CSPE: Cloud Storage Provisioning Decided by Rate of Return and Workload Characteristics

Jianzong Wang*, Rui Hua, Changsheng Xie, Jiguang Wan, Yanjun Chen, Peng Wang, Weijiao Gong

School of Computer Science and Technology, Huazhong University of Science and Technology, China
Wuhan National Laboratory for Optoelectronics,Wuhan,China
* Corresponding and Student Author: jzwang@smail.hust.edu.cn

I. Motivation

As recent report [1] claims, the capacity of digital content on the Internet has amounted to 500 billion GB. What is more, this number is estimated to be double in next year. The emerging of cloud computing offers a rather feasible solution to the problem of information explosion. Thus, for those IT enterprises with high demand of storage, a big concern is to determine whether it is cost effective to lease storage service from the servers over clouds. In this paper, we introduce a Cloud Storage Provisioning Engine (CSPE) to help users rationally evaluate the benefit of purchasing new disk drives and leasing from remote servers offered by Infrastructure as a service (IaaS) providers.

The contributions of CSPE are as follows:

- CSPE evaluates the future storage demand by tracing previous data increment tendency, which is completely customer-made for growth-oriented enterprises.
- CSPE uses the widely-used Internal Rate of Return (IRR) in economics to solve ”purchase or not” problem with regard to storage provisioning .
- In regular services stage, we optimize our engine from workload utilization perspective to further complete workload provisioning for the purpose of cost saving.

II. Background

Recently, Facebook claims it owns over 750 millions users, and times of shared content (including videos, photos, logs, and etc.) reach 4 billions per day. In 2020, Estimates for global data usage shows an incredibly huge number of 35ZB (1ZB equals 1 billion TB). Under the tendency, we can predict that thousands of medium-size enterprises with tens or hundreds servers, which make up almost 50 percent of all data centers installed in the US [2] will be in a critical dilemma (i.e. to cloud or not problem) in the foreseeable future.

A few researches have been done to help evaluate the cost saving over clouds through scientific experiments. Study [3] discusses the viability of Amazon’s Simple Storage Service (S3) to save cost for IT enterprises. However, we observe that there is a lack of generalization for those applications not used in the study. Another example is [4], it shows nearly 20% cost savings by bringing forward ”right-virtualizing” method, but the study is aimed to solve ”to virtualize or not” problem and only considers the licensing fees of virtualization technologies for IaaS providers. Hence, in this project, we attempt to build an integrated engine with generalization and suitability to Software as a service (SaaS) providers.

III. Overview of Modeling Approach

We drive our CSPE working by following steps. Firstly, we bring in Internal Rate of Return (IRR) models - widely used in capital budgeting to measure and compare the profitability of investments in order to help decide whether companies should purchase new disk drives or lease remote cloud computing service. However, merely a solution to ”to purchase or not” problem is not enough for practical applications. The challenge to estimate the unpredictable peak resource utilization (e.g. CPU, Disk, Network) of each workload is always a priority and risk in the local datacenter, because a bursty (i.e. high peak-average utilization ratio) workload actually causes a less dense workload placement possible on the server and hence much lower average server utilization, which renders in deployment of more resources and higher cost. Thus, in the second phase, we come up with a module called ”Burstiness Filter” (shown in figure 1) to identify those bursty workloads and then migrate them to the cloud storage service providers for the benefits of cost savings and risks avoiding.

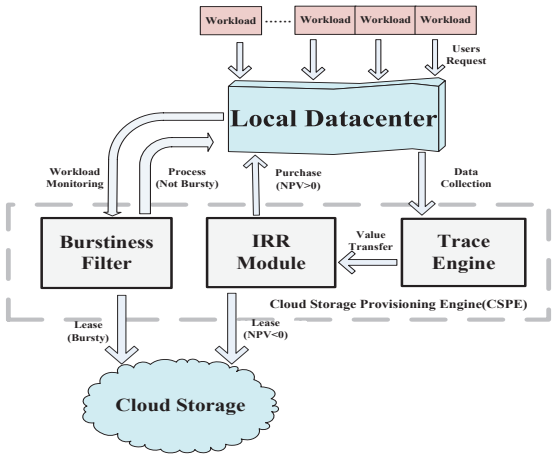

Fig. 1. Framework of CSPE

IV. Model Description

In our CSPE framework, the architecture of which is shown in figure 1, we apply a **Trace Engine** to analyze previous records in the local datacenter to predict data extension tendency and estimate future demand in storage. Then predicted

results will be input into **IRR Module**, where we use our IRR approach to draw graphic images of calculated ΔIRR values for a storage life expectancy from 0 to 10 years (shown in figure 2). Based on the results, the enterprises can make a quick purchase or lease decision in a more straight way.

The internal rate of return makes the net present value (NPV) of all cash flows (both positive and negative) from a particular investment equal to zero. The simplified standard NPV equation is shown in Eq (1), and the parameters are reflected in table I.

$$NPV = \sum_{t \in n} \frac{C_t}{(1+r)^t} \tag{1}$$

TABLE I
NOTATIONS OF MODEL

Notations	Description
C_t	Initial disk drive investment
C_0	Initial disk drive investment
r	Discount rate
t	Time period (years)
n	The life cycle of this project (years)

So we can infer that IRR of purchasing new disk drives are given in the left one of Eq (2), and similarly the IRR of leasing over the clouds in the right of Eq (2):

$$NPV_P = \sum_{t \in n} \frac{C_t}{(1+IRR_P)^t}, NPV_L = \sum_{t \in n} \frac{C_t}{(1+IRR_L)^t} \tag{2}$$

Using secant method Eq (3) and (1), we get IRR_P and IRR_L respectively, then we could calculate the ΔIRR Eq (4) using the equations above:

$$r_{n+1} = (1+r_n)(\frac{(1+r_{n-1}}{1+r_n})^P - 1 \tag{3}$$

$$\Delta IRR = IRR_P - IRR_L \tag{4}$$

Where

$$P = \frac{\log(NPV_{n,in}/|C_0|)}{\log(NPV_{n,in}/\log(NPV_{n-1,in})} \tag{5}$$

In light of the result of ΔIRR, companies are able to make decisions for their benefits. A positive ΔIRR value motivates companies to purchase new storage devices. Conversely, a negative ΔIRR stimulates companies to lease remote cloud storage instead of buying new ones.

After the disk drives bought as a new part of local clusters within the firms, the **Burstiness Filter**, a module in our CSPE, continues to monitor users' request (i.e. workloads), where the bursty workloads are detected and then migrated to the clouds.

V. MODEL CHARACTERIZATIONS AND EVALUATION

For small or medium size enterprises aforementioned, the blue and green lines in figure 2 shows the approximate ΔIRR trend in recent 10 years. We find that the IRR of leasing over the clouds exceeds that of purchasing new disk drives and human capital for operation and monitoring. For the large size enterprises with a datacenter of thousands servers, their trend of ΔIRR is indicated by pink and black lines in figure 2. As described in the curve, the investment of purchasing new devices becomes more profitable after 8 years, , especially for those far-sighted enterprises with servers of long expectancy. The paper [4] shows that most workloads share bursty demands, while most of the time these workloads have low CPU usage. After evaluating workloads in HP customer environment, the average peak to mean ratio for CPU usage was 52.6, while some workloads having a peak to mean ratio above 1000. We can assume that if we set the threshold 52 in our burstiness filter, and define those workloads above 52 as bursty, CSPE is able to save costs up to half of all the applications.

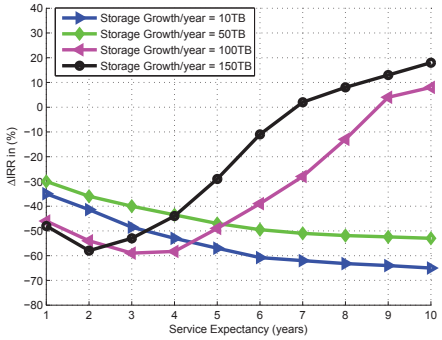

Fig. 2. Preliminary Testing of ΔIRR Tendency

VI. CONCLUSION AND CURRENT PROGRESS

Briefly, CSPE not only serves to solve purchase or lease problem for IT firms, but also automatically monitors and inspects workload condition in the local datacenter and categorizes workloads according to burstiness attribute. We then keep workloads with stable and predictable utilizations in the local datacenter and migrate other workloads with high burstiness to IaaS.

We have executed the CSPE in shared environment using various workloads, the preliminary results show advantages described in section I. We are currently perfecting our IRR approach to better solve the storage provisioning problem, and working on the models to detect the bursty workloads automatically.

REFERENCES

[1] Randal Bryant, Randy H. Katz and Edward D. Lazowska, Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science and Society. *December 2008, pp. 1-15.*

[2] US Environmental Protection Agency, Energy Star Program, Report to Congress on Server and Data Center Energy Efficiency, Public, Aug. 2007

[3] M. Palankar et al. Amazon S3 for Science Grids: A Viable Solution?. *Proc. Int'l Workshop Data-Aware Distributed Computing*, ACM Press, 2008, pp. 55-64.

[4] D. Gmach, J. Rolia, and L.Cherkasova, Resource and Virtualization Costs up in the Cloud: Models and Design Choices. *Proc. of the International Conference on Dependable Systems and Networks, (DSN20011)*,Hong Kong, China, June 27-30, 2011.