Amazon puts network power online

GETTY IMAGES

Plug in: industry supercomputer power on the desktop PC could have a big impact on scientific research.

When Dutch computer scientist Rudi Cilibrasi needed hundreds of hours-worth of computing time to test a data-mining algorithm earlier this month, he went not to his IT department but to Amazon.com. He paid \$60 with his credit card, and in minutes had the equivalent of ten servers installed, which crunched through his job in a couple of days — ten times faster than his desktop PC would have managed.

Large web companies such as Google, eBay and Amazon have far more computing power at their disposal than any academic networks, and have become leaders in massive-scale distributed computing. Many of their innovations can help scientists, and Amazon's computing-on-demand service, which has been running since August, is no exception. It enables customers to create multiple virtual computers on Amazon's massive computing infrastructure for \$0.10 per computing hour, and to store data for \$0.15 per gigabyte per month.

The service is still in a test phase, so few scientists have even heard of it yet, let alone tried it. But it is a movement that experts believe could revolutionize how researchers use computers. In future, they will export computing jobs to industry networks rather than trying to run them in-house, says Alberto Pace, head of Internet services at CERN, the European particle-physics laboratory near Geneva. CERN has built the world's largest scientific computing grid, bringing together 10,000 computers

in 31 countries to handle the 1.5 gigabytes of data that its new accelerator, the Large Hadron Collider, will churn out every second once it is switched on next year.

"I see no reason why the Amazon service wouldn't take off," Pace says. "For a lab that wants to go fast and cheaply, this is a huge advantage over buying material and hiring IT staff. You spend a few dollars, you have a computer farm and you get results."

Cilibrasi, a researcher at the National Institute for Mathematics and Computer Science in Amsterdam, was using Amazon's service to test an algorithm aimed at predicting how much someone will like a movie based on their current preferences. He says he is a convert: "It's substantially more reliable, cheaper and easier to use [than academic computing networks]. It opens up powerful computing-on-demand to the masses."

It's not just computer scientists who could use cheap computing power. Climate researchers can run global-warming models, cosmologists might simulate the inside of a supernova or biologists could scale up the workings of a liver cell to the whole organ.

The South African National Bioinformatics Institute at the University of Western Cape, Belleville, has already been testing Amazon's system to power large-scale genome comparisons. The pay-as-you-go system offers com-

puting power and bandwidth that the institute could not afford to maintain itself, says Inus Scheepers, a systems administrator there.

Cost is certainly one reason observers are excited about Amazon's system. Other companies, including Sun Microsystems, offer computing-on-demand, but Amazon's service costs a tenth as much. But the main attraction is Amazon's use of 'virtualization' technologies, which many predict will change not just research but computing itself.

Virtualization uses a layer of software to allow multiple operating systems to run together. This means that different computers can be recreated on the same machine. So one machine can host say ten 'virtual' computers, each with a different operating system.

That's a big deal. Running multiple virtual computers on a single server uses available resources much more efficiently. But it also means that instead of having to physically install a machine with a particular operating system, a virtual version can be created in seconds. Such virtual computers can be copied just like a file, and will run on any machine irrespective of the hardware it is using. "In the past, we had to install hardware and software for each machine," says Pace.

Virtualization software is becoming widely available from companies such as Microsoft and VMware. Amazon uses Xen, an open-source system developed at the University of Cambridge, UK, which is fast becoming popular with researchers. As well as using Xen to

"You spend a few dollars, you have a computer farm and you get results."

enable virtual computers to run across a grid or cluster, despite different operating systems on the individual machines, researchers can also use applications developed on their lab computer.

At present, to run an application on a large scale, they often need to rewrite it. With virtualization, researchers can create a copy of their own machine and use it to run large-scale simulations or searches, and it should work exactly as it does in the lab.

Amazon's service combines the joys of virtualization with the huge computing power it has at its disposal. It's an approach that looks set to catch on. CERN has started an internal service similar to Amazon's, in which users can create or delete virtual machines on the fly. "Virtualization is revolutionary," says Pace. "It's clear that this is one way to do scientific research in the future."

Declan Butler