Economics of Cloud Computing

Ronnie Brunner, Jason Brazile
Netcetera
420
Main points to take away

> It can be win – win
You can make money in the Cloud by letting somebody save money in the Cloud and that’s not a contradiction

> You need (and want) to consider the big picture
The Cloud is not just about the CPU price per hour, but it’s about preparing your IT to be as flexible as possible, whether you’re a user or a provider

> The basic mechanics are easy to understand
You don’t need a degree in finance to take the first two points home
Economics Primer

Relax, *we won’t go there*, but… some terms must be known

> **Marginal cost and value**
  Think: peanuts for an additional passenger on a plane vs. the price he pays for the ticket

> **Opportunity cost**
  Think: pleasure gained when going to the movies vs. studying (and getting good grades and for these being admitted to college)

> **Time value of money**
  Think: 100 bucks is the same as 105 bucks in a year if you get 5% interest, but what’s 10$ in 2 hours if you’re hungry now?
Cloud Computing in a Nutshell

Cloud Computing supports those who wish to "try first, justify second"  
- James Staten, Forrester

Cloud Computing is
> Self-service
> On-demand
> Pay-as-you-go *

Cloud Computing allows transforming CapEx to OpEx
- William Fellows, 451 Group
6 of 10 Laws of Cloudonomics (by J. Weinmann)

User arguments for the Cloud:
> Utility services cost less even though they cost more
  If it’s pay as you go it costs more when you go, but nothing when you don’t
> On-demand trumps forecasting
  As soon as your forecast is wrong, you’ll realize this is true

Provider arguments for the Cloud:
> The peak of the sum is never greater than the sum of the peaks
  What? Demand peaks don’t all occur at the same time for all users
> Aggregate demand is smoother than individual
  Utilization of resources for many users is better than for an individual user
> Average unit costs are reduced by distributing fixed costs over more units
  Driving a car for two instead of one person reduces the cost for the first by 50%

True for both:
> Don’t put all your eggs in one basket
  Redundancy increases reliability
The Cloud Computing Service Models

- **Software as a Service (SaaS)**
  - enStratus
  - RightScale
  - cloudkick
  - SCALR
  - Path
  - newScale
  - zimory
  - libcloud
  - nimsoft

- **Platforms as a Service (PaaS)**
  - force.com
  - hadoop
  - intuit
  - appscale
  - Engine Yard

- **Infrastructure as a Service (IaaS)**
  - Amazon web services
  - rackspace
  - bluelock
  - terremark
  - CloudSigma
  - Eucalyptus

**Cloud Enablers / Cross platform solutions**
Market Status

> IaaS: Convergence only on basics (compute, storage)
> PaaS: Domain coverage poor (limited to web, map/reduce)
> SaaS: Very diverse, some already established, and "we’ve only seen the beginning"

Moving target:
> Consolidation expected
> New players expected
> Platforms winning importance
> Total market will grow from ~ $40 billion in 2011 to >$240 billion in 2020

Forrester, “Sizing The Cloud”, April 2011
Amazon/EC2 has a lot of competition

Ben Lorica, Amazon’s cloud platform still the largest, but others are closing the gap
Cloud Economics

Basic principles:
> Economies of scale
> Multi-tenancy
> Higher price, for a shorter time
> Utilization/burstiness † key factors
> Characteristics:
  – Self-service
  – On-demand
  – Pay-as-you-go
> Elastic capacity
> CapEx † OpEx
Economies of Scale

> It’s more than a quantity discount (BTW: there is an upper limit)

COST OF A PETABYTE

- RAW DRIVES
  - $81,000
- BACKBLAZE
  - $117,000
- DELL
  - MD1000 $826,000
- Sun
  - X4550 $1,000,000
- NAS-5000
  - $1,714,000
- amazon
  - AMAZON S3 $2,806,000
- EMC
  - EMCS NS-960 $2,860,000

* Amazon S3 Storage over three years (minus electricity, co-location and administration).

Backblaze blog, “Petabytes on a budget”

Rich Miller, “Who has the Most Web Servers” (2009)
On-demand, Self-service, Pay-as-you-go

User (customer)

> On-demand
  - Low contract periods
  - Now, not later
  - No idling capacity
  - Only use what’s really needed

> Self-service
  - No need to wait for service
  - A must for on-demand

> Pay-as-you-go
  - No upfront investment
  - Pay only for what was/is actually used

Provider

> Multi-tenancy
  - Allows self-service, on-demand
  - “Zero” marginal cost for a new customer

> Provision for peak demand
  - Needs some idle capacity

> Elastic capacity
  - Has more than any individual customer needs
Capacity provisioning

> User has need for capacity
  - Underprovisioning (1)
  - Provisioning for peak (2)
> Cloud has **elastic** capacity (3)
  (i.e. way more than what the user needs)
> User should get exactly the capacity from the Cloud that is actually needed (3a)

Why does this work for the provider?
Varying demand is **statistically smoothed out** over (very) many users
Basic value proposition example

User

> A dedicated DC would cost 200’000€
> Utilization would be max 50% (no need for the resources at night)

Provider

> The provider buys the equivalent of the DC for 150’000€ (quantity discount)
> The provider can provide this service for 100’000€ (smoothed out demand, better utilization than 50%)

Here’s the deal

> You pay a premium price to be able to use it on-demand and pay-as-you-go: 150’000€ for the 50% you need it (during the day only)
> You save 50’000€, the provider makes 50’000€: a win-win scenario
**Sample Cases: Calculating pricing**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Cost Overview</strong></td>
<td><strong>Scenario 1</strong></td>
<td><strong>Scenario 2</strong></td>
<td><strong>Scenario 3</strong></td>
</tr>
<tr>
<td></td>
<td><strong>1 week run</strong></td>
<td><strong>2 week run</strong></td>
<td><strong>4 week run</strong></td>
<td><strong>6 week run</strong></td>
</tr>
<tr>
<td>2</td>
<td><strong>External Cost</strong></td>
<td>645,301.51</td>
<td>680,074.87</td>
<td>749,851.73</td>
</tr>
<tr>
<td>3</td>
<td><strong>Internal Cost</strong></td>
<td>912,276.72</td>
<td>912,276.72</td>
<td>912,276.72</td>
</tr>
<tr>
<td>4</td>
<td><strong>Cost with Cloud Service</strong></td>
<td>1,557,578.23</td>
<td>1,592,351.59</td>
<td>1,661,928.45</td>
</tr>
<tr>
<td>5</td>
<td><strong>Cost without Cloud Service</strong></td>
<td>7,114,942.72</td>
<td>4,432,192.72</td>
<td>2,757,442.72</td>
</tr>
<tr>
<td>6</td>
<td><strong>External Cost</strong></td>
<td>6,916.00</td>
<td>6,916.00</td>
<td>6,916.00</td>
</tr>
<tr>
<td>7</td>
<td><strong>Internal Cost</strong></td>
<td>7,108,026.72</td>
<td>4,425,276.72</td>
<td>2,750,526.72</td>
</tr>
<tr>
<td>8</td>
<td><strong>Cost savings with Cloud Service</strong></td>
<td>5,557,364.49</td>
<td>2,839,841.13</td>
<td>1,095,514.27</td>
</tr>
<tr>
<td>9</td>
<td><strong>Percentage</strong></td>
<td>78.11%</td>
<td>64.07%</td>
<td>39.73%</td>
</tr>
</tbody>
</table>

**Configuration**

- **Date range (years)**: From
- **NPV as of year**: 1
- **Currencies**: 1.00 CHF, 1.00 EUR, 1.00 USD
- **Discount rate (for NPV)**: 10%

**EC2 cost ($) vs utilization (%) @1TB/month**

At 25% utilization, cheaper to reserve

At 3% utilization, even 100s CPUs are “cheap”
Cloud SLAs as good as average data centers

<table>
<thead>
<tr>
<th>Who</th>
<th>Promised uptime</th>
<th>Achieved Uptime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Center (avg. over 300)</td>
<td>n/a</td>
<td>99.5% (~50min/week)</td>
</tr>
<tr>
<td>Top 15% Virtual Systems Management Enterprises</td>
<td>n/a</td>
<td>99.999% (~5min/year)</td>
</tr>
<tr>
<td>Amazon EC2</td>
<td>99.95% (~5min/week)</td>
<td>unknown</td>
</tr>
<tr>
<td>Google Apps Premier Edition</td>
<td>99.9% (~10min/week)</td>
<td>99.85% (~15min/week)</td>
</tr>
<tr>
<td>Microsoft Business Productivity Online Suite</td>
<td>99.9% (~10min/week)</td>
<td>unknown</td>
</tr>
</tbody>
</table>

Data Source: Andi Mann, Enterprise Management Associates
Net neutrality and the power of big

Google/YouTube
> 2009 CS: “YouTube $470M loss”
  but…
> Didn’t account for 73% peering
> Cheap hosting e.g. Iowa & Finland

And in reality…
> Gain: “YouTube appears as loss”
> “YouTube key for reducing general Google operational costs”
> “Caught up to e.g. Microsoft”

Netflix/Level 3 vs Comcast
> 2010 Level 3: “Comcast unilaterally set price for what used to be free”
  but…
> Comcast: “fees for 5:1 imbalance”

And in reality…
> Level 3/Comcast used to peer free
> L3: new Netflix “big data” customer
> Comcast: many “last mile” users

Net Neutrality?

Gannes, YouTube Infrastructure Costs Vastly Overestimated, GigaOM, 2009
Miller, Level 3 vs Comcast: More than a peering spat?, 2010
Netflix in Cloud is U.S.'s largest traffic source

Netflix Turns From Oracle, IBM to Amazon to Save Cash

By Chris Kanaracus, EE Times  Nov 24, 2010 8:00 pm

Netflix moved some of its most crucial IT operations over to Amazon Web Services' Elastic Compute Cloud in order to save money and gain flexibility compared to using more Oracle software and IBM iron.

"Our datacenter runs Oracle on IBM hardware, we could have switched to commodity hardware in a data center, but skipped that step by going to AWS," Netflix cloud architect Adrian Cockcroft told the consulting firm Cloudscaling in an interview posted Tuesday. "There are three points on cost, one is that Oracle on IBM is very expensive, so AWS looks cheap in comparison, and we have fully maxed out our datacenter capacity."

In addition, Netflix "could not have hired enough [system and database administrators] to build out a own data center the fast. We have added 4X as many systems in the cloud as the total we have in our data center over the last year," he said.

Finally, EC2's pay-as-you-go model means costs are elastic: "If you own a resource it sits around a long time waiting to be delivered and installed, and if you no longer want to use that type of resource you are still paying for it for three years."

Cockcroft's remarks add some color to Netflix's announcement in May that it would significantly expand use of AWS. The company had already been using the service for various customer-facing and internal applications, but decided to add "critical pieces" of its service, including member move lists, recommendation engine and fine transcoding.

The move enabled Netflix to free up "scarce engineering resources from the undifferentiated heavy lifting of running its own infrastructure," it said at the time.

Netflix's skyrocketing customer count, which now stands at about 16 million, also made the job of running and expanding data centers too unpredictable, according to a presentation Cockcroft gave the recent QCon conference. In addition, the company has been rapidly transitioning from a DVD-delivery outfit to a mainly streaming operation.

For actually streaming the movies to customers, Netflix contracts with companies like Akamai and Limelight, Cockcroft told Cloudscaling. It also runs systems for account sign-up, billing and other needs elsewhere, he added.

Schonfeld, TechCrunch, Netflix now largest single source of internet traffic in N. America, 17 May 2011
## Cloud consumer risks and their consequences

<table>
<thead>
<tr>
<th>Risk</th>
<th>Examples</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual “contracts” via credit card</td>
<td>Critical service is down because key person's individual credit card expires</td>
<td>Service failure, data mess (where’s what?)</td>
</tr>
<tr>
<td>Single actor can chose wrong direction quickly</td>
<td>Introduction of a proprietary SaaS solution that (only) provides a quick fix</td>
<td>Unmanaged service portfolio, not reaching strategic goals</td>
</tr>
<tr>
<td>Costs can't be tracked well</td>
<td>Monthly bills unpredictable due to irregular demand. Lots of hard to track small transactions with many providers</td>
<td>Financial exposure and uncertainty</td>
</tr>
<tr>
<td>Costs slowly increase</td>
<td>Nobody cleans up hard disks or gets rid of unused virtual machines</td>
<td>More expensive over time, unclear what's still needed</td>
</tr>
<tr>
<td>Optimism bias</td>
<td>Assumptions that all costs will go down (and all performance up) just from moving to the cloud</td>
<td>Situation worse and/or more expensive and no plan B</td>
</tr>
<tr>
<td>Data gets leaked</td>
<td>Data protection violation, leak of industry partner’s secrets</td>
<td>Financial liability, loss of trust</td>
</tr>
<tr>
<td>Lock-in dependency</td>
<td>Usage too specifically designed for a specific provider or lack of alternative service</td>
<td>Exposure to outage, higher switching cost</td>
</tr>
<tr>
<td>Data loss</td>
<td>NASA’s moon landing tapes, hacker data vandalism, Provider default</td>
<td>Image/brand damage</td>
</tr>
</tbody>
</table>
Cloud provider risks and their consequences

<table>
<thead>
<tr>
<th>Risk</th>
<th>Examples</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commoditization of everything</td>
<td>20+ providers of cloud compute nodes and storage services</td>
<td>Race to the bottom - mainly about price</td>
</tr>
<tr>
<td>Visible customer gets burned</td>
<td>April 2011: Reddit, Foursquare, Quorum suffered from AWS EBS failure</td>
<td>Financial liability, loss of trust</td>
</tr>
<tr>
<td>Infrastructure gets hacked</td>
<td>Dec 2010: Microsoft BPS Cloud Service Data Breach</td>
<td>Image/brand damage, loss of trust</td>
</tr>
<tr>
<td>Can’t deliver on uptime</td>
<td><a href="http://en.wikipedia.org/wiki/Rackspace#Downtime">http://en.wikipedia.org/wiki/Rackspace#Downtime</a></td>
<td>Lose regular customers and ability to keep tenancy high</td>
</tr>
<tr>
<td>Can’t maintain low marginal costs</td>
<td>Too much per-customer customization prevents streamlined provisioning/operations</td>
<td>Limited niche market</td>
</tr>
<tr>
<td>Legislation threatens business model</td>
<td>Google in China</td>
<td>Change product (e.g. allow censorship) or abandon market</td>
</tr>
<tr>
<td>Can’t grow customer base</td>
<td>Microsoft Azure</td>
<td>Loss of momentum</td>
</tr>
</tbody>
</table>

1 Gray, Microsoft seeks to stem Azure exodus with huge appliance, Informed Virtualization Criticism, 2010
2 http://jpf.github.com/domain-profiler/ycombinator.html
Six Cloud Scenarios - (1/3)

Scenario: **Offline compute, <80GiB/day**
Examples: **Scientific computing, Data mining, Audio recoding, etc**

- Turnaround time deadlines ✱ flexible
- No storage other than “local” disk
- 10 Mbps @80% ✱ 80GiB ~23 hours¹
- Vendor price/performance changes often – helps to constantly measure

**Costs:**
- Cloud deployment: **low/med**
- Disaster recovery cost: **very low**
- Brand/Security risk: **very low**

Scenario: **Data in cloud >80GiB/day**
Examples: **High volume website, Disaster recovery backup, etc**

- Storage/network transfer dominate
- Block vs service e.g. EBS vs S3
- reduced redundancy ✱ cheaper ?
- RAIN (redund. array inexp. nodes) ?
- CDN ✱ faster net, wider dist. ?
- media shipping for import/export ?

**Costs:**
- Cloud deployment: **low/med**
- Disaster recovery cost: **med/high**
- Brand/Security risk: **med** (encryption)

¹ [http://www.wolframalpha.com/input/?i=80GiB+at+(0.8+*+10)+Mbps](http://www.wolframalpha.com/input/?i=80GiB+at+(0.8+*+10)+Mbps)
Six Cloud Scenarios (2/3)

Scenario: **Scalable large content with critical latency**
Examples: (Spiky) media distribution

> Cost is non-goal (can’t DIY)
> Likely case for a CDN (Global reach, Avoid congestion, Great for media streaming)
> Safe to assume that costs decrease

Costs:
> Cloud deployment: **med**
> Disaster recovery cost: **low**
> Brand/Security risk: **med**

Scenario: **Freemium SaaS model**
Examples: Xing, Basecamp, ZoHo...

> Free vs. Premium (Feature limited, Time limited, Capacity limited, Seat limited, Customer class limited)
> If lifetime value of a customer - acquisition cost > cost to operate the service (incl. the free users)

Costs:
> Cloud deployment: **med/high**
> Disaster recovery cost: **med**
> Brand/Security risk: **med/high**
Six Cloud Scenarios (3/3)

Scenario: **Website with little computation, little data**
Examples: **SME web site or small shop**

- IaaS: very cheap (first glance)
- PaaS: very easy

Costs:
- Automation (Platform or DIY on IaaS)
- No worries about scale (Platform)
- Cloud deployment: low
- Disaster recovery cost: med/large
- Brand/Security risk: med

---

Scenario: **SaaS calculation example**
Examples: **Large corp. introducing SaaS**

- “[…] cut travel costs by at least 5%, […] bill for travel over €40M/a”
- Shared desktop and virtual conference software pilot
- Assume 1K€/d and 3 travelers / event
- Costs offset if 500 use it instead of travel once per year

Costs:
- Some One Time Cost (OTC)
- Monthly Recurring Cost (MRC)
Consequences of shifting operations to the cloud

Implications for IT operations
> Developers responsible for ops
> Master data (and backups) in cloud
> IT dept is Amazon/Rackspace/Linode
> Cloud capacity >= datacenter
> No staff doing platform infrastructure

Many traditional IT roles go away
> No System Administrators
> No Database Administrators
> No Storage Administrators
> No Network Administrators

Capacity planning (old thinking)
> Capacity expensive
> Capacity takes time to buy/provision
> Capacity increases not shrinks
> Capacity big chunks, $ up front
> Planning errors big problems
> Systems are clearly assets
> Depreciate assets over 3 years

Implications for application design
> Although uptimes are good in aggregate, must design for failure
> Cloud persistence services have widely varying latency

http://www.slideshare.net/adrianco/netflix-in-the-cloud-2011
Main points to take away

> It can be win – win
   Paying a premium price for less time results in savings allowing the provider to get the premium

> You need (and want) to consider the big picture
   The Cloud helps not only create cheaper, but also better (more reliable) services. All things considered, you’d want to go Cloud even if it was more expensive, just because it’s better

> The basic mechanics are easy to understand
   You don’t need a degree in finance to take the first two points home unless we didn’t do our job today