The Earth Science Big Picture

Jeff Dozier

DEC Review, UC Berkeley
March 3, 1993

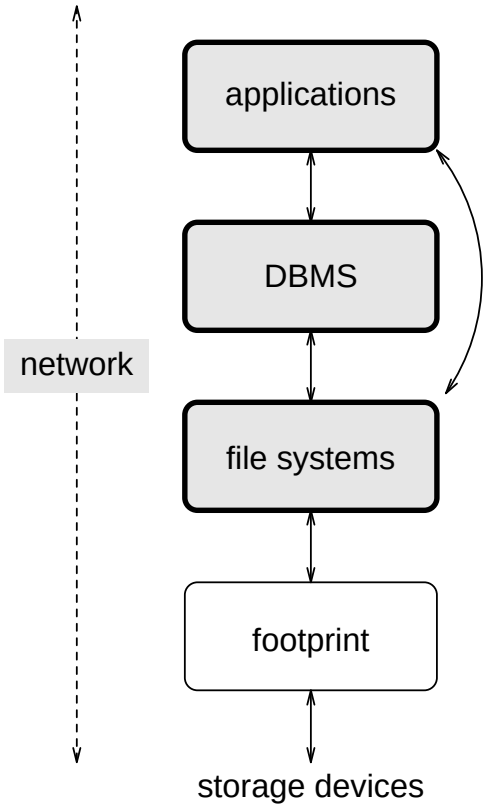
1 Global change researchers and Sequoia

- What do we want?
- What do we need?
- What do our computer science colleagues think we need?
- How do we make these the same?

2 How do we make these the same?

- Learn about the existing tools and the computer science research agenda.
- Teach our computer science colleagues and steer their research agenda.
- Implement the DBMS-centric view.
- Implement new tools and help with their development.

The Sequoia 2000 Layered Architecture



(data flow vertically)

3 What do we want?

A data and information system that

- helps us develop a better understanding of the Earth (better models and better guidance about what to measure);
- fosters interactions between observations and models;
- enables and encourages interdisciplinary research;
- makes us more productive.

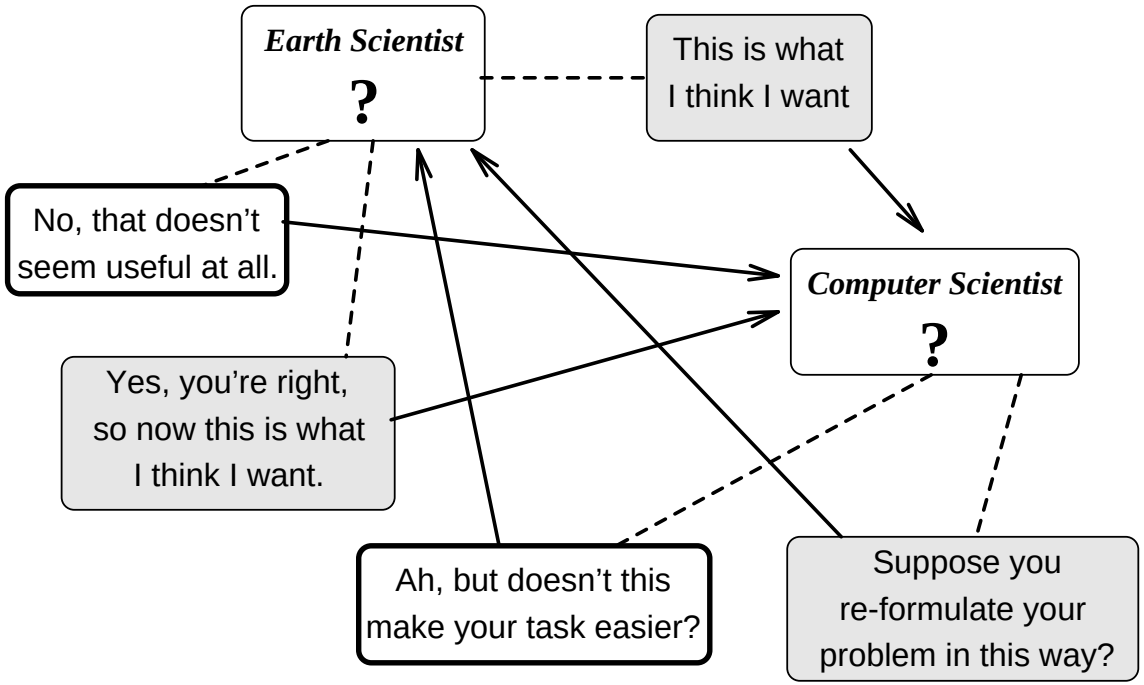
More specifically, a data and information system that

- allows easy access to large volumes of remote data
  - (Sequoia provides storage devices, networks, file systems, etc.);
- provides superior software tools for analysis and visualization of data, along with easy interchange of data among different software tools
  - (Sequoia provides data base management tools and a variety of visualization packages, but need to be integrated);
- enables collaboration among researchers in different disciplines and on different campuses
  - (Sequoia just getting started on this);
- enables painless, error-free application of algorithms to data
  - (Sequoia's recipe management agenda must be steered in this direction).

4 End-to-end Scenarios and the Visualization Agenda

These will help focus the agenda.

By providing a clear picture of (possibly primitive) processing steps, they enable iteration between the computer scientists and the global change scientists.



What's included in the Scenarios?

- selection of data to analyze;
- moving data to the computer where the analysis will take place;
- whatever format conversions are necessary;
- analysis of the data, including software packages and models used;
- evaluation of results, including comparison with other observations;
- writing results into storage;
- preparation of results for publication.

Three Canonical Scenarios

Use Sequoia 2000 hardware and software to:

- integrate general circulation model (GCM) [UCLA];
- analyze regional climate using surface meteorological and hydrological data [Scripps and DWR];
- interpret remote sensing data [UCSB].

Benefits

- a set of processing steps that allow us to build new needed functionality into Sequoia;
- a set of metrics that allow us to measure performance, to be synthesized into the Sequoia Analysis Benchmark (not just a Visualization Benchmark).

5 Showcase Projects

- Scenarios do not address the issue of inter-campus collaboration among researchers with different disciplinary backgrounds. Nor do they, by themselves, constitute "new science."
- Integrated "Showcase" project is being developed (Catherine Gautier, lead) that will use integrated research among Sequoia global change investigators.
- Each investigator will also extend his/her science with Sequoia 2000.