WHAT COST US CLOUD COMPUTING?
A Case Study on How to Decide for or Against IaaS based Virtual Labs

Nane Kratzke
Lübeck University of Applied Sciences
Mönkhofer Weg 239, 23562 Lübeck, Germany

Keywords: Cloud, Computing, Cost, Estimation, IaaS, Decision, Making, Model, Average, Peak, Load, Ratio, *atp*, Weinman, Case Study.

Abstract: Coud computing is characterized by ex ante cost intransparency making it difficult – from a decision point of view – to decide for or against a cloud based approach before a system enters its operational phase. This contribution develops a four step decision making model and describe its application by a performed use case analysis of the higher education domain which might be interesting for colleges, universities or other IT training facilities planning to implement cloud based training facilities. The developed four step decision making model of general IaaS* applicability can be used to decide whether a IaaS cloud based system approach is more cost efficient than a dedicated approach.

1 INTRODUCTION

Cloud computing is one of the latest developments within the business information systems domain and describes a new delivery model for IT services based on the Internet, and it typically involves the provision of dynamically scalable and often virtualized resources.

A performed a literature review showed that most of the overall cost efficiency is deduced by capacity efficiency which is intensively proclaimed as a key benefit by cloud service providers (see (Kratzke, 2011a) and (Kratzke, 2011b)). The simple fact that only the used capacity of a cloud-based service has to be paid inveigles to postulate the overall cost effectiveness of cloud-based approaches. Almost every analyzed publication repeats this more or less unreflected – even Talukader et al. (Talukader et al., 2010). This paper does not denial this postulation in general but advocates a more critical view like (Weinmann, 2011) or (Mazhelis et al., 2011) do.

According to (Truong and Dustdar, 2010) or (Kratzke, 2012) cloud computing is also characterized by an ex ante cost intransparency. This very important weakness (from an IT management and decision making point of view) is even little reflected in literature so far. To answer the question whether a cloud-based approach is more cost efficient than a dedicated approach it has to be answered the question what costs will be generated per month **before** a cloud based approach enters operation (Walker et al., 2010). This is very difficult to answer ex ante because it is influenced by a bunch of interdependent parameters. But for profound decision making exactly this question has to be answered before a system is established in a dedicated or cloud based manner.

Research Methodology. Especially the work of (Weinmann, 2011) is very interesting from this decision making point of view because it shows how to decide for or against a cloud based approach very pragmatically. This contribution is about using the work of Weinman to build up a simple decision making model for or against cloud based approaches. This decision making model has been applied in a presented case study covering the higher education domain.

We want to find out whether it is more economical for practical courses covering web technologies to provide classical dedicated educational labs or to use IaaS in order to provide our students virtual labs for their practical courses.

The analysed case study was a college lecture for computer science students covering primarily web technologies which was held at the Lübeck University of Applied Sciences in 2011. During the practical courses of this lecture students formed groups of 5 or 6 persons in order to build up a website for a scientific

*This contribution follows the IaaS, PaaS and SaaS definition of NIST (Mell and Grance, 2011).

Kratzke N..
WHAT COST US CLOUD COMPUTING? - A Case Study on How to Decide for or Against IaaS based Virtual Labs.
DOI: 10.5220/0003880804510456
In *Proceedings of the 2nd International Conference on Cloud Computing and Services Science* (CLOSER-2012), pages 451-456
ISBN: 978-989-8565-05-1

conference on robotic sailing[1] (project 1) or establish a google map based automatic sailbot tracking service (project 2) for the same conference. All groups were assigned cloud service provider accounts from Amazon Web Services. The groups were asked to use these accounts in order to fulfill their projects in a complete cloud based manner. Both projects were developed according to the timetable shown in figure 1.

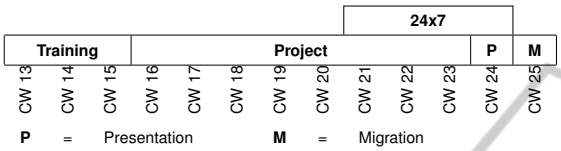

Figure 1: Project phases.

Outline. This contribution has the following outline. The decision making model is described in section 2. The case study is analysed in section 3. This contribution sums up some findings of general as well as educational interest and ends with a summary, conclusion and outlook in section 4.

2 DECISION MAKING MODEL

(Weinmann, 2011) is stressing the following interesting fact which is a crucial input for pragmatic decision making for or against cloud based system implementations especially on a IaaS level of cloud computing:

> *A pay-per-use solution obviously makes sense if the unit cost of cloud services is lower than dedicated, owned capacity. [...]*
>
> *[...]* ***a pure cloud solution also makes sense even if its unit cost is higher****, as long as the peak-to-average ratio of the demand curve is higher than the cost differential between on-demand and dedicated capacity. In other words, even if cloud services cost, say, twice as much, a pure cloud solution makes sense for those demand curves where the peak-to-average ratio is two-to-one or higher.* (Weinmann, 2011)

According to Weinman the peak-to-average ratio is the essential indicator whether a cloud based approach is economical reasonable or not. So it is not necessary to estimate costs per month of a cloud based solution exactly. It is sufficient to proof that cloud based costs are smaller then a dedicated system implementation. And this can be figured out by analysing the peak usage as well as the average usage of a system.

Let us operationalize this by defining a usage characteristic. A usage characteristic is a time sorted tuple. Each element of the tuple state how many server instances are operated within a specified atomic timeframe t. This atomic timeframe is typically set by a cloud service provider. It is the smallest possible usage reporting granularity of a cloud service provider. For example Amazon Web Services has an atomic timeframe of $t = 1h$.

$$uc := (i_{t_1}, i_{t_2}, ..., i_{t_n}) \quad t_1 < t_2 \wedge ... \wedge t_{n-1} < t_n \tag{1}$$

Let us now define the peak usage *peak* as the maximum number of parallel operated server instances for a given analytical timeframe $[T_{Start}, T_{End}]$. T must be an even multiple of t.

$$peak(T_{Start}, T_{End}, uc) := max(i_{t_k}, ..., i_{t_l}) \quad t_1 \leq t_k \wedge t_l \leq t_n \quad T_{Start} \leq t_k \wedge t_l \leq T_{End} \tag{2}$$

In the same way we can define the average usage *avg* for a given analytical timeframe $[T_{Start}, T_{End}]$:

$$avg(T_{Start}, T_{End}, uc) := \frac{\sum_{z=k}^{l} i_{t_z}}{T_{End} - T_{Start}} \quad t_k \geq T_{Start} \wedge i_l \leq T_{End} \tag{3}$$

So we can define the peak-to-avg ratio *pta* of a given usage characteristic *uc* (see equation 1) within a given analytical timeframe $[T_{Start}, T_{End}]$ in the following way:

$$pta(T_{Start}, T_{End}, uc) := \frac{peak(T_{Start}, T_{End}, uc)}{avg(T_{Start}, T_{End}, uc)} \tag{4}$$

In the following we will also use the average to peak ratio atp which is defined as the inverse of the pta:

$$atp(T_{Start}, T_{End}, uc) := \frac{1}{pta(T_{Start}, T_{End}, uc)} \tag{5}$$

According to (Weinmann, 2011) we have to compare the costs of a classical dedicated approach with the costs of a cloud based approach. On the IaaS level it is common to be billed per service usage with a granularity of the atomic timeframe t level. Which would be in case of Amazon Web Service that you are billed for a server instance per complete (or partial) hour usage. Let us name our dedicated costs per atomic timeframe d and our cloud costs per atomic timeframe c. Cloud costs c can be easily figured out because they are provided as pricing by their cloud

[1]World Robotic Sailing Conference 2011, see http://www.wrsc2011.org

service providers[2]. Dedicated costs per atomic timeframe d are a little more complex to calculate. Nevertheless for estimations we can assume, that they can be defined via their amortizations. If a dedicated instance can be procured for p value units[3] their dedicated costs per atomic timeframe can be calculated as follows[4]:

$$\begin{aligned} d_{ATF}(p) &= \frac{p}{ATF} \\ d_{3year}(p) &= \frac{p}{3 \cdot 365 \cdot 24h} \\ d_{5year}(p) &= \frac{p}{5 \cdot 365 \cdot 24h} \end{aligned} \tag{6}$$

Typical amortization timeframes are a 3 year or a 5 year hardware regeneration interval (see equation 6). So a 500 $ server would have the following dedicated costs per atomic timeframe of 1h over a amortization interval of 3 years.

$$d_{3year}(500\$) = \frac{500\$}{3 \cdot 365 \cdot 24h} \approx 0.019 \frac{\$}{h} \tag{7}$$

According to (Weinmann, 2011) the peak-to-average ratio pta should be greater than the relation between the variable costs per atomic timeframe c and the dedicated costs per atomic timeframe d_{ATF} which can be expressed in the following form:

$$\begin{aligned} & pta(T_{Start}, T_{End}, uc) > \frac{c}{d_{ATF}(p)} \\ \Leftrightarrow\ & pta(T_{Start}, T_{End}, uc) d_{ATF}(p) > c \\ \Leftrightarrow\ & c < \frac{d_{ATF}(p)}{atp(T_{Start}, T_{End}, uc)} \end{aligned} \tag{8}$$

In other words equation 4 provides a clear decision criteria to decide for or against a cloud based approach. By knowing your average to peak ratio atp, your hardware procurement costs per instance p as well as your hardware amortization timeframes ATF (which is typically 3 or 5 years) it is possible to calculate a maximum of cloud costs per atomic timeframe c_{MAX} until a cloud based approach is economical (see equation 9). Whenever a cloud service provider can realize instance pricings below c_{MAX} we decide for a cloud based approach[5] – in all other cases we should realize the system in a dedicated approach[6].

$$c_{MAX} := \frac{d_{ATF}(p)}{atp(T_{Start}, T_{End}, uc)} \tag{9}$$

The atp function generates values between $]0.0...1.0[$. Equation 9 shows us that low atp values – that means high peak to average relations (peaky usage scenarios) – will result into very high c_{MAX} values. On the other side: High atp values – that means very low peak to average ratios (constant usage scenarios) – result into decreasing c_{MAX} values. In the absolutely worst case of $atp = 1.0$ (absolutely constant usage) the dedicated costs become the maximum costs which means that the cloud service provider has to provide resources cheaper than the dedicated ones (which is very very unlikely).

3 ANALYSED CASE STUDY

Table 1 shows all costs per group within the analysed timeframe. In total the Lübeck University of Applied Sciences had to spend 847.01$ in providing a (virtually) unlimited amount of server instances to 49 students organized in 9 groups within a timeframe of 13 calendar weeks. This sounds impressive but says in fact nothing about how cost efficient this performed cloud based approach was. Could we had reached the same results with a classical dedicated approach?

Table 1: Group overview.

Group	Size	Project	Costs in $
WRSC 1	5	WRSC Website	88.39$
WRSC 2	6	WRSC Website	265.37$
WRSC 3	4	WRSC Website	88.14$
WRSC 4	6	WRSC Website	162.88$
GM 1	6	Sailbot Tracking	41.17$
GM 2	6	Sailbot Tracking	57.58$
GM 3	6	Sailbot Tracking	57.46$
GM 4	5	Sailbot Tracking	37.42$
GM 5	5	Sailbot Tracking	48.58$

3.1 Usage Analysis

As we have found out in our analyzed use case – main cost driver was server/box usage. Thats why a detailed box usage analysis has been performed and is shown in figure 2.

Figure 2(A) shows the maximum (according to equation 3) and average box usage (according to

[2]E.g. Amazon Web Services publishes these cloud costs per atomic timeframe here: http://aws.amazon.com/de/ec2/#pricing

[3]E.g. US Dollars $ or Euro €

[4]Be aware – this assumption do not account typical additional operational efforts like administration, cooling or electricity. Nevertheless we do not want to calculate exact costs we only want to know whether a cloud based approach is more economical than a dedicated one. In this case it is OK to give the dedicated side an advantage by not accounting aspects like administration, cooling, electricity, etc. although these costs are included in the variable costs on the cloud service provider side.

[5]Only from an economical point of view.

[6]Also only from an economical point of view.

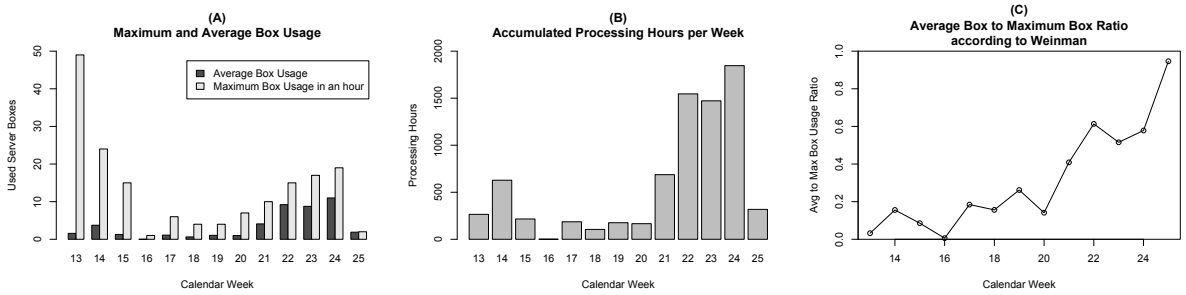

Figure 2: Analysed Box Usage.

equation 2) per calendar week measured within the analysed timeframe (calendar week 13 - 25). **Figure 2(B)** shows the total sum of all processing hours generated by all used server boxes/instances per calendar week (this is the usage characteristic shown in equation 1 aggregated to calendar weeks). **Figure 2(C)** shows the average to peak load ratio (according to equation 5) per calendar week.

Within the **initial training phase (calendar week 13 - 21)** the usage characteristic shows an extremely high maximum box usage but an astonishing low average usage. This characterizes an extreme peak load situation and results in extreme low atp ratios (see equation 5 and figure 2(C)). According to equation 8 or the definition of c_{MAX} (equation 9) this shows a very ideal cloud computing (peaky) situation. **So training phases seems to be very economical interesting cloud computing use cases**.

Within the **project phase (calendar week 13 - 15)** the usage characteristic shows dramatically reduced maximum as well as average box usages. Nevertheless the atp ratios (see equation 5 and figure 2(C)) stay in a very comfortable situation for cloud computing approaches. So we still have a peaky usage situation but on a dramatical lower level. **So also development phases seems to be very economical interesting cloud computing use cases**.

The **24x7 phase** (calendar week 21 - 24) shows raised maximum as well as average box usages (see figure 2(A). Also the atp ratios are raising within this phase – nevertheless the peaky usage remains but less distinctive. The 24x7 phase can be clearly seen in the accumulated processing hours (see figure 2(B)) which shows a clear peak in calendar weeks 22 - 24. We have stated already in section 3 that **24x7 seems to be expensive** and we can harden this conclusion by our usage analysis. This might surprise some readers.

The **migration phase** (calendar week 25) was characterized by transferring the best solutions for the website and sailbot tracking service into the operational environment. Within this phase the systems run in a steady load scenario as can be seen in figure 2(A) (almost the same average box as well as maximum usage) and in figure 2(C) (an atp ratio near 1.0). According to equation 9 this shows an extreme uncomfortable situation for cloud computing situations – **so steady loads seem to be no economical interesting cloud computing use cases.**

3.2 Economical Decision Analysis

As we have seen in sections 3 and 3.1 we can identify different phases which are more cloud compatible than others from an ecomomical point of view. Training and development phases show very low atp ratios (see figure 2(C)) and therefore indicate peaky usage characteristics of resources which advantages cloud computing realizations[7] (check equation 9). Other phases with less peaky usage characteristics (like our 24x7 or migration phase) disadvantage cloud computing realizations.

So we have identified pro and cons for a cloud based realization of our educational labs. But how to decide? Now we are going to apply our decision model presented in section 2.

Step 1: Determine the atp Ratio
First of all we have to calculate our overall average to peak load ratio atp. According to equation 3 we have to define our timeframe basis ($T_{End} - T_{Start}$). Our analysis timeframe covered the calendar weeks 13 - 25. So an intuitive timeframe for average building would be 13 weeks – but this implicates a continual usage of an educational lab over a complete year which is very uncommon. But in the university or college business educational labs are typically used one time per semester. In most cases an educational lab can be used only one time a year (per semester – that means average building over 26 weeks) or even only one time per year (every second semester – that means average building over 52 weeks which is a year). In our analyzed timeframe 7612 hours of instance usage were generated (the sum of all bars of figure 2(B)). So equation 11 (which is an application of equation 3) shows the average amount of servers which would

[7]From an economical point of view.

be necessary to provide 7612 processing hours within a 26 or 52 week timeframe.

$$avg_{26w} = \frac{7612h}{26 \cdot 7 \cdot 24h} \approx 1.74$$
$$avg_{52w} = \frac{7612h}{52 \cdot 7 \cdot 24h} \approx 0.87 \quad (10)$$

Now we can build up our average to peak ratio. Our maximum server usage within an atomic timeframe of 1 hour was 49 servers (please check figure 2(A)). So by applying equations 2, 3, 4 and 5 we got the following atp ratios for a 26 or 52 week timeframe.

$$atp_{26w} = \frac{1.74}{49} \approx 0.035$$
$$atp_{52w} = \frac{0.87}{49} \approx 0.018 \quad (11)$$

Step 2: Determine your Dedicated Costs
First of all we have to find out how much would cost us a dedicated server. At the Lübeck University of Applied Sciences our procurement office could purchase the smallest possible server version[8] for about 3055$. Equation 6 told us to calculate our dedicated costs per atomic timeframe ($1h$) in the following way for a 5 year amortization interval:

$$d_{5year}(3055\$) = \frac{3055\$}{5 \cdot 365 \cdot 24h} \approx 0.0697 \frac{\$}{h} \quad (12)$$

Step 3: Determine your Maximal Economical Cloud Costs
Equation 9 told us to calculate our c_{MAX} costs in the following way:

$$c_{MAX}^{26w} = \frac{d_{5year}(3055\$)}{atp_{26w}} = \frac{0.0697}{0.035} \frac{\$}{h} \approx 1.99 \frac{\$}{h}$$
$$c_{MAX}^{52w} = \frac{d_{5year}(3055\$)}{atp_{52w}} = \frac{0.0697}{0.018} \frac{\$}{h} \approx 3.87 \frac{\$}{h} \quad (13)$$

Do we find appropriate resources within our maximal costs? We have to figure this out in our last step 4 to make a profound decision for or against a cloud based approach.

Step 4: Determine Appropriate Cloud Resources
Now we know our maximal cloud costs and have to look if our cloud service provider can deliver appropriate and comparable resources. In our case this is Amazon Web Services, but it could be any other IaaS cloud service provider as well. We do this exemplarily for a 26 week timeframe[9]. But it works absolutely the same for all other timeframes or IaaS cloud service providers as well.

Table 2 shows all instance types of AWS and their allocated costs. Remember section 3.2 told us, that all instance types cheaper than 1.99 $\$/h$ result into cloud based solutions which are more economical than dedicated approaches.

Table 2: AWS Instance Types and Pricings, according to AWS pricing informations on 28th Oct. 2011, EU (Ireland) Region, On-Demand Instances, Linux/UNIX Operating System.

AWS Instance Type	ECU	Price/h	Economical	Comparable
Micro	< 1	0.025$	yes	-
Small (Standard)	1	0.095$	yes	o
Large (Standard)	4	0.38$	yes	o
XL (Standard)	8	0.76$	yes	+
XL (High Memory)	6.5	0.57$	yes	+
2x XL (High Memory)	13	1.14$	yes	++
4x XL (High Memory)	26	2.28$	no	++
Medium (High CPU)	5	0.19$	yes	o
XL (High CPU)	20	0.76$	yes	++

As you can see in table 2 all provided instance types (except one[10]) of AWS in the EU Region are economical in the sense of section 2 and equations 8 and 9. The most *similar* instance types listed in table 2 are marked as **'o'**. (**'-'** stands for *worser*, **'+'** for *better* and **'++'** for *much better* than a dedicated reference system[11]).

So in our analysed use case the *Medium (High CPU)* or may be even the *Small (Standard)* AWS instance types (see table 2) are the most comparable systems to our dedicated reference system (Dell PowerEdge Server R610). Both provide variable cloud costs clearly below our maximum costs of $1.99\$/h$ (see equation 13). **So in our analysed use case a cloud based approach is more economical than a dedicated approach.**

[8]Dell PowerEdge Server R610, 2.13 GHz Intel Xeon processor, 8GB memory, 140 GB hard drive (valid on 28th October 2011) with approximately 2 ECU – so the AWS instance type pendant would be something between a Standard Small (1 ECU) or Large (4 ECU) instance type. Check out detailed instance type informations of AWS here: http://aws.amazon.com/de/ec2/instance-types/

[9]Because in our special case we can use our educational lab every semester – so two times a year.

[10]4x XL (High Memory) instance type is not economical reasonable but this instance type is not comparable to our reference system because it is much more powerful.

[11]In our case the Dell PowerEdge Server R610

4 CONCLUSIONS

Summary. This contribution presented a pragmatical decision making model for or against IaaS based distributed systems inspired by (Weinmann, 2011). We applied this decision making model (see section 2) in a concrete use case of practical educational labs in the higher education domain (colleges, universities, etc., see section 3) and showed that it is very economical to use cloud based educational labs where ever it is possible. It turned out that cloud based educational labs have a more than 25 to 50 times cost advantage (see section 3.2 [step 3]) to classical dedicated approaches. So cloud computing seems to be a very promising and **economical variant of providing educational labs** for university or college practical courses which is mainly due to an inherent peaky usage characteristics of practical university or college courses (see figure 2).

Conclusions. Nevertheless the decision making model is applicable to all other domains and distributed system development approaches as well. Crucial point of the here presented approach is the first step (determine the average to peak ratio). For a profound decision this average to peak ratio has to be determined before a system enters its long-term operational phase. Problem is that this average to peak ratio is hardly predictable in analysis and development phase because it depends on a bunch of interdependent and hardly predictable parameters (Kratzke, 2011a), (Kratzke, 2012). Our proposal is to plan large-scale distributed systems generally IaaS based so that in a first usage evaluation phase the average to peak ratios can be analyzed from provided cloud service provider usage data and the presented decision making model can be applied. Dependent on the results of this evaluation the system can stay in the IaaS cloud or can be transferred to a dedicated infrastructure.

Outlook. This contribution will not deny some short comings so far. We analysed a box usage intensive use case. In our ongoing research we plan to evaluate how the here mentioned principles can be applied or adapted to data storage or data transfer intensive use cases as well. And finally – this contribution covered only the IaaS level of cloud computing so far – it is a very interesting question for our ongoing research whether the here mentioned principles can be applied to the PaaS and SaaS level as well.

ACKNOWLEDGEMENTS

Thanks to Amazon Web Services for supporting our ongoing research with several research as well as educational grants. Thanks to my students and Michael Breuker for using cloud computing in practical education. This contribution would not exist without their engagement. Let me thank Alexander Schlaefer and Uwe Krohn for organizing the World Robotic Sailing Championship 2011 in Lübeck and their confidence in our students.

REFERENCES

Kratzke, N. (2011a). Cloud-based it management impacts. In *Proceedings of the 1st International Conference on Cloud Computing and Services Science (CLOSER 2011)*, pages 145–151.

Kratzke, N. (2011b). Overcoming ex ante cost intransparency of clouds. In *Proc. of the 1st international Conference on Cloud Computing and Services Science (CLOSER2011, special session on Business Systems and Aligned IT Services - BITS 2011)*, pages 707–716.

Kratzke, N. (2012). Cloud computing costs and benefits. In Ivanov, van Sinderen, and Shishkov, editors, *Cloud Computing and Service Science*, Lecture Notes in Business Information Processing. Springer.

Mazhelis, O., Tyrväinen, P., Eeik, T. K., and Hiltunen, J. (2011). Dedicated vs.. on-demand infrastructure costs in communications-intensive applications. In *Proc. of the 1st international Conference on Cloud Computing and Services Science (CLOSER2011)*, pages 362–370.

Mell, P. and Grance, T. (2011). The nist definition of cloud computing (draft). Technical report, National Institute of Standards and Technology (U.S. Department of Commerce).

Talukader, A. K., Zimmermann, L., and Prahalad, H. (2010). Cloud economics: Principles, costs and benefits. In *Cloud Computing - Computer Communications and Networks*, volume 4, pages 343–360. Springer.

Truong, H.-L. and Dustdar, S. (2010). Cloud computing for small research groups in computational science and engineering. *Computing*.

Walker, E., Brisken, W., and Romney, J. (2010). To lease or not to lease from storage clouds. *Computer*, pages 44–50.

Weinmann, J. (2011). Mathematical proof of the inevitability of cloud computing. http://www.JoeWeinman.com/Resources/Joe_Weinman_Inevitability_Of_Cloud.pdf.