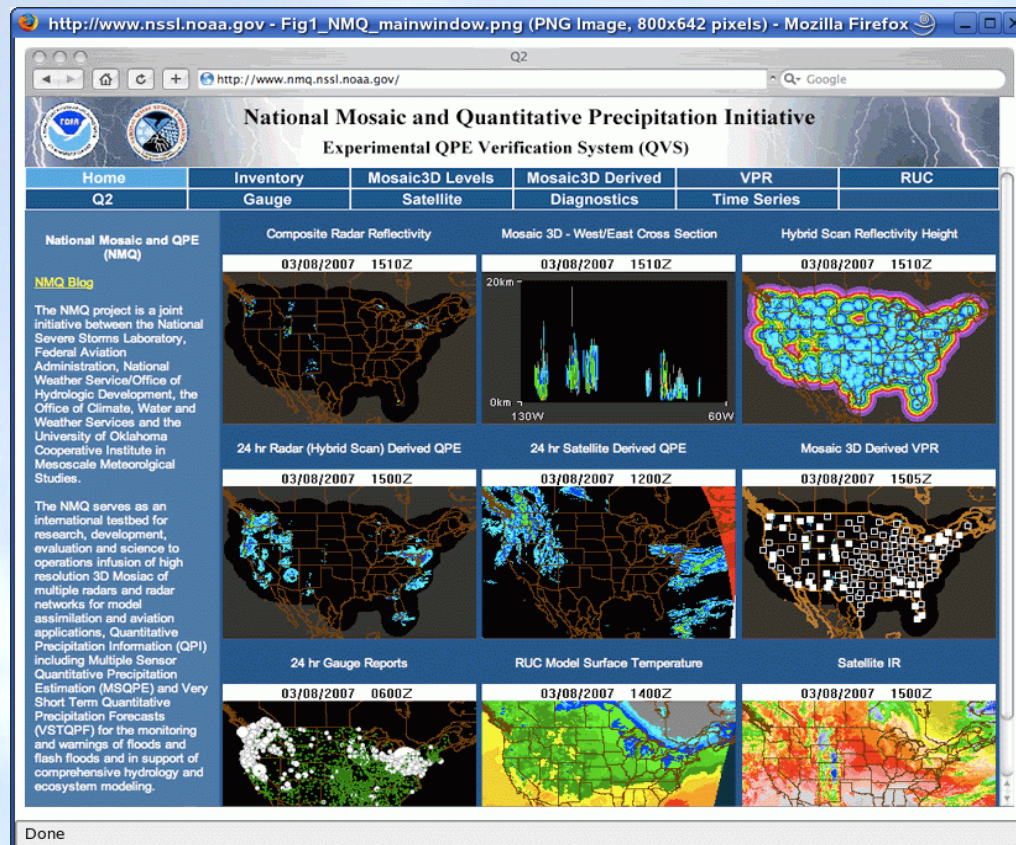
- A unit of work done by a provider to achieve desired end results for a consumer.
- Platform Independent
- Standard way to communicate between services, client applications
  - Access data, produce a forecast
- Link web services via a Service Oriented Architecture (SOA)

# Open Geospatial Consortium (OGC) Web Services

- Standards for data access
  - Web Mapping Service (WMS)
    - To access maps, images
  - Web Feature Service (WFS)
    - To access point or “feature” data
  - Web Coverage Service (WCS)
    - To access gridded data
- Widely adopted but deficient for Earth Science Community

# Web Coverage Service

- composite radar data national coverage every 5 minutes
- retrieve some portion of the data efficiently
- integrate with other data for verification, forecasting, etc







Questions?