

The Dollars and Sense of Cloud Computing

By [Andy Barkett](#), Architecture and Technical Management Consultant

On October 8, 2009, cloud computing will be celebrating its second birthday. It was in 2007 that the headline “Google and I.B.M. Join in ‘Cloud Computing’ Research” first appeared¹. Two years later, cloud computing is still a big buzzword – everyone seems to be talking about it – and yet no large IT department I know of has truly altered how they deliver services to leverage the cloud. Startups are in the cloud, but does *cloud computing* make sense for Corporate IT?

Update: What Does ‘Cloud Computing’ Mean Today?

Cloud computing is used to describe everything from “off-site hosting” to “commodity hardware supercomputing.” Now that the term is two years old, can we finally settle on a meaning for it?

Cloud computing is more than just off-site hosting. To say that moving something “out of our server room” or “out of our datacenter” is the same as saying that something moved “into the cloud” – it misses the point. Cloud computing is also more than just virtualized environments, software-as-a-service, scalable architectures, or utility computing (the quantization and sale of RAM, clock cycles, or disk space as tiny dollops of ‘resources,’ instead of whole machines).

Today, we can settle on a definition that captures all these essential factors and differentiates cloud computing from older alternatives, such as grid computing or off-site hosting. Our definition du jour is from Wikipedia: “It is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet.”

Systems like Amazon’s EC2 and S3, Google’s App Engine, and Microsoft Azure have become the gold standards for cloud computing, and, while they are all different, each fits the definition above.

Why is Cloud Computing Here to Stay?

The macroeconomics of cloud computing are very compelling. While we can all now buy commodity hardware for about the same price, whether in a large or small order, we can’t all *run* the server with the same cost structure. In the datacenter, power and cooling can constitute as much as 25-50% of the cost of running a server. The price for a technician to repair the server, an administrator to patch its operating system, and network connections to connect it to the other servers often add up to cost more than the server itself.

Operating a \$2,500 server might cost \$12,000 over the three-year life of that server. But, Google, Microsoft, or Amazon, with their massive and efficient datacenters, standardized operating systems, trained administrators, and heavily virtualized environments, could probably operate the same \$2,500 server for less than \$7,000.

In fact, the savings can be even greater. The virtualization revolution has taught us that we can pool and oversubscribe resources. With cloud computing, as long as users' needs aren't all likely to hit us on the same day, we can buy 10 terabytes of disk space, but allocate 15 terabytes to our users with none the wiser. Because cloud computing is delivered over the internet and is virtual and scalable, it allows us to pool ever larger groups of users (even users from different companies!) and reap ever larger efficiencies from oversubscription.

Unless your organization garners a competitive advantage from the build-out and operation of datacenters, or unless you have nothing to gain by reducing provisioning time and increasing manageability, you're probably headed for the cloud.

How much can it save me?

The electricity to run a standard, racked server in California might cost \$300 per year. In a typical, small to medium datacenter, that \$300 turns into \$600 per year when we factor in the price of cooling. But power and cooling for the same server in a large, efficient datacenter run by Microsoft or Amazon in Washington state, could cost as little as \$200 per year.

The labor and administration savings are equally compelling. If a user requests a new server today, by what day will it be running and configured for their use? In most IT departments, the answer is that the provisioning and setup process takes anywhere from 30-60 days unless the server is already deployed (which means we are already paying for it!). The time and effort required to order and set up the server cost us \$300-\$500, and if the user doesn't use 100% of the server resources, we've wasted even more money. Compare this to the 10-30 minutes required to deploy a new instance of a server from Amazon's EC2, which we procure only on demand, and stop paying for the instant it is no longer needed.

For a rack of servers that might cost \$100,000 to procure, and another \$300,000 to run for the next three years, the direct cost savings by moving into the cloud could be as high as 75%.

Why Isn't Everyone Doing it?

Startups are flocking to the cloud because they don't have \$400,000 to spend on a rack of servers and so simply have no choice. But the equation for Corporate IT is a little different. The massive cost savings are quickly offset by the problems and difficulties of moving into the cloud.

Moving critical business systems into the cloud has many disadvantages:

- Possible reduction in security. However, you should ask yourself if our security today is truly better than that of Google, Amazon, and Microsoft.
- Loss of control. This is an important factor. Are we truly comfortable with someone else deciding when to patch our OS?
- Loss of "auditability." When the auditor asks me where data is stored and who has access, I might not feel comfortable saying "I don't know; please look in the cloud."

-Loss of “touch.” We are still uncomfortable with the idea that we can’t physically hit the reset switch when we need to.

-Increased bandwidth cost. To use the cloud effectively, we have to move a lot of data into and out of it. This may lead to increased bandwidth costs or may simply make certain performance sensitive systems ineligible for the cloud.

These problems are very real, and, for the most part, have not yet been solved. It is hard to imagine justifying a savings of \$200,000 if the risks of moving a critical system into the cloud are so large. After all, we’d *pay* an extra \$200,000 to make sure that our sensitive financial data is secure, under control, and performing well for our users. This is precisely why many business-critical IT systems haven’t moved into the cloud.

The Power of Salesforce.com

The answer to almost every criticism is the same, and it’s Salesforce.com. Salesforce.com is a dynamically scalable, extensible, on-demand software product delivered over the web. It looks suspiciously like cloud computing. Consider this:

- Salesforce.com holds sensitive customer data.
- Salesforce.com holds sensitive pricing information, price quotes, and other financial data.
- Nearly 100% of data in salesforce.com is hosted “off-site” for the user.
- Over 50,000 companies, large and small, use Salesforce.com every day.

And all 50,000+ customers (with over 1.5 million individual users) are served from about 1,000 servers!

Salesforce.com proves that sensitive, business critical IT systems can be run “from the cloud.”

Which Applications *Can* and *Should* Go Into the Cloud?

Any application that has dynamic scaling needs, both up and down, is a great candidate for the cloud. For example, a web-serving infrastructure that needs to scale up massively to handle short-term PR or product release events might make a great candidate to move into the cloud.

Another good candidate is any system that is already remotely administered – it’s a lot easier for users to accept these moving into the cloud because it doesn’t feel like a big change.

Systems that *don’t* have stringent security or audit-trail needs are easier to move into the cloud.

Systems, which don’t require massive, timely data copies or moves, are easier to move into the cloud *without having to upgrade networks first*.

Finally, compute farms – large grids used for massively parallel processing tasks – are logical candidates *if they currently sit idle some of the time*. If you have 500 machines, even 25% underutilization is very expensive.



How Fast Should We Move?

We've gone two years without moving things into the cloud, and it doesn't seem to have cost us our jobs or even to have hurt our reputations. Perhaps, two years from now, there will still be few corporate IT departments utilizing cloud resources.

It may be safe to say then that no one will get fired for staying out of the cloud. But not getting fired isn't really our goal; streamlining IT, saving money, and delivering better service to our users are our goals. Cloud computing can help achieve these goals, and it can help achieve them now. We should challenge our IT departments to solve the problems of cloud computing and reap the savings.

ⁱ <http://www.nytimes.com/2007/10/08/technology/08cloud.html>