# Evaluating the Operational Risk in the Transition from an Infrastructure Asset to a Commodity: the Case of Cloud Storage

Loretta Mastroeni[a], Maurizio Naldi[b]

[a]*Department of Economics, University of Rome Tre, Via Silvio D'Amico 77, 00145 Roma, Italy*
*E-mail: mastroen@uniroma3.it*
[b]*Department of Computer Science, University of Rome at Tor Vergata, Via del Politecnico 1, 00133 Roma, Italy*
*E-mail: naldi@disp.uniroma2.it*

**Abstract**

The usage of third party storage facilities, known as cloud storage, is going to be a major commodity in information systems management. However, the transition from a privately owned infrastructure to a leased one is not always profitable. The economical risk associated to the transition is analysed through the Value-at-Risk indicator.

*Keywords:* Cloud storage, Information systems, Operational risk

## 1. Introduction

Cloud computing represents one of the major technological trends of the last decade. By cloud computing we mean the possibility that anything we do on a computing platform (on our individual PC or on a corporate workstation) can be done on an external computing platform, owned by a third party and located elsewhere, outside our premises (not in our house, not in our corporate data center). This is expressed by saying that we have moved to the cloud. The term *cloud* is used, since the actual location of the computing resource we exploit is not relevant (and actually unknown). According to the Cloud Computing Manifesto, The key characteristics of the cloud are the ability to scale and provision computing power dynamically in a cost efficient way and the ability of the consumer (end user, organization or IT staff) to make the most of that power without having to manage the

underlying complexity of the technology. [1]. Any resource typically present on a computer can be moved to the cloud and used remotely: we can exploit its computing power (i.e., using a CPU much faster and powerful than our PC); we can use an application resident on the cloud rather than buying the license to store in on our computer; we can use the huge storage capabilities of the cloud rather than relying on the limited space on our PC's hard disk. The trend is also spurred by the exponential increase of mobile devices connected to the Internet [2]. A survey of the several applications of cloud computing is contained in [3].

The cloud is already a reality. In many cases, we rely on an external source for the applications we run and the data we store, e.g., for e-mail or through a social network like Facebook or Twitter. We store files online, and maybe create documents, spreadsheets and presentations entirely online. "A lot of people are in the cloud and do not even realize it" [4].

It has been suggested that this move to the cloud model makes computing the fifth utility (along with water, electricity, gas, and telephone) [5]. Actually, moving to the cloud allows a company to get rid of most of its investments in IT and use the cloud instead, on a pay-per-use basis. The company reduces its IT infrastructure assets to a minimum, replacing them with the computing resources offered as a commodity by the cloud provider. Since the purchase of infrastructure assets is a major expense in data center operations, the move to the cloud represents a major shift in the financial management of IT operations: capital expenses are mostly replaced by operational expenses.

However, the move is not without risks. When the company gives up its own IT infrastructure, it is exposed to the prices offered by the cloud provider. Should they go up, the way back to resuming its own infrastructure would be neither fast nor cheap. The transition from an infrastructure to a commodity must therefore be pondered with a careful long-run economic analysis.

In this paper, we deal with the problem of evaluating the risk associated to that transition. We focus on cloud storage, i.e., the move of just data (not the applications, not the computing power) from the corporate date center to the cloud. We list and analyse all the expense items involved in both paradigms, and provide a methodological framework to compare them and decide whether the move is economically sound. We also provide some results of the application of the methodology to a typical scenario.

The paper is organized as follows. In Section 2, we review the background of cloud storage, and report the main business initiatives in the field. In

Section 3 we analyse all the expense items associated with the decision either to own the data storage facilities or to lease the storage space from the cloud provider. Finally, in Section 4, we employ the well known tools of the Net Present Value and Value-at-Risk to compare the two alternatives.

## 2. Cloud storage

The main thrust to cloud storage is the explosion of data storage needs and the growing requirement of ubiquitous access to data. In this section, we sketch the extent of those two phenomena and describe the cloud storage solution.

Data abound both in the residential and business environments. The development of digital consumer electronics has moved to the digital domain much of the content that was stored in an analog fashion, e.g., documents, photos, videos. The total amount of data generated in the world has grown from 2.6 exabytes in 1986 (1 exabyte equals 1000 billion megabytes, i.e., $10^{12}$ megabytes) to 295 exabytes in 2007, with a 23% compound annual growth rate (CAGR) over the latest 2 decades [6]. More recent data, extending into 2011, show that the total amount of generated digital data was 1699 exabyte in 2011, with a 55.7% CAGR over the years 2006-2011 [7]. As reported by Hilbert and López, roughly half of this volume of data is stored on hard disks, servers, and mainframe systems [6].

People also need to access data (their own and their company's) from anywhere. The growth in computing power has led many to move to laptops, which allows people to work basically anywhere. At the same time, the progress in mobile communications has spurred the use of the mobile communications infrastructure to access the Internet. Data do not need anymore to be physically attached to the computing device, but can be stored anywhere in the world and accessed through the Internet. This is the basic concept of cloud storage: storing the data on a remote data center, which is managed by third parties. The entity storing the data is named the *cloud*, because it may be distributed on several physical locations.

There are several advantages for companies to move their data to the cloud:

- Companies are relieved of their data center management task;
- Investments pertaining to the storage facilities are replaced by operational expenses

| Amount of data | Price per month and GB[$] |
|---|---|
| First 1 TB | 0.125 |
| Next 49 TB | 0.110 |
| Next 450 TB | 0.095 |
| Next 500 TB | 0.090 |
| Next 4000 TB | 0.080 |
| Over 5000 TB | 0.055 |

Table 1: Price list for Amazon S3

- Cloud solutions are highly scalable, i.e., they can easily adapt to the growing or falling needs of the user;
- Companies need only pay for their actual consumption of storage facilities;
- Being based on very large pools of servers and storage devices, cloud solutions are inherently more reliable than the lower-sized individual data repository.

On the other hand, consumers benefit from the cloud by being able to access their data from anywhere in the world through a simple web interface.

Though usage examples abound, an oustanding case is the pilot project *DuraCloud* launched by the US Library of Congress and the DuraSpace organization to test the use of cloud technologies to enable perpetual access to digital content. The project, which includes the New York Public Library and the Biodiversity Heritage Library, started in 2009 [8], and is now part of the National Digital Information Infrastructure and Preservation Program (NDIIPP).

A number of business initiatives are in place to provide cloud storage. The best-known is probably the S3 service (Simple Storage Service) run by Amazon [9]. You can store data on Amazon S3 by paying an amount for each gigabyte on a monthly basis. A recent pricing table is shown in Table 2, where you can see that the unit price falls progressively as the overall amount of data grows.

A popular cloud, aimed mainly at consumers, is Dropbox (`www.dropbox.com`), which is free up to 2 gigabytes. Rival cloud providers for consumers

are Mozy (`mozy.ie`), Carbonite (`www.carbonite.com/en/`), and Crashplan (`www.crashplan.com`), whose performance has been compared with Dropbox's in [10].

The interest in cloud services has also attracted the attention of a giant in the information processing sector. Google, after offering its cloud service to developers (`developers.google.com/storage/`), is launching the Google Drive initiative directed to consumers [11].

However, the transition to the cloud is not straightforward. Among the greatest matters of concern is security: who do the data stored on the cloud belong to in the end ? how are they protected from theft or corruption ? A recent survey shows that the fear is fading for small businesses as well [12].

## 3. Economics of cloud storage

When evaluating the transition to the cloud from an economical viewpoint, we must compare the two alternative situations, where the company decides either to buy and manage its own storage facilities or to lease the storage space provided by the cloud. The company incurs different expense items in the two cases. In this section, we identify the items that concur to the cash flows when the company decides either to buy or to lease.

If the company opts for the Buy decision, it has to purchase a number of disks (and pertaining controllers, i.e., the devices that enable the flow of data to and from the disks) with an overall capacity large enough to host its volume of data. These costs fall into the category of CAPEX (CAPital EXpenditures), since the disks are fixed assets that contribute to the company's activities. Such disks are subject to failures, and their useful life is random. We must take into account that they have to be replaced, so that these CAPEX are incurred not just at the beginning of the period of interest but also throughout it, to replace failed disks. The percentage of disks replaced each year is named the Annual Replacement Rate (ARR). Just a few extensive studies on disk failures are reported in the literature. The ARRs reported in [13] vary from a minimum of 1.7%, for disks in their first year of operation, to a maximum of 8.6% for 3-years old disks. In [14] a wider range is reported, with ARRs varying between 0.5% and 13.5% (3% being the most frequent value). On the other hand, we have to consider that disks owned by the company have anyway a salvage value, that must be accounted for in the comparison. We can assume that the disks still working at the end of period of analysis (a random value, given the random failure process) can

be sold on the secondary market as a fraction of the current market price for new disks of the same capacity.

Disk prices however fall over time because of the technological advances. An extensive survey conducted on disk prices every week over more than 5 years has shown that the price follows a decaying trend which is approximately exponential, with local fluctuations around that trend [15]. While in [15] prices were assumed to change deterministically according to that trend, in [16] the deterministic model has been abandoned in favour of a more realistic random model. Namely, the price has been assumed to follow a Geometric Brownian Motion, whose expected value decays exponentially over time.

Additional expenses incurred by the company, if it decides to buy its own disks, are the operational expenses to manage the disks. Namely, the company has to pay for the power consumption and the personnel in charge of managing the data center. A 24/7 service has been assumed for the disks, with a fixed unit cost per kWh.

Instead, if the company leases storage space from the cloud, the CAPEX disappear (a major advantage of the transition to the cloud), but are replaced by a larger set of OPEX. Namely, the company has to pay a leasing fee, which also caters for all the operational costs strictly associated to the storage device itself during the leasing period, such as the electrical power needed to run the devices (which under the Buy decision is instead an individual OPEX item for the company). The leasing fee may be expected to be proportional to the market price for the storage devices. The other relevant OPEX is the personnel, which is strongly reduced with respect to the Buy case, but has still to be present to manage the interactions of the company's computing structure with the cloud.

The set of expenses under the two alternatives Buy (B) or Lease (L) are summed up in Table 2.

## 4. Risk evaluation

The company moving to cloud storage faces the transition from the ownership of storage infrastructure to the usage of a leased infrastructure. It needs a criterion to decide whether the move looks worthwhile and a risk assessment tool to evaluate the consequences of a wrong decision. In this paper, following the approach of [17], we adopt the differential Net Present Value (NPV) to compare the two alternatives, and the Value-at-Risk (VaR)

| Item | Category | Buy-or-Lease |
|---|---|---|
| Disks | CAPEX | B |
| Controller | CAPEX | B |
| Power | OPEX | B |
| Personnel | OPEX | B/L |
| Salvage | Revenues | B |
| Lease | OPEX | L |

Table 2: Cash flow components

to assess the risk. In this section, we provide some results for typical situations.

In the case we are examining, we have considered just the expenses (with the exception of the salvage value recovered by selling the disks on the secondary market), so that both alternatives have a negative NPV. The revenues deriving from the stored data are irrelevant, since their contribution to the NPV is equal under both alternatives (the operations and the commercial exploitation we can achieve are expected to be the same irrespective of where the data are stored). When comparing two alternatives, we can either select that exhibiting the largest NPV (or, better, the least expensive), or simply compute the difference between their NPVs (the differential NPV). This is the criterion we employ here: we compute the difference

$$\Delta\text{NPV} = \text{NPV}_{\text{buy}} - \text{NPV}_{\text{lease}}, \tag{1}$$

and hold the Buy decision (owning the storage infrastructure) if $\Delta\text{NPV} > 0$, and the Lease decision in the opposite case.

Actually, the $\Delta$NPV is a random variable, depending, in particular, on the price fluctuations and on the random failures affecting the disks. If we take into account the random nature of most cash flows intervening in the NPV, we end up with a probability distribution for the values of the NPV. Given the large set of variables involved in the cash flow computation, we do not obtain an analytical expression for the probability density function (pdf) of the NPV. Rather, we can obtain an empirical probability density function by resorting to MonteCarlo simulation, where we generate sample of all the random variables considered in Section 3. A typical outcome is shown in Figure 1, which has been obtained for a large company, whose storage needs

grow from 10 terabytes to 50 terabytes over a time horizon of 5 years. In this,

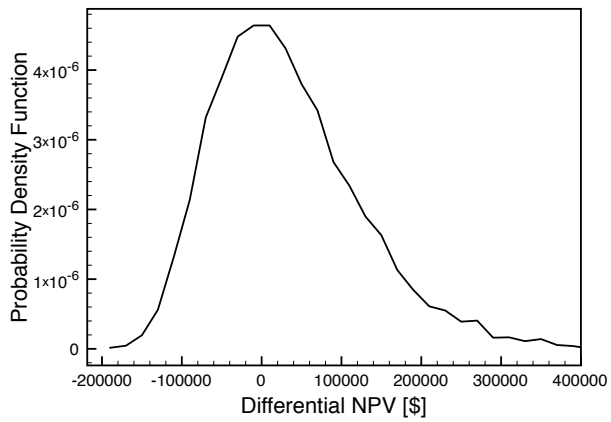


Figure 1: Empirical Probability Density Function of Differential NPV

as in a number of other cases we do not report here for brevity, we obtain a unimodal distribution, lightly skewed to the right.

Though the empirical pdf gives us a picture of the relative convenience of moving to the cloud, we need to define a decision criterion based on some statistics of the NPV. A natural choice is to resort to a measure of the central tendency of the NPV. Of the two well known central tendency measures, the mean and the median, the mean $\Delta$NPV (denoted as $\overline{\Delta\text{NPV}}$) has been shown in [17] to lead to a substantial reduction of risk. Hereafter, we consider therefore that quantity as the indicator of the convenience of the Buy decision. Namely, the decision criterion is formulated as follows:

$$\begin{aligned} \overline{\Delta\text{NPV}} > 0 &\Longrightarrow \text{BUY} \\ \overline{\Delta\text{NPV}} < 0 &\Longrightarrow \text{LEASE} \end{aligned} \tag{2}$$

Whatever the choice taken according to the principle (2), the decision may prove wrong. In fact, there is always a non-zero probability that the actual differential NPV has a sign different from its mean. In that case, we take the decision leading to the most, rather than the least, expensive solution. We can evauate that risk by computing the Value-at-Risk. Since we can err in either of two ways (by taking the wrong Buy or Lease decision) we have to

define the VaR for a confidence level $\alpha$ in the following two alternative ways:

$$\mathrm{VaR}_\alpha = \begin{cases} l : \mathbb{P}[\Delta\mathrm{NPV} < -l] \leq 1-\alpha & \text{if } \overline{\Delta\mathrm{NPV}} \geq 0 \\ l : \mathbb{P}[\Delta\mathrm{NPV} > l] \leq 1-\alpha & \text{if } \overline{\Delta\mathrm{NPV}} < 0. \end{cases} \tag{3}$$

Two sample curves, pertaining to the confidence levels 95% and 99%, are reported in Figure 2, where the leasing surcharge represents the factor by which the leasing fee is increased with the current purchase price for disks. Again, we consider here a large company, but over a time horizon of 10 years. Roughly, the left hand portion of all the curves shown in those figures refers to the case where we opt for the Lease decision; the right hand portion refers to the opposite case. In fact, the left region corresponds to light leasing fees, which lead to the Lease decision. The skewed shape of the pdf of the NPV, shown in Figure 1, gave us a hint that the Buy decision is less risky, since the right tail, with positive $\Delta$NPV favouring the Buy decision, is much longer than the left one. This intuition is confirmed here, since the left region (corresponding to the Lease decision) exhibits a much larger VaR: the risk incurred when taking the Lease decision may be much larger than when deciding to buy. When we proceed from safe regions (i.e., regions of leasing surcharge values where the risk is low) towards the threshold leasing surcharge, we see that the risk increases. The risk reaches its peak when the Differential NPV is very close to zero, and then suddenly falls to the lower values corresponding to the Buy decision on the right region.

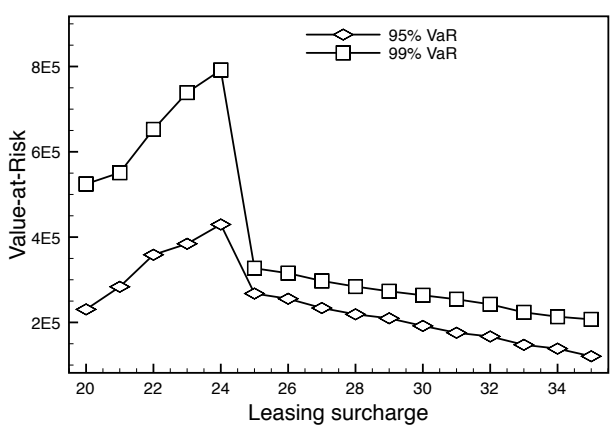


Figure 2: Value-at-Risk for a large company

## 5. Conclusions

Cloud computing is going to represent the fifth utility, which brings along the commoditization of computing resources. This is particularly true for the data storage service, which can be easily transferred to the cloud and employed on a pay-per-use basis. We have provided a methodological framework to compare the alternatives, owning the data storage facilities vs. leasing them from the cloud. The decision to move to the cloud is not void of significant economical risks. In particular, by computing the Net Present Value of the two alternatives, we show that we can evaluate the risks connected with either decision, as a function of the expected leasing fee.

## References


[1] VV.AA., Open Cloud Manifesto, Available at `http://opencloudmanifesto.org/opencloudmanifesto1.htm`, ????

[2] T. E. E. Staff, On the periphery: The clouds communications with its clients will become ever more intelligent and interactive, The Economist (23 October 2008). Available at `www.economist.com/specialreports/displaystory.cfm?story_id=12411896`.

[3] VV.AA., Cloud Computing Use Cases, Technical Report, Cloud Computing Use Case Discussion Group, 2 July 2010.

[4] J. Brockman, Counting On The Cloud To Drive Computing's Future, 27 March 2009. Available at `http://www.npr.org/templates/story/story.php?storyId=102453091`.

[5] D. C. Wyld, The Utility of Cloud Computing as a New Pricing and Consumption - Model for Information Technology, International Journal of Database Management Systems 1 (November 2009) 1–20.

[6] M. Hilbert, P. López, The Worlds Technological Capacity to Store, Communicate, and Compute Information, Science 332 (April 2011) 60–65.

[7] A. V. Makarenko, Phenomenological model for growth of volumes digital data, CoRR abs/1102.5500 (2011). ArXiv preprint series.

[8] E. Allen, C. M. Morris, Library of congress and duracloud launch pilot program using cloud technologies to test perpetual access to digital content., Library of Congress, News Release, July 14 2009. Available at www.loc.gov/today/pr/2009/09-140.html.

[9] D. Robinson, Amazon Web Services Made Simple: Learn how Amazon EC2, S3, SimpleDB and SQS Web Services enables you to reach business goals faster, Emereo Pty Ltd, London, UK, 2008.

[10] W. Hu, T. Yang, J. N. Matthews, The good, the bad and the ugly of consumer cloud storage, SIGOPS Oper. Syst. Rev. 44 (2010) 110–115.

[11] A. Efrati, Google near launch of cloud storage service, Online on Wall Street Journal website, 9 February 2012. Available at http://online.wsj.com/article/SB10001424052970204369404577211961645711988.html.

[12] B. Tansey, Fear factor decreases for cloud computing services, Online on CNBC website, 17 February 2012. Available at www.cnbc.com/id/46431304.

[13] E. Pinheiro, W.-D. Weber, L. A. Barroso, Failure trends in a large disk drive population, in: 5th USENIX Conference on File and Storage Technologies, FAST 2007, pp. 17–28.

[14] B. Schroeder, G. A. Gibson, Disk failures in the real world: What does an mttf of 1, 000, 000 hours mean to you?, in: 5th USENIX Conference on File and Storage Technologies, FAST 2007, pp. 1–16.

[15] E. Walker, W. Brisken, J. Romney, To lease or not to lease from storage clouds, IEEE Computer 43 (2010) 44–50.

[16] L. Mastroeni, M. Naldi, Storage Buy-or-Lease Decisions in Cloud Computing under Price Uncertainty, in: 7th EuroNF Conference on Next Generation Internet, Kaiserslautern.

[17] L. Mastroeni, M. Naldi, Long-range evaluation of risk in the migration to cloud storage, in: 13th IEEE Conference on Commerce and Enterprise Computing, CEC 2011, pp. 260–266.