

The Public Cloud: Is There Still Room for Google?

Hakan Jakobsson
hakan@hakan-jakobsson.com

Version 2016-04-03

Introduction

When a product gains popularity it is common that a pattern emerges. In the beginning, there are a small number of pioneering companies producing it, but as its popularity grows, more and more vendors enter the market to get a piece of the action. After a while, the number of vendors peaks as consolidation sets in and eventually, the market becomes dominated by a small number of players. For instance, there were around 80 U.S. automakers in in 1920 and more than 80 U.S. television manufacturers in the early 1950s. More recently, the number of disk drive manufacturers peaked in 1984 at almost 80 and the number of PC manufacturers about 3 years later when there were around 100 hundred vendors according to Michael Mauboussin's book "More Than You Know." Today, there are far fewer manufacturers in any of those categories.

It's more than likely that the public cloud will see a similar pattern with early growth in the number of players followed by a peak and then consolidation with a small number of dominant players emerging as the big winners. In fact, we have already seen the emergence of AWS and Azure as major players, and consolidation in the form of HP shutting down its Helion public cloud and Rackspace refocusing on its role as a support provider. However, while AWS and Azure may well turn out to be the eventual winners in the cloud, it's still early in the game and cloud computing is only at a small fraction of its potential market size. The question is if it's early enough that Google can become a dominant player.

Just like for Amazon, any growth in Google's cloud revenue constitutes real revenue growth for the company. That's not the case for Microsoft. When companies like Microsoft and IBM brag about how fast their cloud businesses are growing, part of that growth is cannibalization of other income streams that doesn't increase the total revenue for the company. It would seem less valuable to have a huge growth in cloud revenue if that revenue is just what used to be the company's on-prem revenue that is now booked through cloud computing. That is one issue to take into account when comparing Google's cloud numbers with those of Microsoft.

One can easily make the case that Google has enormous potential in the cloud. It's a rapidly growing area and we are still only in the beginning of the race. Google, with all its resources, including a huge cash flow and tremendous employee talent, should be able to take on the cloud as well as anyone. Besides, it has a huge internal use case and pioneered a lot of cloud-related technology that later found its way into popular phenomena like Hadoop. And then there is the presence of popular apps like Gmail that help give Google some degree of cloud mindshare. So obviously, there is a good argument for why Google might become a winner in cloud computing. However, that argument is not terribly exciting nor is it overly compelling – look at what happened to its effort to become a player in social networks. The counter argument that Google might fail to live up to its potential in the cloud is far more interesting.

Google and the art of staying relevant

Google, like any high-tech company, does its best to try to stay relevant in the rapidly changing world of technology. Its major cash cow, search advertising, faces challenges including

- The increasing prominence of mobile devices where the dynamics of advertising are different from the desktop.
- The drain of lucrative product searches to Amazon. An Amazon Prime member will often assume that a purchase will be on Amazon and will search for a product there directly instead of using a general search engine. <http://searchengineland.com/googles-eric-schmidts-berlin-speech-biggest-search-competitor-amazon-205761>
- Increasing competition from Facebook for advertising dollars. Facebook's social network information about its users has an enormous potential advertising value and Facebook is still in the process of figuring out how to maximize that value.

So Google has been branching out in all kinds of directions and launching all kinds of projects in the hope that some will turn out to be wildly successful. But as expected, a list of some past projects like Wave, Glass, Google+, Buzz, Orkut, Page Creator, Lively, Answers, Print/Audio Ads, Jaiku, and Notebook will probably have more losers than big winners. Of course, there are some projects that have found major traction as well, like Android, Gmail, and YouTube. So with the proliferation of Google projects, questions have always been raised about Google's level of commitment. Is Google really "all in" on a project or is it more of an experiment? There used to be a lot of skepticism about Google's public cloud offerings along those lines, but sometime in 2015, it would seem that Google decided to increase its commitment to its public cloud significantly and improve its credibility as a serious player in the enterprise space. For comparison, Microsoft dramatically intensified its focus on the cloud after the appointment of Satya Nadella as CEO in February 2014.

In November 2015, Google announced the hiring of Diane Greene, the cofounder of VMware, to lead its cloud effort. It showed that Google was serious about going after the enterprise market with its Google Cloud Platform. And in March of 2016,

Google's GCP NEXT 2016 event generated quite a few headlines. In conjunction with the event, there was an announcement of a serious expansion of the geographical scope of its platform with the addition of new regions. At the event, there was buzz about customer wins, announcements about exciting new AI technology, the whole nine yards. The stock of longtime speech recognition technology stalwart Nuance – the original technology behind Apple's Siri – dropped over 6 percent on the news that Google would publish an API for its speech recognition technology as part of a GCP service.

So why now all of a sudden? In 2005-2006, Google probably knew more about large-scale computing than any other company on this planet because of its internal usage. If it had made the right moves then, it could most likely have avoided getting far behind a bookstore in the public cloud race. Perhaps the reason for the intensified cloud effort, apart from a general desire to diversify its revenue streams, is an increasing realization of where cloud computing is heading – instead of merely providing commodity services, the IaaS clouds are evolving into ecosystems. One could well see why Google, used to fat margins for its search advertising, initially wouldn't have been overly enthusiastic about entering what may have seemed to be a race to the bottom in terms of margins for providing commodity services. Now the company seems to have found religion, but is it too late for Google to become a major ecosystem in the public cloud?

The cloud barriers to entry

Any vendor that wants to be in the public cloud faces the issues of competing with the two frontrunners AWS and Azure. AWS has a large lead, both in market share and technology. Azure has the benefit of Microsoft's entrenchment in the enterprise space, complete with a broad software stack, enterprise-grade support, a large skill pool, and a sales force with existing relationships with a large customer base. Google, on the other hand, is weak on enterprise credibility and some of its offerings have odd quirks like its database pricing. (See <http://hakan-jakobsson.com/archive/DatabasePricing.pdf> for an example.) So Google faces a major fight going up against two major players that are both bigger and have broader offerings.

One major issue that will affect Google's ability to catch up is the fact that the original IaaS clouds are becoming increasingly complex ecosystems rather than data centers for commodity computing. Hence, the choice of cloud provider becomes more of a long-term decision. And it will be a challenge for a smaller player to compete with entrenched ecosystems that are both larger and richer.

Another issue for Google's ability to catch up is the notion of *data gravity*. The term was coined in analogy to physical gravity where larger, heavier objects have stronger gravitational pull than smaller ones and will tend to suck in the smaller ones. In computing, sending information over a data network is often slow and expensive, so there is every reason to do as little of it as possible. Hence the idea that a larger object, data, will suck in smaller objects in the form of application code so

that the code will be executed where the data is and obviate the need to send the data over a computer network.

Ecosystems in the cloud

There is a meme that the cloud is a mere commodity approach to IT infrastructure. According to various versions of this meme, different cloud providers offer pretty much the same thing – computing and storage – and while one vendor may offer slightly faster computing and another slightly lower prices, the vendors are pretty much interchangeable. Just use virtual machines or containers for your computation and send them to whatever provider that offers the lowest price. As a result of this meme, people have believed in concepts like hybrid clouds and multi-cloud computing. The mind-boggling idea behind the hybrid cloud is that companies would have private cloud infrastructure in their data centers sufficiently similar to that of one or more public cloud providers that they could effortlessly shift their workloads between different clouds as needed. Cloud bursting would be an example.

The hybrid cloud idea has mainly been a tremendous sales effort by traditional IT vendors to sell infrastructure components to their customers to create private clouds in their data centers. Public cloud providers, like AWS, typically make their own hardware, so the shift to cloud computing is quite bad news for the traditional vendors of on-prem IT infrastructure. Hence their push for the hybrid cloud. In practice, the concept has found little traction but it's still heavily promoted by some very large legacy IT vendors. The enthusiasm for this idea seems to have been dampened by an increasing realization of the implausibility of private clouds being able to mirror the rapidly growing functionality of public clouds. Very few companies will be able to have private clouds that can keep up with the level of innovation that is going on in the public cloud, so the private cloud component of a hybrid cloud will likely be an inferior environment compared to the public cloud component. So why waste a lot of money trying to build something as complex as a cloud in your own datacenter? A much more reasonable concept than that of the hybrid cloud is that of *hybrid IT* – the idea that legacy applications will continue to run in private data centers and coexist with newer deployments in the public cloud. Legacy applications may eventually move to the public cloud but it may well be a rather slow process.

The main problem with the “commodity” meme is that it reflects a very unsophisticated use of the cloud that may have been common in the early days but is now becoming obsolete because of services. Services are what will turn public clouds into ecosystems that are not mere “platforms” but also include third-party apps and tools, skill pools and the availability of professional services. Ecosystems are not commodities that you switch between willy-nilly. Take Windows: On the surface it's an operating system that does approximately the same things that other operating systems do. But to a Microsoft shop, it's likely more of an ecosystem of integrated components ranging from the OS to .NET to Active Directory, Analysis

Services, Excel, SQL Server, etc. Add to that the skill pool of people with Microsoft training, apps developers, and the availability of professional services and you have an ecosystem that is far more than just an operating system. It's the ecosystem that has dominated the desktop and has had a fair share of data center customers for the last few decades. Other areas have seen similar formations of some form of ecosystems. Android and iOS would be examples. Or take Hadoop, which was originally based mainly on HDFS and MapReduce, but has now evolved into an ecosystem with a wide variety of components.

Right now, AWS is the overwhelming cloud leader in services. These range from very basic services for compute and storage to content delivery, networking services, load balancing, database services, analytics, mobile, the Internet of Things, "server-less" Lambda, transcoding, you name it. And these services and their integration often provide a really good value proposition that you can't fully realize unless you go "native" on that cloud rather than just using containers to run as some compute tasks. (See <http://hakan-jakobsson.com/archive/DatabasePricing.pdf> for examples.) Other cloud providers will likely eventually duplicate the functionality of the AWS services they lack, but those versions probably won't work exactly the same way, will have different APIs, will require a separate learning curve, and, thus, will likely not allow for easy interchangeability between clouds. There will be high switching costs.

AWS's lead in services, market share, and mindshare seems to be creating a classical virtuous cycle of increasing third party support, an increasing number of professionals with AWS skills, etc. that makes the platform more and more appealing and helps it attract new customers. The customer growth, in turn, makes it even more attractive to third parties and as a skill set and so on. The following question arises: Could Amazon be on the verge of creating an ecosystem that could become a juggernaut in computing just like what happened with Microsoft's Windows ecosystem during the PC revolution back in the 1980s and 1990s? That possibility must certainly have occurred to Microsoft given its strong Azure push and its apparent willingness to embrace Linux for fear of becoming a mere Windows-shop niche in the cloud. Apparently, Microsoft, under Satya Nadella, no longer thinks Windows will take over the world and recognizes the huge role of Linux in enterprise computing.

So how does Google take on the two larger players? One approach has been to compete on price. That might work for computing as a pure commodity, but as the ecosystems evolve, there will be increasing vendor lock-in that will make price less of an issue. Microsoft and Oracle sell a lot of software, but hardly because their prices are always the lowest. Moreover, AWS has shown in the past that it's not afraid to lower its prices if deemed appropriate.

Google has also made an effort to differentiate itself by introducing advanced AI-related cloud services, and it may well have better technology there than anyone else. The problem is that for most IT organizations, beating the World Champion in

the game of Go is far less important than very pedestrian issues like security, databases, load balancing, etc. Google has a penchant for doing really cool, advanced stuff, but that may not be helpful if it distracts the company from the fundamentals of enterprise computing.

So the central and yet to be answered question is this: *Is there room for Google and other providers like IBM to become major players in the cloud at this point or is the cloud mainly going to be an ecosystem showdown between AWS and Azure?* The cloud revolution, like the PC revolution, will likely take decades to play out in full, but we will probably get a good idea of the answer to the question within the next few years.

Data gravity in action

Any time you want to move around large amounts of data, you face challenges. The most obvious one is that data movement is somewhat expensive. Public cloud providers are quite generous when it comes to letting you upload data to their cloud storage. Typically, the upload itself is free. However, you still have to figure out how to get the bandwidth to do so efficiently. In some cases, the old-fashioned “sneaker net” – the movement of physical disks from one location to another – may be the most economical option. AWS, with its Snowball appliance, lets you copy your data locally onto a disk device and ship that device to Amazon. To quote the AWS pitch: *“Even with high-speed Internet connections, it can take months to transfer large amounts of data. For example, 100 terabytes of data will take more than 100 days to transfer over a dedicated 100 Mbps connection. That same transfer can be accomplished in less than one day, plus shipping time, using two Snowball appliances.”*

In addition to the upload bandwidth issues, there are cost issues with the cloud providers themselves for downloads: While public cloud providers encourage you to send data into their clouds, they don’t encourage you to extract it in large quantities. Storing a Gigabyte in the cloud will typically cost you a couple of pennies per month. If you want to extract that Gigabyte to outside of that cloud, be prepared to pay your cloud vendor close to a penny in addition to paying whoever provides the bandwidth for your side of the extraction. In other words, if you move data around a lot, it may well cost you much more than simply storing it.

But in addition to the obvious bandwidth and cost issues, there are other reasons why people might want their data to stay put:

- Consistency – keeping a consistent version of all your data has its own challenges even when the data is stored in a single, non-distributed database. Shipping data around adds a layer of complexity to the consistency issues.
- Security – shipping data around can sometimes make security-minded people a little nervous.
- Ease of management – moving data between different platforms and ecosystems is most likely an impediment to manageability.

- Backup and recovery – how does moving data around fit into the picture?

So for workloads that involve large quantities of data, it's most likely that the data will stay put and the workload will be executed in whatever cloud in which the data resides. This scenario is in sharp contrast to the hybrid cloud vision of easily being able to move workloads between different clouds. And data and database services are very important in the cloud. When AWS launched Redshift, its data warehousing service, it became the fastest growing service in AWS history but was later surpassed by the Aurora database service in that category. The accumulation of large quantities of hard-to-move data in AWS represents a problem for other vendors that want to compete with Amazon for existing AWS customers.

An interesting example of data gravity is Teradata's move to become a service on AWS and Azure. Teradata is a longtime successful vendor of on-prem database appliances and it has its own hosted cloud service. Its combination of proprietary hardware and software has been one of the leaders in highly scalable database technology. So the move to offer the software part on the AWS and Azure marketplaces running on Amazon and Microsoft infrastructure is likely because it thinks that those clouds will contain large quantities of data in the future. To paraphrase Sir Francis Bacon: *"If the data won't come to Teradata, then Teradata must go to the data."* So far, there has been no announcement of Teradata running on GCP.

Conclusion

It will be interesting to see how the cloud plays out in terms of successful ecosystems. Very likely, there will be consolidation where a small number of players will dominate and the rest of the field is made up of some insignificant niche players. So if this scenario plays out, will Google be one of the dominant players? It's too early to tell, but just as Google's efforts in the social network space were derailed by the success of Facebook, it's very possible, and perhaps more likely than not, that its public cloud effort will fail to catch up with AWS and Azure.