

# Head in the clouds

'Cloud computing' is being pitched as a new nirvana for scientists drowning in data. But can it deliver? **Eric Hand** investigates.

**D**ennis Gannon, a computer scientist at Indiana University in Bloomington, knows all about bringing huge amounts of computer power to bear on complex scientific problems. He has at his disposal, for the purpose, Big Red, one of the world's largest supercomputers, right there on the campus.

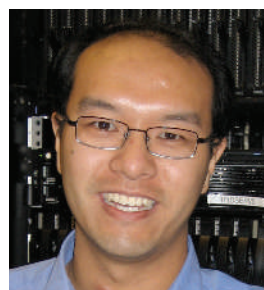
But when Jong Youl Choi, a graduate student computer scientist at the university, approached him with a bioinformatics program that he had written, Gannon suggested they run it on Amazon's EC2 Beta program, as translating it for Big Red would be too time-consuming. Last year, the Seattle-based e-commerce firm introduced a 'cloud-computing' option that provides access to an ever-expandable 'cloud' of powerful computer servers. Gannon and Choi set up three virtual computers and uploaded their program, which seeks matches for an unknown protein sequence from a massive national database. The job took about 15 minutes and cost them less than US\$2.00.

"It was easy," Gannon says. "I don't think Choi even thought about going to Big Red."

Earlier this month, Google and IBM announced their own approach to cloud computing. They say that they will offer free use of a cluster of several hundred servers to the computer-science departments of six top US research universities.

The machines will provide researchers with vast amounts of computer power for their data-crunching. Advocates of the concept say that it will fill a middle ground between the computers that most researchers currently have access to, and machines such as Big Red.

"Right now, we have no choice," says Randal Bryant, dean of computer science at Carnegie Mellon University in Pittsburgh, Pennsylvania, "except from stepping up from desktops to using these giant supercomputers."



**"Students don't have experience dealing with this order of magnitude of data."**  
— Dennis Quan



**Aim for the sky: cloud computing offers options for researchers to work with vast amounts of data.**

IBM and Google say that their pilot scheme will teach computer-science students how to write software that uses the interlinked computers to scour and analyse the large data sets that back up Internet services such as e-mail, maps and social networking. In one exploratory project, for example, students developed software to scan the contents of Wikipedia for entries that might be malicious. "Students don't have experience dealing with this order of magnitude of data, or the ability to make their computer tasks work in a parallel fashion," says Dennis Quan, leader of the IBM group responsible for setting up the clusters.

The companies hope that the concept will evolve into a valuable resource for researchers who need powerful computers to crunch their data, but don't want the hassle of adapting their applications for supercomputers. They eventually plan to expand the cluster to 1,600 processors.

For now, Quan says, the firms have no plans to charge researchers for access to the cloud computers. But their investment, which has been estimated at tens of millions of dollars, may be intended to test the waters and see how badly university scientists crave more computing power. "They're probably donating this to see if there's a market there," says Richard Loft, a physicist who works with supercomputers at the National Center for Atmospheric Research in Boulder, Colorado.

On 16 October, Amazon responded to IBM and Google by expanding its program to more developers, and enhancing the service

in terms of speed and storage. Their pay-as-you-go model, unveiled in August 2006, asked researchers to pay small fees of about 15 cents to store a gigabyte of data for a month, or 10 cents and upwards for an hour of processing time.

All this activity from the computer business's biggest names reflects the recent explosion in the sheer volume of data that scientists have to deal with. Rapid growth in the precision and speed of sensors and instruments, as well as in the computers themselves, has fired this expansion.

## Virtual reality

Dealing with the data often falls to supercomputers such as Blue Ice, in Boulder, or Big Red. But critics say that these are more attuned to running huge models of complex systems — such as Earth's climate, or gas flow in a jet engine — than they are to the sort of intensive database mining that researchers in many disciplines now want to do. And although the centres that house them do all they can to fight the perception, supercomputers aren't seen as the most user-friendly devices around. The centres still "favour the efficiency of the machine over the efficiency of the humans that are using the machine", claims Bryant. He also says that their limitations can make using them "like the old punch-card days".

Other universities, such as the University of Texas in Austin and the University of Tokyo, have sought to address their scientists' needs by building their own supercomputing clusters. But Google, IBM and Amazon reckon that their experience in high-power applications will make it cheaper and more efficient for them to provide it as a service. If the experience of Gannon and his student is anything to go by, the cloud computer could be the workstation of the future. ■