Some Data Systems: Too Big and Too Fuzzy
(Roy Jenne, May 29, 2001)

Some data systems may sound wonderful to many people. They promise to do almost everything. They are like a dream.

✧ These “dream” systems come along each few years.

✧ The systems are too big, too fuzzy, and have a poor history.

✧ But governments and companies may get “hooked” by them.

✧ They can adversely affect the handling of large datasets.

✧ They displace practical data plans that are faster, better, and cheaper.

Roy Jenne
Oct 2001
Some Data Systems: Too Big and Too Fuzzy

Roy Jenne
29 May 2001
Rev Oct 2001

A new "big idea" for data systems seems to come along each few years and is sold heavily. In 1993, the big idea was "client server" and it was sold as if it would solve all problems. I wonder why we do not hear much about it anymore. These "big ideas" often have more fuzz and hype than real content. The problem is that the grand ideas often lead to systems that are too big and problem-prone. The ideas can displace a good problem analysis that leads to quicker and better results.

This set has 6 main items with 27 pages and 2 pages in front.

1. Some people like the idea of a dream data system (May 2001, 1 p)

2. Some grid projects in Europe (March 2000, 3 p)
   - Almighty grid promises shortcut to knowledge; from Financial Times (Mar 2000, 1 p)
   - Europe joins race to turn Internet into one vast computer (Nature, 16 Mar 2000, 1 p)

3. Lost at sea (Federal Computer Week, May 2000, 2 p)
   - The Navy wants $16b for an intranet project called N/MCI.
   - Congress is wary of another DOD grand design IT project, a type with a history of huge cost overruns, poor performance, delays, etc.
   - Me: Congress is right to be worried.


5. Seven more papers about the "Grid"
   - Scientists weave new-style Webs to tame information glut (Science, 29 Sep 2000, 2 p)
   - The Grid projects (New Scientist, 21 Oct 2000, 1 p)
     - Data will come from the new CERN collider at a huge rate (starts 2005)
   - The big CS Grid projects (RJ, Oct 2000, 1 p)
   - Global Grid gets DOD not (Federal Computer Week, Sep 3, 2001, 1 p)
   - New DISS data Grid
     - A proposal to NASA (Feb 2000, 3 p here)
     - I do not recommend this.
   - The NSF teragrid project (Aug 2001, 2 p)
   - IBM tries to sell grid computing (Computerworld, Aug 6, 2001, 1 p)

6. Creating a digital earth (Federal Computer Week, July 17, 2000, 4 p plus 1 p)
   - The USA can actually achieve something very close to a "digital earth."
   - But it should not be done as one big project.
   - And it should not be standardized very much or coordinated too much.
     - And more pages
Some People Like the Idea of A Dream Data System

- The dream system is sold to accomplish almost everything as if by magic.

- The ideas are not concrete, but why worry about this? Don’t upset my dream!

- And the dream may pull in lots of funds
  - But the ideas are fuzzy.
  - And these systems have a bad history.

- Something different is needed to solve real data problems.
  - A data plan is needed that will really work.
  - It has to be possible to summarize how the systems will really work.
  - The ideas need to be practical ones that can result in good services within a sensible time period, and at a good price.

THE HARM FROM A DREAM DATA SYSTEM:

- It often displaces a good data effort that is stable, and which will deliver better results at a lower cost.

- So we give up our good solutions.

Roy Jenne
May 25, 2001
Some Grid Projects in Europe

- Handle data from CERN

Europe joins race to turn the Internet into one vast computer

Almighty grid promises a shortcut to knowledge

The successor to the internet could put unprecedented computing power at scientists' fingertips, says Clive Cookson

Note: CERN will have one of the world's biggest databases. This sort of planning can lead to big trouble.

Raf Jenne
March 2000
Almighty grid promises a shortcut to knowledge

The successor to the internet could put unprecedented computing power at scientists’ fingertips, says Clive Cookson

March 4, 2000
Financial Times
London

The scientists who brought you first the internet and then the web are now cooking up the “grid.” This new technology for connecting up the world’s computers will be millions of times more powerful than anything available today — and far easier to use.

“Imagine being able to sit at your desk and access all the information in the whole world to answer your question, in the format you need. Not just a list of links to relevant web sites or even selected online articles, but every piece of data ever collected on the subject, in a relevant, user-friendly way.”

That vision of the next internet generation comes from Britain’s Particle Physics and Astronomy Research Council, which is working with scientists from Cern, the European centre for particle physics research in Geneva, to take the first step towards a global grid.

The impetus comes from scientists who need to share previously unimaginable amounts of data between laboratories. This led to the creation of the grid, set up by the US Advanced Research Projects Agency in 1988, and to the worldwide web, invented at Cern in 1990.

Cern’s latest atom smasher, the $1.8bn Large Hadron Collider due to come on stream in 2005, will present the greatest computing challenge of the universe — thousands of billions of bits of computer data will have to be analyzed collectively by 5,000 scientists in 150 universities.

A new approach is needed and, after evaluating various ways industry is already using the grid, Cern has settled on the grid. American scientists, led by Ian Foster of the University of Chicago and Carl Kesselman of the University of Southern California, have been developing this concept over the past three years, and the US government is providing about $100m a year to set up several experimental grids.

The grid is a high-speed network that links supercomputers, databases, specialized processors and personal machines. It differs from today’s internet mainly through its “middleware” — programs that make collaborative computing much easier and more reliable.

“Just as the electrical grid enables you to plug in a device without worrying about power generators, the computing grid lets you extract information from anywhere in the world without knowing where it is,” says Jim Sadler, who is co-ordinating FPARC’s grid programme. “When you for...

Imagine being able to sit at your desk and access all the information in the whole world to answer your question

● Environmental scientists will model complex systems such as climate change.

A standardised grid, created for science, would quickly be adopted by industry, says Chris Jones, head of technology transfer at Cern. He sees a close analogy today with the situation at the end of the 1980s, when the US National Science Foundation laid down a standard “internet protocol” and then Tim Berners-Lee at Cern invented the web, with “hyperlinks” between documents. (The web took off commercially in 1993 when Marc Andreessen, the founder of Netscape, wrote a program that enabled untrained users to navigate by clicking on words or icons.)

"Public funding seeded the creation of the internet as we see it today and I think the time has come to spend more public money preparing for the next 10 years," says Mr Jones.

Bert Dekkers, head of IBM’s European internet laboratory in the Netherlands, says industry is already taking an interest in grid architecture. One oil company, for example, is considering it as a way to distribute the analysis of seismic and geological data around the world.

“The distinction between scientific and commercial computing is starting to blur, particularly for applications that involve remote visualisation,” he says.

A huge rise in the capacity of the communications infrastructure will be required to bring the benefits of the grid to the consumer market. But this is achievable. Mr Dekkers points out that mobile internet connections will be 300 times faster in 2003 than today.

A forerunner of the sort of distributed computing that is difficult to achieve today but would become commonplace with the grid is Seti@home. This enlists more than 1m personal computers worldwide to process radio telescope data, in the hope of picking out signals from an extraterrestrial civilisation.

The scientists at the University of California, Berkeley, who are masterminding the search for extraterrestrial intelligence, had great trouble devising a program that would use the internet to divide up the data, send it out to volunteers and then get it back. With the grid, anyone with a problem that captures the world’s imagination could command virtually infinite computer power.

Go to 2
Europe joins race to turn the Internet into one vast computer

Paris
CERN, the birthplace of the World Wide Web, aims to make Europe a leading player in the largest distributed computing project in history. The Large Hadron Collider (LHC) at CERN, the European Laboratory for Particle Physics, will produce a deluge of data. Now CERN is coordinating a proposal that would use the challenge of analysing all these data to help develop a concept known as the grid.

The grid idea was created in the United States (see Nature 402, suppl., C67–C70; 1999). Its goal is to develop software and Internet protocols to transform the Internet into a single gigantic computer. Researchers throughout the world could work on shared data sets on a network running thousands of times faster than today's best. Eventually, the technology could be applied to the public Internet.

The United States already has a $500 million five-year grid effort involving 50 research centres and coordinated by the National Computational Science Alliance. CERN now plans to submit a proposal to the European Union (EU) for a grid infrastructure costing 30 million euros (US$29 million). It has set itself a deadline of 10 May. The proposal is likely to be favourably received, as European Commission officials have actively solicited it.

David Williams, formerly CERN's head of computing and networks and now responsible for relations with the EU, points out that the web has made the physical location of information irrelevant, yet scientists still mostly use the Internet just to search for information. Real-time computing and data handling have hardly been explored, he says. John Taylor, director-general of the UK research councils and a convert to grids, has coined the term 'e-science' for this new way of working.

The CERN-led proposal will initially focus only on particle physics. But EU funding may be extended next year to other disciplines such as biology. Some European Commission officials see the challenge presented by the LHC's deluge of data as an opportunity for an all-out bid to catch up with the United States in constructing what many believe will be the successor to the web.

Andrea Dahmen, spokeswoman for research commissioner Philippe Busquin, says he "wants to see a reliable high-performance EU Internet network in place as soon as possible".

The $1.8 billion LHC, which will come online in 2005, is an ideal test bed for the grid concept. Collections of protons steered into head-on collisions will spew out around 7 petabytes (10^15 bytes) of data every year. The raw data from just one of the LHC's detectors represent the equivalent of every person on Earth talking into 20 telephones at once. The challenge is to transmit this information — requiring 1,000 times more computing power than CERN can currently deliver — in a usable form to thousands of users in more than 40 countries.

Under the CERN plan, data would fan out across a high-speed network to 10 national and regional data centres, and in turn out to hundreds of local centres and universities. The innovation will be the glue holding all of this together: a new suite of middleware, software and protocols designed to allow real-time distributed computing. The Internet would be made to function as if it were a single computer and database rolled into one.

"Getting ourselves liberated from the geographical constraints will be crucially important" for the success of the LHC, says Williams. "We need to use any and all possible resources to process the data. Not all of the people nor all of their computing and data-handling resources can be installed permanently at CERN."

Imagine an LHC scientist sitting at his or her desk in California, say, or Budapest. One click on a web wizard, and time is automatically reserved and purchased in real time from supercomputers and clusters of personal computers around the world. Another click, and data sets worldwide are scoured for all the Higgs two-photon events recorded so far. The interface invisibly converts all the datasets into a compatible format. One keystroke, and a menu pops up offering a suite of advanced visualization techniques that will allow the data to be analysed interactively with colleagues elsewhere.

Turning this dream into reality was the aim of a meeting at CERN last week between officials from the major European research councils, the European Space Agency and the European Bioinformatics Institute. The proposal that emerged would build on middleware developed from the existing Globus distributed software effort, a joint project of the Argonne National Laboratory in Illinois and the University of Southern California's Information Sciences Institute. The Linux operating system would be used throughout, says Williams, to ensure that software remains open-source.

By coordinating the national efforts that are being planned — Britain is likely to approve £100 million (US$158 million) in funding for grids later this year — the proposal could quickly be scaled up to include other disciplines and industry.

"It is clear that there are people in other sciences very interested in doing the grids together," says Chris Jones, CERN's head of technology transfer.

Support for the development of grids is also awakening within industry. Jones last week discussed the grid proposal with a delegation of British industrialists, and Williams is optimistic that concrete industrial support for the proposal will be in place before the 10 May deadline.
Congress is very wary –
- Another DOD grand design IT project
- A type often with bad costs and poor performance

LOST AT SEA

THE NAVY FACES ROUGH SEAS IN ITS QUEST TO CONVINCE CONGRESS IT NEEDS – AND CAN AFFORD – THE $16 BILLION INTRANET CALLED N/MCI

BY DAN VERTON

May 2000

Congress never saw it coming.
The Navy has pieced together a plan for what would be the government's most costly information technology project — a $16 billion intranet that would serve as the Navy's primary telecommunications link to 350,000 users on bases and ships deployed worldwide.

The so-called Navy/Marine Corps Intranet (N/MCI) would replace a hodgepodge of two dozen Navy and Marine Corps networks with a seamless network owned and operated by a single contractor. The Navy argues that it needs the intranet to prepare for network-centric warfare and to beef up information security on its computers.

But with a little more than a month until it plans to award the contract, the Navy is fending off increasing criticism of N/MCI. Congress — wary of another Defense Department grand design IT project, a type that has had a history of enormous cost overruns, delays and performance shortfalls — wants to know how the Navy can pay for N/MCI without asking for new money in the fiscal 2001 budget.

Last week, the Senate Armed Services Committee planned to attach language to the fiscal 2001 DOD appropriations bill that would require the Navy to follow a strict procurement management process for N/MCI, according to a source familiar with the markup. The Senate also may require the Navy to reintroduce N/MCI as a new program, which could delay the project for two years as the Navy obtains proper approvals for the project.

Rep. Curt Weldon (R-Pa.), who chairs the House Military Research and Development Subcommittee, said last week that N/MCI is a "very controversial" issue on the Hill, and, despite the threats of language being added in the fiscal 2001 Defense Authorization bill, "it will probably come down to a conference issue."

Federal unions have questioned the rationale and benefits of a large outsourcing, or seat management, contract such as N/MCI. And many of the Navy's front-line IT professionals have accused the Navy of using N/MCI to simply enforce standards, expressing concern that handing the entire infrastructure over to a single vendor presents serious security risks.

Acquisition professionals, meanwhile, have said the Navy would be foolish not to test the intranet concept in a small pilot project rather than risk losing N/MCI altogether if the global project runs aground.

Despite the criticism, the Navy plans to do what it has promised with N/MCI next month: transfer ownership of its entire IT infrastructure (including that which belongs to the Marines) to a single commercial vendor under what will be one of the largest, all-encompassing networking deals ever in government.

The Navy believes the naval intranet is not only affordable but necessary.

Sticking By Its Guns
Navy officials believe N/MCI is the only way they can fully leverage the so-called "Revolution in Military Affairs" — a term used to describe the advances in command, control and communications brought about by the use of high-tech computing and communications systems.

Simply put, the Navy has too many networks. And too many networks have created a Navy enterprise that is
efficient and too difficult to secure, and is riddled with interoperability and configuration control problems, according to the Navy.

The original N/MCI plan called for the contract to support 700,000 users at 300 bases in the continental United States and Hawaii for five years with a three-year option at a cost of $10 billion. The new estimate places the eight-year contract at $16 billion and the number of users at 350,000. "I think we were much more aggressive on the front end until we actually looked at what we had," said Ron Turner, the Navy's deputy chief information officer for infrastructure, systems and technology.

N/MCI will be mandatory for all Navy and Marine Corps commands, according to Turner. "It is our network," he said.

Even the Marine Corps, which had already made significant strides toward its own network by the time N/MCI came around, is on board with the program, Turner said. In fact, the Marines are particularly well positioned to be plugged into N/MCI because they already have a single enterprise Network Operating Center up and running at Quantico, Va., according to N/MCI briefing documents.

"It's like Ragu"

Still, the funding question has haunted the Navy from the beginning.

Congress is questioning where the Navy will find the new money to pay for a $16 billion project, and how the service will ensure the investment does not go to waste.

A "contract of this magnitude constitutes a major acquisition [and]...all budgeting guidelines must be followed," the House Armed Services Committee told the Navy in March in an official memo.

Rep. Herbert Bateman (R-Va.), chairman of the Armed Services Committee's Military Readiness Subcommittee and one of the most vocal detractors of the Navy's silence on the intranet deal, followed up on the memo by saying that he was "very concerned" about the lack of any new funding request from the Navy and the absence of any proof that the contracting approach is in line with congressional regulations.

The Navy's basic assumption for how it will afford the intranet is that it already spends enough money on IT every year to pay for N/MCI. The Navy's IT budget for this fiscal year is $2.1 billion, according to Federal Sources Inc., a McLean, Va.-based market research firm. But the department doesn't allocate the money smartly, creating an environment of technology haves and have-nots, says Dave Wennengren, the Navy's deputy CIO for the Year 2000 problem and information assurance.

Wennengren says the Navy can find the money in its existing budget. "It's like Ragu; it's in there," he said.

In fact, the Navy argues that N/MCI is cheaper than the alternatives. The Navy started down the N/MCI road after it looked at what it would cost to maintain, upgrade and refresh
Data Ideas During 1976 – 81

1. What were leading edge ideas, 1976 – 81
   ► These ideas had pizzazz
   ► Most data centers did not do this (lucky break)

2. The idea: Just load all data into a DBMS
   ► This will solve all problems
   ► And “make your life happy”
   ► Argue over which DBMS should be “standard”

3. The advocates were impervious to timing arguments
   ► People had to find the problems the hard way

4. The DBMS people finally found out
   ► You can’t transfer data in a DBMS format

5. Finally, reality set in (about 1981 – 82)
   ► But then people tried other big problem methods

6. The big NASA data systems came later
   ► They did not fall into this trap
   ► But they did cost NASA lots of grief and lots of money

Roy Jenne
Sep 2000
Rev May 29, 2001
DATA NETWORKS

Scientists Weave New-Style Webs To Tame the Information Glut

Physicists collaborating on a new generation of big experiments may drown in a data waterfall unless they find a way to channel the flow. A consortium of 16 universities has just received an $11.9 million federal grant to build a shared computational network, or data grid, that they hope will serve as the right sort of pipeline—and lead to even better science.

The idea behind data grids is to allow users to tap into a universe of electronic information, regardless of its location or origin. The grids are often compared to the popular music file sharing program Napster, which enables Internet surfers to exchange files. But Napster still relies on a central server to keep track of which music clip is on whose PC. A better comparison is a rival program called Gnutella, which allows users to share any file format in a totally decentralized system.

University researchers want to do the same trick with supercomputers and large data sets. To do so they've created a consortium, funded in part by the National Science Foundation, called the Grid Physics Network or GriPhyN (pronounced "griffin"). "GriPhyN will solve problems more demanding than any individual can solve," says Ian Foster, a computer scientist at Argonne National Laboratory in Illinois and co-principal investigator of the GriPhyN project. Biologists and medical researchers have also seen the value of peer-to-peer networking (see below) and want to make their data available over grids, too.

Right now, physicists can share big databases, but it is a nightmarish task. "We've been doing this for a long time, but it requires a lot of special expertise," says Fabrizio Gagliardi, a CERN physicist heading DataGrid, a European project that will join with GriPhyN. "Right now you have to know the exact locations and access procedures for each computer system." He compares it to e-mail 15 years ago: "When I was working at Stan-

ford, I had to log in to five different machines just to read my mail at CERN." Data grids will make global data sharing painless, Gagliardi says.

GriPhyN is arriving just in time to serve several large physics projects. Initially it will join the Sloan Digital Sky Survey (SDSS), the Laser Interferometer Gravitational Observatory, and two experiments at CERN, called ATLAS and CMS, that will run on the Large Hadron Collider (LHC). Each project offers the type of challenge that GriPhyN hopes to conquer: oceans of data that thou-

EXPERIMENTS SERVED BY GRIPhY N PROJECT

<table>
<thead>
<tr>
<th>Application</th>
<th>Data volume (terabytes/year)</th>
<th>Type of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDSS, 1999</td>
<td>10</td>
<td>Catalogs, image files</td>
</tr>
<tr>
<td>LIGO, 2002</td>
<td>250</td>
<td>Multiple channel time series, Fourier</td>
</tr>
<tr>
<td>ATLAS/CMS, 2005</td>
<td>5000</td>
<td>Events, 100 Gb/sec simultaneous access</td>
</tr>
</tbody>
</table>

Experiments served by GriPhyN project

sands of collaborators around the world must analyze to pick out painfully small signals from a noisy and cluttered background.

When the LHC comes online in 2005, for example, the collisions of its subatomic particles will generate a data stream of 5 petabytes every year. One petabyte is roughly equivalent to the capacity of a million

Download the Human Brain, With Security

Neuroscientists collect huge quantities of data on the human brain. But compared with their colleagues in physics, they are traditionally much less likely—for professional and personal reasons—to want to share them (Science, 1 September, p. 1458). Now a group at Rutgers University in Newark, New Jersey, is proposing a two-step, encrypted process for sharing information that would open the door to all legitimate researchers while imposing tough safeguards on its use.

The Rutgers group is writing software for a Napster-like Web site that would make possible "peer-to-peer" sharing of brain-scan images, but would not contain images itself. Instead, the site would house an index of available data sets and a protocol for accessing them. Researchers willing to share brain images would register with the site and describe what was available and under what condi-

The protocol would differ from Napster's, Bly points out, by letting donors control the release of data and keeping a record of each handshake. If a person claims to have data of a certain type but refuses to share it, "this protocol makes it immediately obvious," Bly says. He designs the software with encouragement from the National Institute of Mental Health and hopes to have a test or "beta" version ready by January.

—ELIO T. MARSHALL
personal computer hard drives or a stack of CD-ROMs nearly 2 kilometers high. Particle physicists have always had to deal with reams of data flooding from their detectors, but nothing like LHC. “The current experiments at CERN’s Large Electron-Positron collider generate a few terabytes per year,” says David Strickland, a Princeton physicist working at CERN. “The new experiments will create 1000 times more data than that”—data that thousands of collaborators will need to find and use.

The researchers also need massive computer muscle to crunch the numbers. In SDSS, for example, an astronomer trying to puzzle through a possible case of gravitational lensing might need to sort through 10 million galactic objects in order to find an effect, using sophisticated statistical wizardry and careful mathematical filtering.

“The resources required, for economic or political reasons, just cannot be created at any single location,” says Foster.

Physicists now spend heaps of time locating data files and getting them processed. In searching for images representing how new particles created in a collision were slamming around the detector chamber, for example, a researcher may have to punch in a horrendous chain of computer commands to translate the raw numbers into a useful picture of the results.

GriPhyN and DataGrid will also work together with the Particle Physics Data Grid (PPDG) at the California Institute of Technology (Caltech) in Pasadena to provide a connected computational lattice for big physics experiments. Harvey Newman, a physicist at Caltech, is a principal researcher on PPDG and also a senior member of the GriPhyN group. “We’d always planned to study data transfer and file caching in the short term, then build a longer life system,” Newman says. “Then GriPhyN came along.” Further afield, the European Commission’s Information Society Technologies program has just invited DataGrid to apply for EU9.8 million (about US$8.6 million) to build research grids in Europe.

The international culture of physics fosters such grand sharing, but pitfalls may loom. “We’re all nervous about this,” says physicist Paul Avery, co-PI of the project at the University of Florida, Gainesville. “My experience in large software projects is that unless you sit on top of this all the time, you do diverge.” Newman was also nagged by doubts initially after hearing people discuss the idea at a meeting. “Some people said, ‘We’ll build our national grids and then make them work together.’ But this does not work,” Newman says. “Fortunately we haven’t built anything yet, so there is a good chance that we all build the same thing.”

Strickland, who is not directly involved in the grid construction, says that the grid builders appear to be taking the right tack by funding software engineering rather than just buying lots of new hardware: “They seem to be throwing the right resources at the problem.” But that alone is no guarantee of success, he cautions. “Obviously, we’ve got a long way to go.”

—David Voss

**ANTHROPOLOGY**

**Misconduct Alleged in Yanomamo Studies**

E-mail has been ricocheting among anthropologists as they nervously await the publication of a new book that charges some prominent researchers with professional misconduct—and much worse—in their studies of the Yanomamo, a native people in the Brazilian and Venezuelan Amazon. Written by journalist Patrick Tierney, *Darkness in El Dorado* (W. W. Norton)—an excerpt of which is scheduled for publication in *The New Yorker* next month—accuses anthropologists of creating a false picture of the Yanomamo, manufacturing evidence, and perhaps setting off a fatal measles epidemic.

“This is the Watergate of anthropology,” says Leslie Sponsel of the University of Hawaii, Manoa. “If even some of the charges are true, it will be the biggest scandal ever to hit the field.”

Although few anthropologists have actually read the book, which will not be published until mid-November, it has already stimulated an enormous reaction. The American Anthropological Association (AAA) has promised to hold a session on the book at its upcoming annual meeting. Napoleon Chagnon, a prominent Yanomamo specialist now at the University of California, Santa Barbara, whose research is challenged by Tierney, has already refused to participate in what he calls “a feeding frenzy in which I am the bait.” (Instead, he is consulting libel lawyers.) Meanwhile, other researchers are recruit-
massive dinosaur bones in a museum in New York. Patients in remote areas could see a doctor. And once haptics—touch simulators—are built in, people could use tele-immersion to come together in even stranger ways. A woman in Europe could reach out and touch her newborn grandchild in the US.

But not yet. Problem is, today’s Internet can’t ship data fast enough. To look anything like reality, tele-immersion will have to be able to move mountains of data—spatial and visual information about people and their environments—across the Internet in a trice. Today’s 56-kilobit-per-second connections can’t do that. Even the bare-bones demonstration at Chapel Hill needed 60 megabits per second. High-quality tele-immersion will require even more—around 12 gigabits per second.

**Pioneering spirit**

Fortunately, that kind of capacity is on its way. Leading the charge is Internet2, a consortium of American universities, government agencies, private companies and international organisations that is trying to recreate the collaborative spirit of the early Internet. The group is building high-speed networks and developing software with the aim of transmitting data at speeds an order of magnitude faster than the Net does now. The project’s leaders say it is a unique test bed for Internet applications of the future, including tele-immersion.

In addition to high-speed networks, tele-immersion will require supercomputers to perform the trillions of calculations that are needed to portray environments in 3D. This kind of computer power would have to be on tap over the Internet. Something like that is on the way, too, in the form of a network called the Grid (see “Power sharing”).

In the next tele-immersion experiment, the teams will open two-way portals so that all three locations can see and hear one another. They will then try out a dummy collaborative project: rearranging the furniture in a doll’s house.

Eventually the researchers would like to make tele-immersion even more naturalistic, perhaps by jettisoning the headgear and glasses altogether. One possibility is to use a screen that transmits different

**POWER SHARING**

**HAS YOUR COMPUTER run out of processing power? Wouldn’t it be great if you could just tap into someone else’s. Anyone else’s. This is the idea behind the Grid. The name is borrowed from the US electricity grid, and the goal is to make it as easy to access processing power over the Internet as it is to plug a TV into the wall.**

Once the Grid is up and running, users will be able to call on undreamed-of resources simply by switching on their PC. Processing power, software and data from computers across the world—all this will be at their fingertips. And just like the electricity grid, it won’t matter where the juice actually comes from.

“A tremendous amount of computing power is sleeping when the US is sleeping, and when Europe is sleeping,” says computer scientist Fabrizio Gagliardi of CERN, the European particle physics lab near Geneva.

“If we could tap into that, we could do very powerful things.”

Gagliardi is head of the Datagrid project at CERN, a grid that links computers at 21 science and industry centres throughout Europe. Scientists at CERN need the grid because in 2005 when they start smashing particles together in their new accelerator, the Large Hadron Collider, raw data from the collisions will pour out at a rate equivalent to a million feature-length movies a second. Processing this data would swamp the CERN computers, so the researchers need a way to share the burden. The Grid will help them to distribute the data to computers across the continent.

Computer scientists in the US are working on a Grid of their own. The Grid Physics Network (GriPhyN) links 16 universities and will be used to crunch data from four major physics projects, including two Large Hadron Collider experiments. Last month, the National Science Foundation awarded 11.9 million to the universities of Chicago and Florida to pursue the project. GriPhyN should be up and running by 2005.

Managing shared resources isn’t easy, and the developers’ immediate task is to develop a layer of software called “middleware” to keep order on the grid. “The middleware is the guts of it. It is the broker that goes out and finds what is available,” says Chris Jones, head of technology transfer at CERN. This software also ensures that all the computers are speaking the same language. “It allows me to identify myself,” says Gagliardi. “It establishes a secure connection, so I can be charged. It asks: where are the resources? It looks for the best way to do the job. It will also predict how long the job will take, and how much it will cost.”

It is possible to gain access to other computers at the moment, but incompatible systems and languages, bureaucratic rules and suspicious network administrators make it complicated. A pervasive Grid would change all that. “All the authentication, looking for appropriate data, software, space and payment will happen automatically, without us seeing it,” says Carl Kesselman from the University of Southern California in Los Angeles. Kesselman, along with Ian Foster from the University of Chicago, introduced the concept of the Grid in 1996. Both are now working on GriPhyN. In ten years, they say, we will all be part of one giant, global Grid.

If they’re right, it won’t be long before we can access the Grid at home, simply by installing middleware on a PC and linking up to an ultra-fast network. But what would we use it for?

“We are still trying to realise what the technology is really good for,” Kesselman admits. Foster sees working or playing together as the killer applications. “What people really like to do is communicate,” he says. “It is those applications, for example interactive gaming or tele-immersion, that will really take off.” Such activities require huge amounts of processing power and will only be possible if you can siphon it off the Grid.

Joanna Marchant

21 October 2000

New Scientist • www.newscientist.com

*The Grid idea started 1996*
The Big CS GRID Projects

- The main ideas:
  - Do much processing at distant sites
  - Use fast networks

- Plans:
  - Apply this to biggest datasets in world

- Status:
  - The physics community has bought into this (I think)

But

- It has a very fuzzy feel.
- So be very careful.

Two refs:
Magazine          Date        Pages
Science           29 Sep 2000  2250 - 51
New Scientist     21 Oct 2000  54

Roy Jenne
A global grid for DOD

- Does someone know the history of this?
- Does it try to integrate too much?
- Will it hurt the more direct approaches to solving IT problems?

- Roy Lema

Global grid gets DOD nod

BY CHRISTOPHER J. DOROBEK

The Pentagon has approved its first enterprise architecture. The Global Information Grid architecture Version 1.0 is a worldwide architecture for providing data to military forces around the world, from regional commanders to soldiers on the front lines, said Margaret Myers, the Defense Department’s acting deputy chief information officer.

“This is something that we’ve never really had before,” she said.

The architecture will provide the first take on an integrated DOD enterprise information technology architecture, Myers said Aug. 29 during a forum sponsored by Federal Sources Inc. DOD has already started working on Version 2.0 of the GIG architecture, she said.

“The department is coming up with innovative ways to use that architecture, she said. “It should allow DOD to acquire IT faster. It could have an impact on how we acquire information technology.”

The GIG architecture includes an operational view, which DOD officials said describes and connects the operational elements, tasks and activities, and information flows required to accomplish mission and business operations.

Meanwhile, the system view describes and associates the systems and how they interconnect to the operational view and its requirements.

FEDERAL COMPUTER WEEK

Sept 3, 2001

Page 1
NewDISS Data Grid

Alex Woo
awoo@arc.nasa.gov

Numerical Aerospace Simulations
Research Branch
Ames Research Center

http://www.nas.nasa.gov/IPG
02/24/2000

Presentation to the New DISS committee
Planning for new NASA data systems
Feb 2000
Overview

- NewDISS needs to be building an open systems & engaging the larger community building "data grids"
- NewDISS needs to move beyond a data archive to a on-demand data, instrumentation & analysis service
- Information Power Grid is a substantial testbed with application experiences with relevance to NewDISS.

This "Grid" is in the other group at AMES that is not associated with the simulator at AMES.
Motivation

- The “Grid” will enable distributed resource sharing and applications based on high-speed coupling of people, computers, databases, instruments, and networks.
- Many facilities are moving toward making resources available on the Grid.
- The Information Power Grid is NASA’s push for a persistent, secure, robust implementation of the “Grid”.

Information Technology

And more slides
Plug-and-play
supercomputers

TeraGrid project aims to foster sharing of data-intensive scientific research

BY DAN CATERINICCHIA
AND COLLEEN O’HARA

When you plug an appliance into an outlet, you don’t need to know that it draws power from a large electrical grid shared by many other appliances. If such an inconspicuous system can be a reliable source of electrical power for toasters, radios and hair dryers, why not apply the same concept to computing power?

That’s what researchers at a consortium of institutions plan to do as they team with several private-sector technology heavyweights on a $53 million project funded by the National Science Foundation.

Four research centers will work primarily with IBM Corp., Qwest Communications International Inc. and Intel Corp. to build the Distributed Terascale Facility (DTF). A high-speed TeraGrid network will connect the centers and enable scientists and researchers across the country — and eventually around the world — to share resources, scan remote databases, run applications on geographically dispersed computers and view complex computer simulations in real time from multiple locations.

The network will transform the research world, said Dan Reed, director of the National Center for Supercomputing Applications (NCSA) and the National Computational Science Alliance.

“What we’re really building is a high-speed backbone to connect the four centers,” Reed said. “The TeraGrid provides seamless access without people knowing where the infrastructure is and how it works. The DTF is the beginning of 21st century infrastructure for scientific computing.”

The four research centers that will make up the DTF are the NCSA at the University of Illinois at Urbana-Champaign; the San Diego Supercomputer Center (SDSC) at that city’s University of California campus; the Argonne National Laboratory in Argonne, Ill.; and the California Institute of Technology in Pasadena (see box).

The TeraGrid could have a dramatic impact on scientific research. SDSC director Fran Berman foresees scientists investigating the fundamental processes of the human brain, sharing simulations of new cancer drugs and ushering in an era of human genome research.

The grid will create vast pools of computing resources by connecting widely distributed supercomputers using the Internet or high-speed research networks, as well as open-source protocols from Globus, a consortium of government and educational institutions working on grid-computing research.

“Globus is kind of the glue that ties together all the computers,” said Michael Nelson, director of Internet technology and strategy at IBM.

IBM Global Services will deploy clusters of IBM eServer Linux systems at the four DTF sites beginning in the third quarter of 2002. The servers will contain Intel’s Itanium next-generation microprocessor, said Robert Fogel, strategic marketing manager for Intel’s high-performance computing division.

The system will have a storage capacity of more than 600 terabytes of data. The Linux clusters will be connected via a 490 gigabit/sec Qwest network, creating a single computing system able to process 13.6 trillion calculations/sec (13.6 teraflops).

The initial TeraGrid is scheduled for completion in three years, but pieces of the network should be operational by the end of this year, Reed said.

As the grid grows to include regional and then international research facilities, it will eventually be used for commercial applications, in much the same way the Internet is, Reed said.
NSF Launches TeraGrid For Academic Research

Promising benefits to researchers working on everything from drug discovery to climate forecasting, the National Science Foundation (NSF) last week launched what will be the nation's most powerful network for scientific computing. NSF has pledged $53 million to four U.S. research institutions and their commercial partners to build and operate a system expected to be up and running by 2003. Its official name is the Distributed Terascale Facility, taken from its targeted capacity to perform trillions of floating-point operations per second (teraflops) and store hundreds of terabytes of data. But if it's a success, it may go down in history as Internet 3.

The institutions—the University of California, San Diego; the University of Illinois, Urbana-Champaign; the California Institute of Technology in Pasadena; and Argonne National Laboratory outside Chicago—are no strangers to supercomputing. San Diego and Illinois, for example, are home base for NSF's Partnership for Advanced Computational Infrastructure program. Last year NSF gave $45 million to the Pittsburgh Supercomputing Center for a 6-teraflops machine. But the TeraGrid, as it's been dubbed, is touted as a new breed of supercomputer, with software that will allow high-speed, high-bandwidth connections previously not possible.

"It's not just size or speed," says Fran Berman, head of the San Diego Supercomputer Center. "This will change how people use data and how they compute." Her counterpart at Illinois's National Center for Supercomputing Applications, Dan Reed, says the TeraGrid will "eliminate the tyranny of time and distance."

It's already changed the sociology of supercomputing, with its cutthroat competition to have the biggest and fastest machine on the block. The winning institutions were the only entrants in what was scheduled to be a competition. "We were under a lot of political pressure to get this out by September," says an NSF official, "and we only gave [applicants] 3 months to put together their bid. We knew that would be a tough deadline for people to meet." Despite being the only applicant, the winners put together a proposal "that passed [peer review] with flying colors," says Bob Borchers, NSF's head of advanced computing.

The TeraGrid will build on an existing 40-billion-bits-per-second fiber-optic network, the so-called Internet-2 created by Qwest, one of three key industrial partners in the facility. It will rely on clustered Linux servers from IBM powered by thousands of Itanium-family processors from Intel. Each of the four institutions will contribute elements to the TeraGrid; by April 2003, it is expected to deliver 13.6 teraflops of computing power and more than 450 terabytes of storage.

NSF officials are hoping that this fall Congress will give the agency enough money to connect the Pittsburgh center to the grid in a few years' time. That will be followed, says Borchers, by a "deepening" of the network to connect a steadily rising number of regional and local sites. That's the path NSF followed to help create its previous research backbone that became the Internet.

JEFFREY MERVIS

17 Aug 2001

www.scientemag.org SCIENCE VOL 293 17 AUGUST 2001

Note: I am not sure if this is a "digital grid" project or not.
- Roy Penne
IBM Tries to Sell Grid Computing

Aims to cluster servers, but analysts are skeptical about corporate uses

BY TODD R. WEISS

After making a big investment in Linux technology, IBM is now betting corporate users will buy into another emerging but still unproven open-source concept: grid computing.

The idea, which is also being eyed by Sun Microsystems Inc. and Microsoft Corp., provides a way to link servers in multiple locations and combine their computing power. Grid computing originated in universities and research institutions, but IBM last week announced an initiative aimed at both scientific and business uses.

IBM detailed plans to build a worldwide grid using systems at its various data centers. Users would pay for processing time on an "e-utility" basis. The company is also grid-enabling key products and offering to help develop and manage grids for users.

IBM said it has been chosen by the British government to work on just such a system as part of a national grid being built at various universities for collaborative scientific research.

But much remains up in the air. IBM officials didn't disclose a schedule for completing the company's own grid or how usage of it will be priced.

Grid computing itself is also still a work in progress. "This is a vision, and visions don't necessarily come true," said Tony Hey, director of the U.K.'s Office of Science and Technology. "But it has the potential to be as significant as the Web."

"We see this as being very applicable to e-business," said John Patrick, vice president of Internet technology at IBM. "It has some further maturing to do, but we're quite confident... that we can get there."

Analysts were more skeptical about whether corporate users will accept grids, which are built with clusters of servers, standard security measures and open-source protocols and software. Users could get performance improvements by dynamically shifting processing workloads across grids of servers, said Dan Kusnetzky, an analyst at IDC in Framingham, Mass. But the challenge for vendors, he added, "is that most commercial applications don't work [that] way."

Jonathan Eunice, an analyst at Illuminata Inc. in Nashua, N.H., said the clearest use for grid computing remains in the research field, at least for now.

For business users, Patrick said, benefits could include the creation of a virtual computing environment providing e-commerce Web sites with almost limitless capacity backups. Companies could use grid network resources on demand and pay usage fees, he added.

* IBM is working with the British Government on this
* This story talks about distributed computing and not about data issues
* Another emerging but still unproven concept
Inside

July 17, 2000

A digital earth

www.fcw.com for daily news
The vision looks something like this: Go into virtual orbit around Earth. Zoom in on any part of the planet and grab a detailed view of a road system, vegetation, weather, even an image of a single house or garden. Flick a virtual wrist, and you can even know the political layout of the community, all in 3-D.

That capability is what the Digital Earth program, a multi-agency effort headed up by NASA, wants to deliver. Digital Earth, first announced by Vice President Al Gore two years ago, sounds futuristic, but some elements of this vision might soon appear as real products. Later this month, the U.S. Geological Survey, one of the sponsors of the project, will hold a crucial workshop to determine what an alpha version of Digital Earth might look like.

Digital Earth is expected to eventually encompass all the geological, geographical and demographic information collected about Earth and its inhabitants. That would enable scientists, teachers, policy-makers or anyone else to develop detailed analyses of features and phenomena as small as one meter in size.

But Digital Earth is more than just a mapping application. The program has the potential to become a framework for an enormous range of information that can be pinpointed to a single geographic point on Earth, including cultural and scientific data. By layering this information over the geographic framework, proponents say, Digital Earth could lead to a much broader understanding of events that affect people around the world, from natural phenomena to cultural and political issues.

More immediately, it’s seen as the basis of projects such as the Global Disaster Information Network. This initiative, launched by the White House in April, would create an electronic system that would provide geological, demographic, political and other information at any place a disaster occurs. That way, federal, state and local officials will have quicker access to more information to make better decisions about what services to provide.

On the commercial side, Digital Earth could revolutionize the markets of travel and tourism, real estate, media, and sports and entertainment by providing more geographic information.

“Digital Earth is far more than the next ‘killer app’ on the Net — it will become the next Internet,” said David Lorenzini, director of geospatial data for Skyline Software Systems Inc., a company in Waltham, Mass., that provides Internet-based 3-D georeferenced information.

The Internet, according to Lorenzini, is defined by World Wide Web browsers, which display 2-D information and are largely text-based. Digital Earth could form the basis of the next Internet because it is a 3-D, visually based “Web” format that will be browsable. This Internet form will be more useful because studies have shown that information is more meaningful to people when presented in a visual rather than text format.

Making Desktops Obsolete

Gore, announcing the project in January 1998 at the California Science Center in Los Angeles, said a pressing problem facing the nation was how to turn the huge amounts of geospatial data that had been collected via satellites and other means — a volume that would only continue to expand — into information that people could use.

The answer Gore proposed was a mechanism for tying together data from multiple, networked sources and making that integrated resource available visually as a browsable, 3-D version of the planet. It would make use of broadband Internet technology now under development and other emerging technologies.

This model of information management could make obsolete the current “desktop” metaphor employed by existing PC operating systems. PCs depend on a “files and folders” representation to find and manipulate information — an inherently hierarchical approach. The Digital Earth concept is non-hierarchical: It contains all information in plain view, and finer-grained elements of that information will be made available by “zooming in” on finer and finer resolutions of areas of the planet.

“If we are successful,” Gore said, “[Digital Earth] will have broad societal and commercial benefits in areas such as education, decision-making for a sustainable future, land-use planning, [and] agricultural and crisis management. [It] could allow us to respond to man-made or natural disasters, or to collaborate on the long-term environmental challenges we face.”

More than 30 government agencies that either produce or use geospatial data are involved in the initiative, including USGS, the National Oceanic and Atmospheric Administration, the National Science Foundation, the Environmental Protection Agency and the Army Corps of Engineers. NASA is the
lead agency for the project.
In the long run, Digital Earth will become important to government because sharing geospatial data among agencies — currently a rare occurrence — will become the norm, said Ivan DeLoach, chief of the EPA's data acquisition branch. This is particularly true because funding and other resources are expected to become even more constrained, he said.

"The sharing of data is done only haphazardly, currently," DeLoach said. "We have some agreements with various agencies, but we want to develop this [sharing] on an enterprise and organizational level so we can leverage the resources we have. Standards, statutes and stovepipes are all barriers to doing that now."

The critical element to sharing data is interoperability, said Alan Gaines, senior science associate for spatial data and information at the NSF. A wide range of data, software and systems has to work together. Government and universities are working on many projects to solve some of the interoperability problems, he said, "but each addresses only a small part of the whole, and none can integrate with any of the others."

The Federal Geographic Data Committee, for example, is examining ways to make it possible for systems to share different kinds of data, even though that data is stored and presented in different and sometimes incompatible formats. The Open GIS Consortium Inc., a nonprofit organization of public and private groups in Wayland, Mass., is working on the software side to integrate geospatial systems around the world.

The Web Mapping Testbed, of which the Open GIS Consortium is overseeing development, will provide one of the underpinnings of Digital Earth. In a demonstration last September, the testbed showed that it was possible to automatically combine, via the Web, different map overlays of the same geographic region. Previously, such a project required skilled technicians working with sophisticated software on stand-alone systems.

"We are pleased that the Digital Earth initiative has embraced the idea of standard as wholeheartedly as it has," said Lance McKee, vice president of corporate communications for the Open GIS Consortium. "It didn't seem to be that way at first, as it looked to be going with multiple, separate projects. But we perceive both the Open GIS and Digital Earth approaches to be based on interoperability, which means they need to be built on a foundation of standards."

Selling Government
Still, despite the vast potential of the project, agencies are expressing varying levels of interest in Digital Earth. Although some government agencies, such as the EPA, are enthusiastic, others stand to benefit are less so.

Charles Challstrom, director of the National Geodetic Survey at NOAA, acknowledged that people at his organization are skeptical. Before trying something as ambitious as Digital Earth, he said he believes the government needs to focus on developing standards and procedures for exchanging information among geographic, mapping and imagery programs spread throughout federal agencies. Others outside of government are taking a longer-term view. Michael Goodchild, chairman of the geography department at the University of California, Santa Barbara, said Digital Earth is a vision of what might be 20 years from now. Nevertheless, he said it's capable of motivating a large number of people to work together toward a common goal that might show results sooner.

"If Digital Earth is to be a repository of everything that's known about the Earth, then it must be capable of storing and handling models, or digital representations, of Earth processes," he said. "We are working hard on the problem of storing and accessing models, and that is likely to produce some exciting results in the next year or so."
This is of more importance to the research community, he acknowledged. The first thing the public may see as a result of Digital Earth, he said, is a better system for access to geospatial data, through the evolution of projects such as the National Geospatial Data Clearinghouse, the Global Spatial Data Infrastructure, Environmental Systems Research Institute Inc.’s Geography Network and other private-sector initiatives.

Indeed, Digital Earth could become the primary way of accessing information about the planet, said Skyline Software’s Lorenzini.

“As various Digital Earth initiatives grow and open access to rich, local content as well as to global information sources, Digital Earth will become more useful than any GIS or data network,” he said. “Digital Earth will evolve into a real-time, collective, visual, spatial decision-support system. It will become the best tool for addressing ill-structured geo-related queries and for promoting informed, collaborative decision-making among nontechnical audiences.”

Other companies are also putting a lot of hope into the eventual success of Digital Earth. Walnut Creek, Calif.-based GlobeXplorer Inc. describes itself as an aggregator of other companies’ geospatial content, which it sells on demand and in real-time to users via the Web, including to other dot-coms. The company has tried to map as much of its content as it can to standards and frameworks expected to be used by Digital Earth, said GlobeXplorer chief operating officer Michael Fisher.

Commercial-side enthusiasm could be the key to Digital Earth’s future. The initiative is moving “at a reasonable pace,” said Thomas Taylor, program manager for Digital Earth at NASA. “The real measure of its progress...will be the extent of the interest shown by the private sector. Companies are already looking at developing products that fall within the concept of Digital Earth. If the initiative is to succeed, it’s the commercial sector that will have to drive it.”

**Patience Needed**

That realization should be raising red flags for the government, according to Jack Pellicci, vice president of global service industry development for Oracle Corp., which recently hosted a Digital Earth “community meeting” at its Reston, Va., facility.

There is a lot of “genuine goodness” associated with Digital Earth, he said, but the reality is that, if government wants industry to participate, “they need to provide the incentives and make business opportunities available.”

“There are a lot of resources out there, but they need to be focused,” he added. “Maybe pulling it together to focus on such things as e-government will provide that focus. Right now, when I talk about Digital Earth to our sales force, their eyes glaze over. E-government is not something you need radically new ideas or technology for. What you need is implementation.”

The fear is that developing Digital Earth will drag on for too long. “It won’t mean very much unless they speed things up and get some practical applications out there in the near future,” Pellicci said.

Other things, outside the control of government, could put a drag on the evolution of Digital Earth. Even though open standards approaches are starting to gel, that might not mean much if long-promised, affordable broadband connections take much longer to reach the general public, according to some observers.

Nevertheless, people are weighing those near-term concerns with the long-term potential of Digital Earth, and coming down on the side of patience. “We are also concerned about the process going from the talking phase to one of real implementation,” said Dan Dubno, an Emmy award-winning producer for CBS, who helps in preparing disaster-related information for CBS broadcasts. “But brilliant ideas take time, and Digital Earth is brilliant. It’s also essential and inevitable. It’s just a question of how soon the government and the politicians can get their house in order about this.”

Lorenzini also said it will be slow-going. But he expects that the program will pick up speed as organizations see how their data “can be visualized and understood by the masses of Net-connected, nontechnical audiences. It’s visual data mining for the video-game-playing generation.”

Robinson is a freelance journalist based in Portland, Ore. Paula Shaki Trimble contributed to this article.
Digital Earth lives on
Private sector helps movement keep apace

BY BRYANT JORDAN

S

SO FAR THERE'S BEEN
NO SLACKENING
OF THE EFFORT THAT I
CAN TELL.
Paul Bryant, FEMA

upporters of the Digital Earth movement say the plan to create a digital real-time representation of the planet is moving forward despite losing the influence of its leading proponent, former Vice President Al Gore.

What has been largely a government-led program is now finding more advocates in the private sector. "It's slowly moving that way," said Jack Pellicci, a vice president of Oracle Corp. and the company's point man on Digital Earth. Industry sees the value of a wide world of spatially referenced data, Pellicci said.

"Location and spatially aware enterprises are the thing of the future," he said.

Pellicci intends to drive that message home in June when he speaks before an international conference on Digital Earth in Canada. "My message is going to be strong on commercialization, partnerships between industry, government and academia," he said.

Digital Earth is an ambitious concept to incorporate maps and data — everything from topographical and population maps to data about migration and weather patterns — into a seamless geospatial system accessible worldwide to citizens and scientists alike. Today, 17 federal agencies participate in the project.

Since the Digital Earth project was launched in 1998, bringing together a number of similar programs across the country and internationally, it has been driven in the United States mostly by federal agencies. NASA — the lead agency — the Federal Emergency Management Agency and others have each maintained small budgets for the program.

At NASA, Digital Earth program manager Thomas Taylor said the private sector's embrace of the project, and even its assumption of the lead role, is how the program is supposed to work.

"It's its own self-sustaining market that is evolving," Taylor said. "We are along for the ride, and it's our choice if we want to participate in it as part of a new market." NASA has budgeted about $1 million a year for Digital Earth activities, Taylor said. He's hopeful that the funding level will be maintained in the fiscal 2002 budget, but specifics have yet to be released.

Paul Bryant, a physicist with FEMA and a representative to the Digital Earth project, said the program "is still a going concern as far as we know."

Congress, he said, has instructed the agency to develop hazard maps that will become part of the geospatial data making up Digital Earth.

"Gore did have a big interest" in Digital Earth, he said. "But so far there's been no slackening of the effort that I can tell."

Mark Rechardt, director of marketing and public-sector programs for the Open GIS Consortium, said he expects the Bush administration will like what it sees in Digital Earth.

"When I look at the Bush-Cheney campaign material, there are so many issues of concern that beg for spatial information to help with the decision-making process," he said.

"I think it's a foregone conclusion that regardless of brand name, the excitement of spatial information is there — people understand it," he said. "Digital Earth is helping to move spatial information from the experts into the hands of everyday decision makers — business people and citizens."

Pellicci said there was concern regarding what a Gore loss in November would mean for the program. But in retrospect there was little cause for concern, he said.

"Over the last year and a half, Gore paid no attention to it anyway — he was busy running for office," Pellicci said.

But it's embedded in NASA, and the rest of the world is involved now."

DOJ unveils antitrust e-mail

BY BRYANT JORDAN

T

he increasing use of the Internet as a means of communicating with government has prompted the Justice Department to issue an online call for leads involving antitrust violations.

"We've been considering this for quite some time. It's simply a way to make ourselves available more readily to people who use the Internet as a means of communication," said Jim Griffin, deputy assistant attorney general for criminal enforcement in the Antitrust Division.

"We still have telephones, and we still receive letters," he said, "but more and more people communicate with each other and with government agencies through the Internet."

The call for information has been added to the Antitrust Division's New Case Unit. The posting is also an enticement to those who might have committed violations. "Individuals or companies with concerns that they may have been involved in criminal antitrust violations may cooperate with the Antitrust Division and avoid prosecution if they meet the conditions of our individual or corporate leniency [amnesty] policies," according to the Web site announcement.

Anyone wanting to contact the department regarding violations can write to newcase.atr@usdoj.gov.
Grid Leaders

From: HPCwire [mailto:hpcwire@tgc.com]
Sent: March 10, 2000 11:17 AM
To: kellie@ucar.edu
Subject: 17239 Grid Leaders Convene at Grid Forum Workshop 03.10.00

Grid Leaders Convene at Grid Forum Workshop
NEWS BRIEFS
HPCwire

San Diego, CA -- The San Diego Supercomputer Center (SDSC) and the National Partnership for Advanced Computational Infrastructure (NPACI) will host the 3rd Grid Forum meeting (GF-3) (http://www.sdsc.edu/GridForum/) March 22-24, 2000 at the Price Center Ballroom on the UCSD (http://www.ucsd.edu) campus.

Grid Forum (http://www.gridforum.org) is a community-initiated forum of individual researchers and practitioners working on distributed computing, or "grid" technologies. Grid Forum focuses on the promotion and development of Grid technologies and applications via the development and documentation of "best practices," implementation guidelines, and standards with an emphasis on rough consensus and running code.

These "Grid" technologies are critical to such activities as the PACI National Technology Grid, being developed through NPACI's Metasystems efforts (http://www.npaci.edu/Thrusts/Metasystems/) in collaboration with the National Computational Science Alliance; NASA's Information Power Grid; the Department of Energy's Accelerated Strategic Computing Initiative; and other activities worldwide. Grid Forum is organized in nine working groups focusing on such areas as security, scheduling, and Grid user services.

"This workshop is an opportunity for the scientific community to come together and find common ground in computational practices, as well as to discuss the continued development of computational infrastructure," said Anke Kamrath, GF-3 organizer and SDSC's associate director for distributed computing. "Grid Forum will help steer NPACI's efforts, as well as those of similar programs, toward the long-term goals of the Grid, to impact, influence, and help shape the Grid into the sophisticated, seamless system we envision. As importantly, having people from various grid projects come together will promote interoperability at the technical level and collaboration at the human level."

The Grid Forum evolved out of a Birds of a Feather (BOF) session at the Supercomputing '98 (http://www.supercomp.org/) conference, in an effort to foster a dialogue on the development of technology grids. GF-1 was held at NASA Ames in June 1999 and at this meeting the initial working groups were formed. At GF-2, held at Northwestern University in Chicago, initial working group charters were developed and an operating structure for Grid Forum was developed. At GF-3, working groups will spend most of their time making progress toward the objectives outlined in their charters.

"The best way to make progress in this community is to have working prototypes, to determine what will work the best," said SDSC's Reagan Moore, who is a co-chair of the Remote Data Access working group. "The Grid Forum is still coming together, but we're already making headway."

The National Partnership for Advanced Computational Infrastructure (NPACI) unites 46 universities and research institutions to build the computational environment for tomorrow's scientific
discovery. Led by UC San Diego and the San Diego Supercomputer Center (SDSC), NPACI is funded by the National Science Foundation’s Partnerships for Advanced Computational Infrastructure (PACI) program and receives additional support from the State and University of California, other government agencies, and partner institutions. The NSF PACI program also supports the National Computational Science Alliance. For additional information about NPACI, visit http://www.npaci.edu/

HPCwire SPONSORS

Product specifications and company information in this section are available to both subscribers and non-subscribers.

[ ] 944) Raytheon
[ ] 936) Sony
[ ] 948) Tera Computer Company
[ ] 909) Fujitsu
[ ] 937) Compaq
[ ] 946) QSW
[ ] 947) Etnus Inc.
[ ] 921) SGI
[ ] 934) HP, High Performance Systems Division
[ ] 905) Sun Microsystems(formerly MAXSTRAT)

Send info requests (an X-marked copy of this message) to more@tgc.com

Copyright 1993-2000 HPCwire. Redistribution of this article is forbidden by law without the expressed written consent of the publisher. For a free trial subscription to HPCwire, send e-mail to trial@hpcwire.tgc.com.
Information about Data Systems, and Failures

The Government Looks At Why Huge Technology Efforts Often Fail

State builds budget infobahn

White House ultimatum to fed IS: Shape up systems or lose funding

Airline turbocharges schedule efficiency

$1B award to fix air traffic systems

To Integrate Systems Or Not: The Latest Corporate Riddle

Tech goofs

IRS CIO: Agency will consider outsourcing

Lessons From The FAA’s Costly Bid To Upgrade Air Traffic Control
COMPUTING

BIG BLUE'S TECH ON TAP

IBM wants business to farm out its computing—to IBM

While Microsoft and America Online battle for control of the consumer Internet, IBM has quietly been constructing its own grand plan to rule the lucrative world of corporate computing. The idea is to put Big Blue at the forefront of a movement in which companies farm out their computing needs to utility-like providers. Instead of having to constantly buy, maintain, and upgrade the latest technology, IBM envisions a simpler world in which companies would buy computing power and programs on an as-needed basis, just as they do electricity from power companies. IBM is investing billions to turn the vision into reality.

The latest development: On Aug. 2, IBM announced its Grid Computing Initiative, a project to provide any desktop with supercomputing potential via the Net. IBM Vice-President Irving Wladawsky Berger, the plan's lead architect, says grid computing promises to make personal computers more powerful by harvesting idle power from other computers. It will also let workers collaborate on sophisticated applications via high-speed networks. "Just as you saw Web technologies move from the scientific world to the commercial world," says Berger, "we see grid computing doing the same."

HUGE POTENTIAL. So how would it all work? A pharmaceutical company, say, could tap into a grid to let a group of researchers run a complex simulation that would help design a new drug. Right now, a drugmaker would have to fork over millions of dollars for a supercomputer to perform that task. Such prospects have piqued interest in the corporate world. Steve Neiman, J.P. Morgan Chase & Co.'s head of high-performance computing, believes on-tap technology would not only help the bank cut computing costs but also let it more quickly ramp up or cut back computing power. "It will be a great thing for us," says Neiman.

The potential market for on-demand computing could one day be huge. International Data Corp. (IDC), the technology-research firm, estimates the market could reach more than $30 billion later in the decade. Big Blue, of course, isn't the only company seeing that fat opportunity. Over the past two months, Compaq Computer, Hewlett-Packard, and Sun Microsystems have all announced plans to work on early stages of on-tap computing. Tech consultants such as Electronic Data Systems Corp. (EDS), and even telecommunications companies, also want a piece of the action. But many industry experts insist that IBM's financial muscle, world-class research labs, trusted brand, and breadth of products and people put it in the best position to profit from the shift. "They just have all the right ingredients," says Karen Benson, an analyst with Gartner Group Inc.

BERGER: The burden of making the grid initiative work is on his shoulders

Grid computing is only one part of IBM's overall strategy to dominate the future of corporate computing. At the same time the company is building the software and hardware needed to move computing fully on-tap, it is readying its army of consultants to build and sell grid-based services. Those services could help workers perform mind-boggling computing jobs in industries from manufacturing to banking. Already, the company's Web hosting unit, the world's largest, is investing some $4 billion to build 50 new corporate data centers—in addition to its existing 175 data centers—which when linked together will power the grid. "All of these capabilities are absolutely needed to make e-business successful in the future," says Berger, who heads the grid project as well as other key development efforts.

To be sure, tech-on-tap is years away. It is also rife with challenges. The most pressing problem is to develop the software needed to reliably manage the complex grid. But the ingredients are starting to gel: IBM's $1 billion investment in the Linux operating system looks like a smart bet, since much of the software being developed to run grids is based on open-source models, in which computer code is shared. IBM rivals such as Microsoft Corp. and Sun, meanwhile, have spurned open-source computing in order to promote their own proprietary systems. Early this year, IBM's consulting arm, the industry's largest, began selling Web-based services that manage purchasing, storage, and other tasks for corporations. It's still early, but IBM's grand plan is off to a good start.

By Spencer E. Ante in New York

IBM'S GRAND PLAN

Big Blue wants to sell corporations on Web-based computer networks. Here are the highlights:

GRID COMPUTING will allow users to tap supercomputer power via the Net

SOFTWARE such as open-source Linux is getting a big push from IBM

CONSULTANTS offer tech skills to build the grid

HARDWARE such as IBM servers will run software for the grid