

CLOUD RULES #99

# Nature never worries about capacity. Neither should you.

Opscode + Cycle Computing = 10,000 nodes in 45 minutes. Science!

Spinning up an Unprecedented High Performance Computing Cluster Using Cycle Computing's CycleCloud Technology Provides Biomedical Giant with Instant Datacenter for Protein Binding Research Project.

## ► The Challenge

Since 2005, Cycle Computing has delivered proven, secure and flexible high performance computing (HPC) and data solutions, helping clients maximize existing infrastructure and speed computations on desktops and servers on-demand in the cloud. Cycle helps customers of all sizes gain fast and efficient access to the kind of supercomputing muscle usually reserved for universities and large research organizations. Cycle's founder, CEO and HPC expert Jason Stowe wanted to test the envelope of cloud computing and take its technology to the next level - spinning up a 10,000-core cluster in the cloud. Stowe teamed up with a prominent life science company who needed supercomputing power to research how proteins bind to each other, and utilized Amazon Web Services to begin provisioning 10,000 cores in an HPC cluster on Amazon's EC2 compute cloud.

## ► The Solution

Using a CentOS Linux cluster, Cycle built its 10,000 cores using 1,250 instances with eight cores each. Scaling up roughly 2,000 cores at a time, Cycle enlisted its own CycleCloud software, the Condor scheduling system, and Opscode's Chef cloud infrastructure automation engine. Using Chef, the total time to provision the entire cluster was 45 minutes and cost approximately \$1,000 per hour for a 10,000-core cluster.



LOCATION : Westerfield, CT

INDUSTRY : Software and  
High Performance  
Computing

INFRASTRUCTURE : Amazon EC2

## ► The Results

- Set up time significantly reduced from weeks or months to minutes
- Increased agility and reliability to more efficiently manage platform configuration
- Improved configuration to increase the scale and enable creation of hundreds of execute nodes at once, without any errors
- Increased cost savings by eliminating cluster ownership/maintenance costs of building out a 10,000-core cluster infrastructure

Cycle Computing has delivered proven, secure and flexible high performance computing (HPC) and data solutions since 2005. The company's technology helps clients maximize existing infrastructure and speed computations on desktops and servers on-demand in the cloud. With its CycleServer HPC management software and CycleCloud fully-supported and secured HPC clusters, Cycle clients experience faster time-to-market, decreased operating costs, and unprecedented service and support. Starting with three initial Fortune 100 clients, Cycle has grown to deploy proven implementations at Fortune 500s, SMBs as well as government and academic institutions including JPMorgan Chase, Purdue University, Pfizer and Lockheed Martin.

*"We amplified our Chef infrastructure and streamlined configuration scripts to help scalability. We developed an approximate timing architecture which benchmarked and facilitated the increased peak number of nodes and the nodes per unit time. Opscode's Chef enabled us to easily and reliably configure our systems to meet our demanding scale requirements.."*

▶ Jason Stowe  
CEO, Cycle Computing

Cycle Computing founder, CEO and HPC expert Jason Stowe wanted to test the limits of HPC computing on the cloud. No stranger to building clusters on Amazon's Elastic Compute Cloud (EC2), having scaled up several thousand cores using Cycle's own CycleCloud software, Stowe and his engineers wanted to push it to the next level —provisioning 10,000 cores on Amazon in an HPC cluster using batch-scheduling technology and running an HPC-optimized application.

Building that kind of muscle in a 10,000-core physical server cluster can take weeks or months, and run up costs into the millions of dollars. Stowe felt strongly that the cloud offered a

solution, though HPC customers have long been skeptical. With the right management and security tools, Stowe felt Cycle could offer an economical, reliable and secure solution.

As reported in articles in Bio-IT World, NetworkWorld, and The Register, enter biomedical giant Genentech, now part of the Roche pharmaceutical company, which needed supercomputing power to do an extensive protein analysis that would eventually consume more than 80,000 hours of computing time. With extensive consultation between Cycle engineers and the company's computational scientists and IT teams, and after thorough price-performance analysis was done by Cycle, the solution architecture was established and the path forward was set.

Cycle used the open-source Condor grid-management system and two of its own tools, CycleServer, its cluster telemetry analytics and visualization engine for HPC management, and CycleCloud, its cloud service offering that rides atop of the EC2 compute cloud at Amazon to launch large scale clusters on-demand. Deploying those tools, it took 15 minutes to fire up the first 2,000 cores in the virtual cluster, and within 45 minutes all of the 10,000 virtual cores on the 1,250 servers were up and running.

To configure them reliably and quickly, and automate software installation and configuration of such large clusters, Stowe's team used Opscode's Chef cloud infrastructure automation engine to tweak the settings on each virtual server instance on EC2.

"We beefed up our Chef infrastructure and streamlined our configuration scripts to help with scalability," said Cycle's Stowe. "We also benchmarked and discovered that with an increasing converge timing scheme that we invented, we could both increase the peak number of nodes and the nodes per unit time. With post-optimizations and infrastructure changes, we can now support simultaneous launches of up to 256 instances or 2,048 cores every five minutes. This is up more than four times from the previous limit of 64 instances at a time. And we even have ideas on how we will increase that number dramatically."

The researchers installed their code and ran the job for eight hours at a total cost of approximately \$8,500, including the fees for storage and for using the Cycle Computing tools as a service. That pencils out to just over \$1,000 per hour for a 10,000-core cluster. Were a large life science company to ramp up its own datacenter for a project of this size, including physical servers, storage, and switching, plus operation and maintenance costs, Stowe estimated that the final tally would have been somewhere near \$5 million. For researchers who need to run simulations only periodically, and not incur hard costs for clusters or have to worry about keeping them busy, it makes significantly more sense to utilize the research capacity offered by Cycle Computing, spend less money and achieve results in a matter of hours.

*"What we're really trying to do is make it as easy as possible for researchers and our other customers to run their code," says Jason Stowe, CEO, for Cycle Computing. "Instead of having to purchase servers and create a datacenter, they can gather their results in hours while saving a significant chunk of change."*

### ▶ About Opscode

Opscode is the leader in cloud infrastructure automation. We help companies of all sizes develop fully automated server infrastructures that scale easily and predictably; can be quickly rebuilt in any environment, and save developers and systems engineers time and money. Opscode's team comprises web infrastructure experts responsible for building and operating some of the world's largest websites and cloud computing platforms. Opscode is headquartered in Seattle.

More information can be found at:

[opscode.com](http://opscode.com)

1008 Western Avenue Suite 600  
Seattle, WA 98104

PH: 206.682.7928