CLOUD: ECONOMICS

Prepared by:
Duncan Rutland, Advisory Services Enterprise Architect, Rackspace® Hosting
# Introduction

Many organizations perceive cost reduction as one of the primary benefits of adopting a cloud hosting model. In practice however, this is not always an accurate assumption. There are nuances to the financial analysis: public cloud computing is not necessarily cheaper than traditional dedicated hosting. There are additionally considerations around the impact on Total Cost of Ownership (TCO), Transformation/Migration costs and the position of an organization within the IT lifecycle. In this whitepaper, we will explore each of these areas in order to illustrate some of the factors an organization needs to consider when contemplating a migration to the cloud.

## Table of Contents

1. Introduction
2. Cloud: Cheaper than Dedicated, Right?
   2.1 Not Necessarily...
   2.2 So How Do We Reduce costs?
   2.3 What Shape is Your Demand Graph?
   2.4 The Diurnal Cycle
3. Will We Reduce Costs?
   3.1 CAPEX to OPEX
   3.2 Measuring TCO
   3.3 TCO Reduction Areas
     3.3.1 TCO Perspective
     3.3.2 Human Perspective
   3.4 Transformation Costs
   3.5 Position in the IT Lifecycle
4. Calculating Investment Return
   4.1 ROI Formula
   4.2 ROI Example
   4.3 Net Present Value
   4.4 NPV Example
5. Conclusion
2 Cloud: Cheaper than Dedicated, Right?

2.1 Not Necessarily…
A common misconception in the marketplace today is that public cloud computing is cheaper than dedicated hosting. Although it surprises many to hear this, a unit of cloud computing is in fact more expensive than a unit of dedicated hosting. The reason is due to the very short period of commitment that cloud service providers require for cloud; typically, a public cloud is metered in 1 hour increments. However, a traditional dedicated hosting contact requires a 12, 24 or 36 month commitment from the customer. Therefore, to capitalize on the short commitment period offered to the consumer and offset the risk of unused resources, public cloud service providers can charge a premium. To use an analogy from the automobile industry: if you need a vehicle for 10 minutes, you take a taxi (e.g., a 10-minute trip that costs $10 = $60/hr unit cost). However, if you need a car for a day it is more economic to rent one (e.g. a $50/day rental = ~$2/hr unit cost).

2.2 So How Do We Reduce Costs?
Given that the unit cost of public cloud is higher than dedicated hosting, how can consumers reduce costs? The answer is to exploit variation in demand. If there is significant variation in demand, there is opportunity to exploit this differential and reduce operating expenditure by matching the supply of resources to the level of demand.

The pattern of demand for an application and density of the supply side (i.e., number of compute nodes or servers) are the primary drivers of whether there is a financial advantage to running that application in a multi-tenant environment, where metered billing can allow the supply of resources to scale up/down in order to meet demand. To realize this benefit:

- The peak/average demand ratio must be greater than the cost ratio of public cloud over dedicated hosting.
- The application must be able to scale horizontally.
- There must be sufficient density of compute nodes to allow the scaling of resources.

There are also licensing implications when transitioning from a scale-up architecture to a scale-out architecture: some applications are licensed per-instance or per-CPU, often over an annual term. In this instance, there can be significant cost implications of adding new instances to a pool of resources. In time, application vendors will follow infrastructure service providers in moving to more flexible pricing models such as per core/hr or per request/transaction. The alternative is to use Open Source Software (OSS) where the license cost issue is non-existent.
2.3 What shape is your demand graph?

Example 1: Periodic Usage

The periodic or “on and off” workload is characterized by periods of relatively high activity interspersed with periods of little or no activity. Common examples of this demand pattern are:

- Batch processing
- Analytics
- Data warehouse

The very high peak-to-average utilization ratio of this workload usually translates to the potential for significant cost savings under a purely public cloud model, since the majority of resources can be “spun-down” when not in use.

Example 2: Spiked Usage

The bursting workload is characterized by baseline periods of “normal” activity interspersed with significant spikes in activity (often of one or more orders of magnitude). Common examples of this demand pattern are:

- Seasonal demand (e-commerce)
- Event driven (sports/entertainment)
- Campaign driven (marketing/advertising)

Since there is a baseline of demand, the hybrid model is often most cost-effective in this scenario: dedicated resources satisfy normal demand, supplemented by public cloud resources during spikes.

Example 3: Cyclical Usage

The cyclical workload is characterized by a regular ebb and flow of demand, usually following a sinusoidal pattern. This pattern usually occurs over a period of 24-hours, and is linked to local patterns of consumption that occur within geographic proximity to the physical infrastructure. Ultimately, this pattern is tied to the natural rhythms of the human sleep/work cycle.

The hybrid model is usually the most cost-effective in this scenario, since there is a baseline of demand throughout the cycle (albeit lower than in the bursting scenario). However the ability to exploit this effect is dependent on the application being able to quickly scale up/down (preferably autonomously) to meet demand.
One thing that all three of the demand patterns on the previous page have in common is that they are not flat. Since the unit cost of public cloud hosting is higher than dedicated hosting, a flat pattern of demand will always be cheaper using fixed or dedicated resources on a traditional 12/24/36 month contract term. So, if demand is flat the utility billing model will not reduce costs.

2.4 The Diurnal (Daily) Cycle
Fortunately, most workloads (when viewed locally) do not exhibit a flat demand graph. This is due firstly to the fact that most enterprise applications are consumed by users within local proximity to the underlying infrastructure (for example, users within an office or within the same country or continent). Combine this fact with a basic feature of human behavior: people need to sleep. The result is the typical diurnal cycle of demand over a 24-hr period that correlates with the sleep patterns of the end consumer.
3 Will We Reduce Costs?

3.1 Capex to Opex
One of the most over-used but inadequately discussed mantras when extolling the benefits of cloud computing is “Convert CAPEX to OPEX”; or put differently: rather than make large capital expenditure on the physical infrastructure (servers, network, storage) required to run an application, why not convert this to operational expenditure and rent capacity from a service provider. While this statement is justifiably relevant for green-field projects and cash flow sensitive organizations such as start-ups, for enterprises with established application portfolios the situation is more complex. The reason for this is that the majority of costs associated with an IT project are operational; CAPEX is typically only a fraction of the TCO (Total Cost of Ownership) of an application. Since the bulk of expenditure over the lifetime of an application is not related to the purchase of physical infrastructure, it therefore makes sense to examine the impact of cloud computing on TCO rather than just CAPEX.
3.2 Measuring TCO

When examining the TCO of an application, there are many factors to incorporate:

- **Real Estate**
- **Utilities**
  - Power
  - Water
- **Tangible Assets**
  - Mechanical / Electrical Infrastructure (Power/Cooling)
  - Server Hardware
  - Network Hardware
  - Storage Hardware
- **Intangible Assets**
  - Hardware Maintenance Contracts
  - Software License Contracts
  - Software Support Contracts
- **Human Labor**
  - Data Center Engineers / Operatives
  - Network Engineers / Administrators
  - Storage Engineers / Administrators
  - Systems Administrators
  - Data Center Administrators
  - Software Developers / Architects
  - Help Desk Operatives
  - General & Administrative
  - Managerial
- **Auditing / Compliance**
  - Power
  - Water
- **Downtime / Outages**

The pertinent question: can you quantify your current TCO? For most enterprises, the answer is “Not really.” While accounting should have a good handle on assets and utilities, most organizations are not able to accurately measure the TCO for a particular application. This is primarily due to inadequate auditing and accounting of resource consumption specific to a particular application or service. For example, does your organization measure for the amount of human labor consumed to operate each application in your portfolio? In most cases, the answer is, “No.” For this reason, breaking down TCO accurately per application is very tricky if not impossible.
3.3 TCO Reduction Areas

Outsourcing the hosting of an application to an external cloud service provider affords the opportunity to reduce TCO. The degree of reduction that is realizable is dependent on the type of cloud hosting model being considered. There are three main categories of cloud computing: IaaS, PaaS and SaaS. The NIST\(^1\) definitions are included below:

### IaaS

"The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls)."

### PaaS

"The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment."

### SaaS

"The capability provided to the consumer is to use the provider’s applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings."

Depending on the model chosen, the consumer outsources varying degrees of the technology stack underpinning their application. With each additional layer that is outsourced, the consumer can potentially reduce costs by leveraging the service provider’s economies of scale while sacrificing control and flexibility.

---

### 3.3.1 TCO Perspective

The table below shows the technology stack underpinning an application according to broad TCO areas, and describes the areas of responsibility for both consumer and service provider in terms of the three key cloud models IaaS, PaaS and SaaS:

<table>
<thead>
<tr>
<th>Component</th>
<th>IaaS</th>
<th>PaaS</th>
<th>SaaS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Process</td>
<td></td>
<td>CONSUMER</td>
<td></td>
</tr>
<tr>
<td>Business Logic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middleware Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application Licensing / Support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS Management</td>
<td></td>
<td>SERVICE PROVIDER</td>
<td></td>
</tr>
<tr>
<td>OS Licensing / Support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Server / Storage / Network HW / Maintenance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME Equipment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real Estate</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is clear from this picture that under an IaaS model, while infrastructure related costs are assumed by the service provider, the consumer is still responsible for a significant portion of TCO that occurs at the OS layer and above. This situation is improved considerably under a PaaS model, but the consumer is then effectively locked into a particular PaaS ecosystem such as Google AppEngine, Microsoft Azure, Force.com, Engine Yard, Heroku and Amazon (when using their “platform” services such as SQS, SNS, SimpleDB/RDS). This lock-in comes at a price: the consumer must cede varying degrees of flexibility and control. For many organizations, this is an unacceptable compromise and the IaaS model is chosen as it allows the greatest degree of customization. The SaaS model provides the greatest degree of outsourcing of responsibility, but comes at a far heavier cost in terms of loss of control and customization.
3.3.2 Human Perspective

The table below shows some of the more common roles that humans perform when managing an application in-house, and describes the areas of responsibility for both consumer and service provider in terms of the three key cloud models IaaS, PaaS and SaaS:

<table>
<thead>
<tr>
<th>Component</th>
<th>IaaS</th>
<th>PaaS</th>
<th>SaaS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Process</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business Logic</td>
<td>CONSSUMER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middleware Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application Licensing / Support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS Licensing / Support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Server / Storage / Network</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HW / Maintenance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME Equipment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real Estate</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This view shows the human perspective of TCO as it relates to technology. Notice that under IaaS, the consumer must still employ expensive resources such as System Administrators, DBAs, developers etc.
3.4 Transformation Costs

When conducting a Cost/Benefit Analysis or producing a business case for an application to be migrated from a traditional in-house hosting context to an outsourced model, the transformation costs involved in moving the application from source to destination must be evaluated. There are several crucial areas that need to be considered:

- **Application Development/Re-architecture:** Modifications to the logic and/or architecture of an application are often required when performing a migration between different hosting contexts. This can be due to different technology stacks at source and destination, integration with back-end infrastructure management systems, variations in latency that affect application performance and many other factors. Time consumed by Project Managers, Developers, Release Managers & Testing must be incorporated into the overall transformation cost.

- **Migration Planning and Execution:** Moving a complex system with many moving parts between hosting contexts must be planned and choreographed carefully to minimize the risk of disruption or failure. Both the planning and execution phases of a migration require participation with various stakeholders such as Project Managers, Account Managers, Support/Help Desk, Systems Administrators and so forth.

- **Migration Downtime:** Most application migrations will require a total service outage while data is synchronized, tests performed and clients redirected to the new location. The business impact of this outage must be assessed and defined in terms of revenue impact, and factored in to the transformation cost.

- **Auditing and Compliance:** If your application is subject to a compliance standard such as SSAE-16/SAS-70, ISO-27001, PCI-DSS, HIPAA or FISMA, then the costs associated with re-auditing and re-certifying the application in its new hosting context must be factored into the transformation cost.

- **Process Re-engineering:** Being components that reside within a larger frame of reference, the business process, migrating applications to another hosting context will require these processes to be examined and re-engineered accordingly. One prime example is around Business Continuity Planning (BCP), where such a migration has profound implications on the business processes around detecting, evaluating and executing a failover to a disaster recovery solution. Support/Operations are also fundamentally affected by changes in hosting context.

If you are using the traditional Return On Investment (ROI) formula to determine the cost effectiveness of migrating an application to an outsourced hosting model, all of these costs must be incorporated into the “Investment” portion of the calculation. When using the Net Present Value (NPV) calculation, these costs should be represented at t=0 as a negative value. An example using both of these types of calculation will be provided in Section 4.
3.5 Position in the IT Lifecycle

Typically, a technology solution is accompanied by a large capital expenditure required to purchase the tangible and intangible assets required to run a business application: servers, network/storage devices, OS and application licensing and support maintenance contracts are all required when hosting an application. These assets are typically depreciated or amortized over a fixed period of time, usually 36-60 months. Thus from a P&L/Income Statement perspective, the age of these assets is another factor to consider in determining the economic impact of migrating.

- What degree of depreciation has occurred for your physical assets such as servers and network/storage devices?
- What degree of amortization has occurred for your intangible assets?

For assets that are not fully depreciated/amortized, and that cannot be re-purposed, a write-off may be required which should also be factored into the transformation cost equation.
4 Calculating Investment Return

Having an understanding of the Total Cost of Ownership (TCO) of your current application, the expected savings from moving to an outsourced hosting model, and the expected cost of transforming the application to a hosted context, it is possible to estimate how much your organization can save (or put another way, estimate the return on investment realized) by such an outsourcing decision.

4.1 ROI Formula

The standard method of calculating the financial return on a technology investment is the traditional ROI formula

\[
\text{ROI} = \frac{\text{gain from investment} - \text{cost of investment}}{\text{cost of investment}}
\]

Gain from investment = Current TCO - Target TCO
Cost of investment = Transformation costs

This method simply takes the gains realized by migrating to the target context (Current TCO minus Target TCO), and divides this figure by the Cost of Investment, or the transformation costs.

4.1 ROI Example

In the example below, we assume that a TCO saving of $200K per year is possible, which equates to $600K over 3 years. The transformation costs required to realize this migration are assumed to be $400K. Thus, over a 3 yr period the return on the initial investment of $400K is 50%.

\[
\text{ROI} = \frac{600,000 - 400,000}{400,000} = 50%
\]

Gain from investment (TCO saving) = $200,000 per year
Cost of investment (Transformational costs) = $400,000

The main flaw with this method is that it does not take account of the time value of money: simply put, a dollar today is worth more than a dollar in the future. This is due to factors such as inflation, and that a dollar today can be invested and generate a return over time. Most accounting/finance departments would prefer to see a formula that takes into account the time value of money, such as Net Present Value (NPV)
4.3 Net Present Value

The NPV formula below is commonly used to appraise the Present Value (PV), the return on an investment over longer periods of time. It does this by discounting the future, or taking into account the time value of money. This is the preferred method used by accountants in determining the true return on an investment in present terms.

\[
\text{NPV} = \sum_{t=0}^{n} \frac{(\text{Benefits} - \text{Costs})_t}{(1 + r)^t}
\]

where:
\( r \) - discount rate
\( t \) - year
\( n \) - analytic horizon (in years)

4.4 NPV Example

The following example uses the same assumptions as the ROI example earlier: an initial transformation investment of $400K (represented at \( t=0 \)), an annual TCO saving of $200K and a 3 year analytic horizon. Additionally, it assumes a discount rate of 10%. The corresponding undiscounted cash-flow is shown for comparison purposes.

<table>
<thead>
<tr>
<th>( t )</th>
<th>Cash-Flow</th>
<th>PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-400,000</td>
<td>-400,000</td>
</tr>
<tr>
<td>1</td>
<td>200,000</td>
<td>181,818</td>
</tr>
<tr>
<td>2</td>
<td>200,000</td>
<td>165,289</td>
</tr>
<tr>
<td>3</td>
<td>200,000</td>
<td>150,263</td>
</tr>
<tr>
<td>NET</td>
<td>200,000</td>
<td>97,370</td>
</tr>
<tr>
<td>ROI %</td>
<td>50.00%</td>
<td>24.34%</td>
</tr>
</tbody>
</table>

The first thing to observe is that the $200K TCO saving is worth less and less over the 3 years period, being reduced to $150K in year three. Secondly, the effective return of $97K that is realized after 3 years is only 24% of the initial 400K investment, a significant correction of the 50% return obtained from the traditional ROI calculation. For this reason, the NPV method is usually preferred when assessing return on a technology investment.
5 Conclusion

As we have seen, there is more to cloud economics than the oft-touted “convert CAPEX to OPEX” platitude. Factors such as the pattern of demand, TCO and transformation costs need to be factored into any financial analysis of a potential cloud computing solution:

- Sufficient variability in demand is required to realize cost savings using a public cloud hosting model. Flat workloads are cheaper on dedicated infrastructure over a fixed contact term.
- It is not possible to accurately measure the current TCO of an individual application without first tracking the utilization of resources per application within your IT organization.
- Transformation costs associated with migrating an application from current to target context must be factored into the cost equation.
- Take account of the time value of money when performing ROI Calculations, using methods such as Net Present Value (NPV).

If you need help in assessing the feasibility of performing a migration of your legacy applications to a cloud hosting model, Rackspace can help. Contact the Advisory Services Team at 1-800-440-1249 or visit www.rackspace.com/AdvisoryServices for more details.