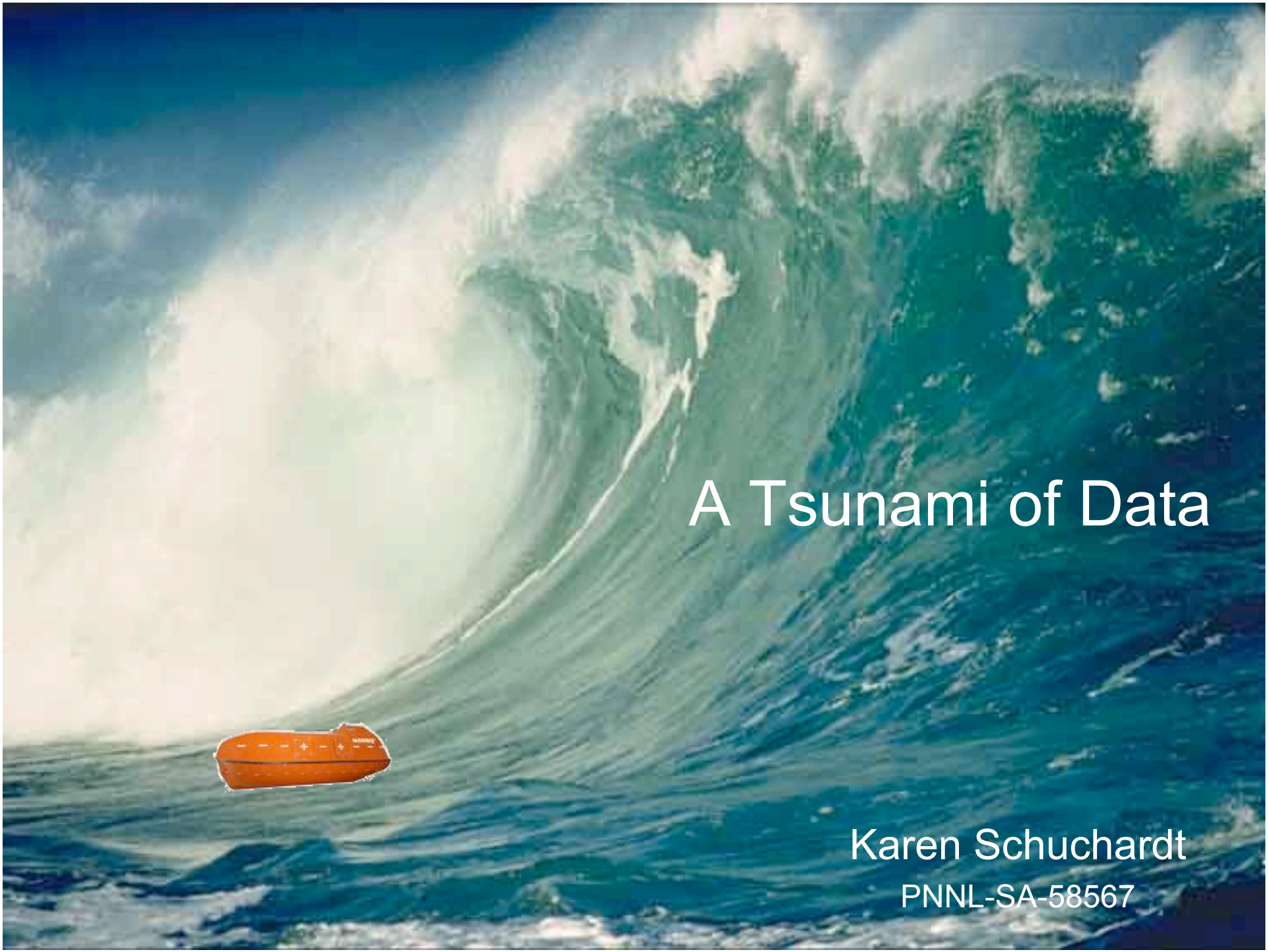

A Tsunami of Data
Karen Schuchardt
PNNL-SA-58567

The GCRM Tsunami

4 km, 100 levels, hourly data

~1 TB / simulated hour

~24 TB / simulated day

~9 PB / simulated year

2 km, 100 levels, hourly data

~4 TB / simulated hour

~100 TB / simulated day

~35 PB / simulated year

Other Data Tsunamis

- 30 TB/night: Large Synoptic Survey (LSS) Telescope (2014)
- 15 PB/year: CERN's Large Haydron Collider (May 2008)
- 1 PB over 3 years: EOS (Earth Observing System) data (2001)

How Big is a PetaByte?

1,000,000,000,000,000 (10^{15} or 2^{50})

2006: 1 PB Disk Array using 9 racks

2008: 1 PB about 4-5 racks

Typical Super Computer Centers: 100's TB disk on parallel file systems; 1-2 PB total

(CERN) going to 4PB disk

Ram disk;
Racetrack memory

Moving Data From Core to Disk

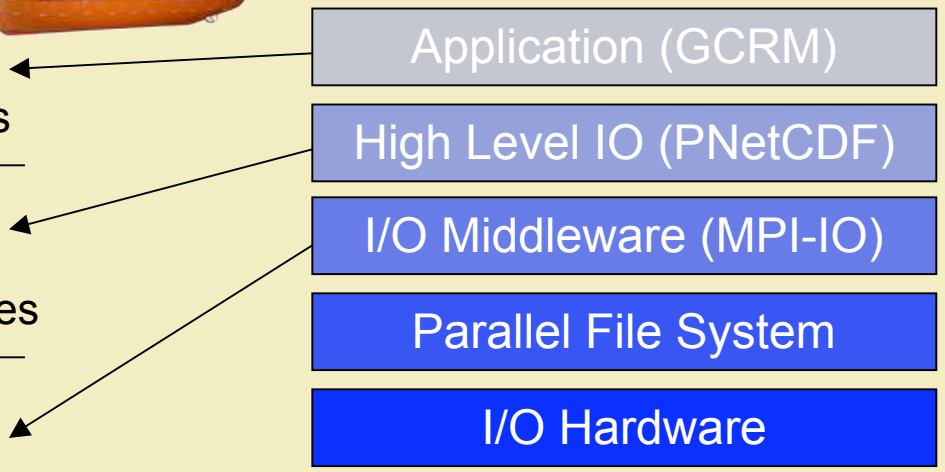
Traditional Approach - All data moves through processor
- blocks computation
- Not feasible with thousands of processors

Current Practice - Dump data in per/processor files
- Poor scaling behavior for tens/hundreds of thousands of processors?
- Expensive to merge data for analysis
- Extra step may introduce errors

- Non-standard; More work to program
- Easier to optimize block writes
- May impact inter-processor communications

- Immature technology
- Familiar programming model
- Could perform poorly due to many small writes

- Immature technology
- Hard to program
- Could perform poorly due to many small writes



Moving Data from Disk to Tape

If the model run generates more data than can be stored on available disk space, data must be moved to tape storage WHILE the simulation runs

150 TB Disk
~ User Alloc. 1TB

450 MB/s pk;
100 MB/s avg

22 PB HPSS;
100 TB cache

1 TB/hour

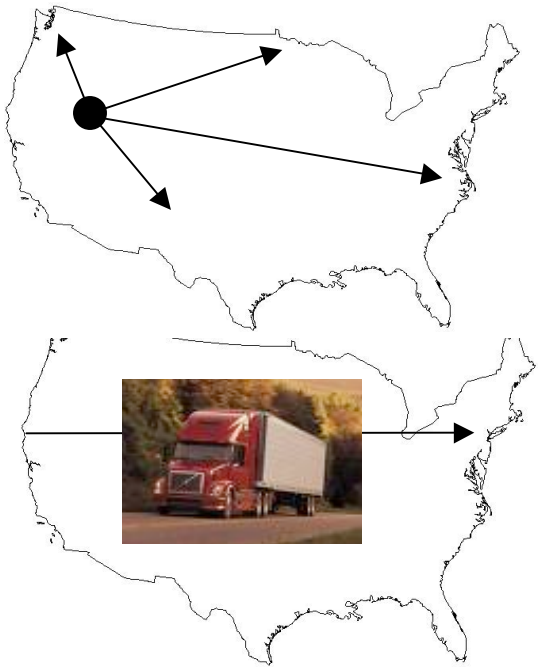
3.2 GB/s HPSS Theoretical Peak;
Larger disk cache; compression;
Memory technology advances

Moving Data to Users

How long would it take to move a PetaByte?

High Speed Networks (Internet 2 = 10Gbps)
TeraGrid throughput: 3 Gbps (last 3 days)[1]
Time: 30 days

SC07 Bandwidth Challenge: 8 Gbps (80%)
Time: 11 days



Typical university/lab connection (1 Gbps)
Typical throughput: < 50 MB/s
Time: 231 days

Brute Force
Time: 11 days (load/unload)
Throughput: 1 GB/s

Data organization changes;
Compression;
Data reduction services;
Moving computation to data

[1] http://www.teragrid.org/about/opinfo.html

Moving Data from Tape for Analysis

Assume user wants

1 variable, hourly data from a 2 month simulation (4km)

Data Set size is 23 TB

*HPSS to Disk Time: 63 hours

*Network Time (1Gbps link): 130 hours

Also dump low res or lossily compressed data to minimize tape access for many uses

* Assuming avg, not peak times

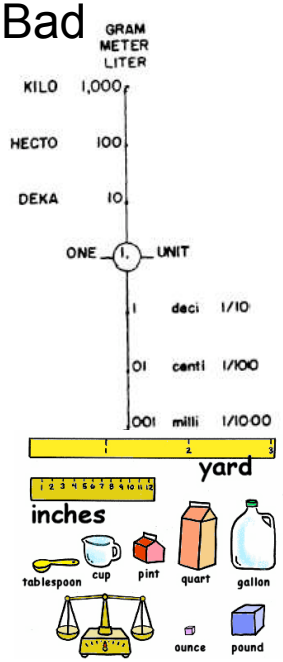
Model Validation/Comparison

Goal: compare model output with observed data or other model output

standards

Good

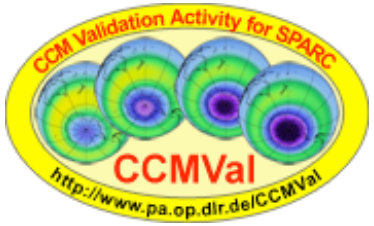
Bad



Climate and Forcast Conventions (CF)

Minimal conventions

- ✔enables end users to determine what is comparable
- ✔ facilitates processing and graphics by locating data in space time and as a function of other variables



. . .

But how much work is CF?

"Conventions have been developed only for things we know we need"

Grids

alternative coordinates, cell bounds, cell_measure…

Discovery metadata

Title, Institution,Source (e.g. model version)…

Data

Unambiguously identify a variable, assign units, associate data

Libcf – a library to make it even easier to conform to CF

Learn more: http://cf-pcmdi.llnl.gov/

The amount of data GCRM can produce will strain every aspect of computing: generation, storage and access, delivery, and analysis

- IO may be a limiting factor for computations on large numbers of processors (10-100K)
- Data from GCRM simulations at 4km or higher is too large to easily move between sites or to store at multiple sites.
  - Easy access to small subsets of data is critical to making simulation data available to the wider community
  - Analysis of entire data set must be done at location where data is stored
- Transfer from archive to disk may be limiting factor for analysis of large portions of data set
- Facility improvements are required but don't count on hardware to resolve all the problems

I’d like to encourage model developers to pay attention to CF standards (to minimize pain later) and to evolving best-practices for model output.