Cloud Computing and Internet of Things Fusion: Cost Issues

Lubna Luxmi Dhirani  
Department of ECE  
University of Limerick  
Ireland  
lubna.luxmi@ul.ie

Dr. Thomas Newe  
Department of ECE  
University of Limerick  
Ireland  
thomas.newe@ul.ie

Prof. Elfed Lewis  
Department of ECE  
University of Limerick  
Ireland  
elfed.lewis@ul.ie

Dr. Shahzad Nizamani  
Department of Software Eng  
Mehran University of Engg and Technology, Pakistan  
Shahzad.nizamani@muet.edu.pk

Abstract—The Internet of Things (IoT) has presented many new dimensions to information technology and data communications and has helped to develop the concepts of Smart City, Smart Travel, Smart Surveillance, Smart Health, Smart Energy, Smart Agriculture, etc. IoT offers lots of opportunity to alter conventional monitoring methods through the use of Smart IT, but it has performance limitations in terms of computational resources, limited storage and processing big data. By merging IoT and cloud computing the industry can overcome the low processing power and storage limitations of IoT, since, cloud computing is ubiquitous, comprises high computational and storage capacity ability, has unlimited virtual resources available and is capable of processing big data. However, the cloud is not a free resource and its costs need to be managed. In this paper, we discuss various cost issues which need to be smartly managed for Industries adopting the Cloud with IoT.

Keywords—Internet of Things, Cloud Computing, Pricing, Virtual Machines

I. INTRODUCTION

The vision behind the Internet of Things (IoT) was to create real-time connection and data communication with people or anything, anytime, anywhere using any network. This was to facilitate overcoming the limitations of legacy systems and ways of data communication and processing.

As stated in [1], “The IoT infrastructure allows combination of smart objects (i.e. wireless sensors, mobile robots, radio frequency identification systems, etc.), sensor network technologies, and human beings, using different but interoperable communication protocols and realizes a dynamic heterogeneous network that can be deployed in unreachable, or remote spaces (oil platforms, mines, forests, tunnels, pipes, etc.) or in case of emergencies, i.e. earthquake, fire, floods, radiations areas, etc. In this infrastructure, these different entities, or things, discover and explore each other and learn to take advantage of each others’ data by pooling of resources and significantly enhancing the scope and reliability of the resulting services”.

Cloud computing on the other hand provides a virtual, scalable, efficient and flexible system for context aware computing and online services. This provides the IoTs with a cost effective solution to deal with data storage, data management and the capability for big data processing [2][3].

This paper focuses on the different cloud instance pricing issues associated with IoT and cloud fusion.

A. Cloud Anatomy

Cloud Computing offers three different services namely Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [4]. Services are offered by giant cloud vendors, i.e. Amazon, Microsoft and Google. Globally the cloud attracts a wide range of customers from different business sectors, industries, IT and Government. In order to cater for the different customers the cloud offers different models (i.e. Public, Private, Hybrid and Community) and such services are provided at cost effective prices. The price is affected by various factors, level of convenience of service, availability, scalability, elasticity, storage capacity and security, etc [4][5]. Fig. 1 illustrates the generic cloud structure offering different applications, platforms and services.

Fig. 1. Cloud Computing Anatomy

B. Cloud Computing and IoT Trending

Cloud computing and IoT services have both been growing areas of technology in recent years. As shown in Table I, the amount of investment in cloud services has increased...
approximately threefold from 2012-2017 and is expected to grow further by 2020 [6].

<table>
<thead>
<tr>
<th>Cloud Model</th>
<th>2012</th>
<th>2017</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrastructure as a Service</td>
<td>$43B</td>
<td>$117B</td>
<td>$167B</td>
</tr>
<tr>
<td>Platform as a Service</td>
<td>$30B</td>
<td>$80B</td>
<td>$120B</td>
</tr>
<tr>
<td>Software as a Service</td>
<td>$28B</td>
<td>$65B</td>
<td>$100B</td>
</tr>
</tbody>
</table>

In 2012, the number of connected IoT devices was 8.7 billion but it is estimated to massively increase by 2020 as shown in Table II [7]. In such a situation, IoT applications will not be able to fulfill the application demand based on their current legacy systems as they will be limited in terms of storage, real-time data processing, capacity, availability and security. To overcome this limitation the cloud can be utilized.

<table>
<thead>
<tr>
<th>IoT Devices</th>
<th>2012</th>
<th>2017</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>IoT Connected devices</td>
<td>8.7</td>
<td>28.4</td>
<td>50.1</td>
</tr>
</tbody>
</table>

C. IoT Limitations and Cloud Computing

IoT applications have limitations as operators do not have standard specifications to follow and each operator installation may vary significantly from one another. It is important to have a standard architecture to encourage a uniform deployment of IoT on a cloud platform. In addition interoperability, scalability, security, availability, and big data are some other limitations that the IoT has [8] that can be addressed somewhat by the cloud.

D. IoT Applications and Cloud Computing

IoT applications have changed our everyday lives in terms of mobility, work and home life [8] as shown in Fig. 2. Radio Frequency Identification (RFID) systems [9] and Wireless Sensor Networks [10] are two key components in aiding data collection, transmission and growth of IoT applications and devices.

E. IoT and Cloud Computing Platforms

In order to promote IoT and Cloud integration, some commercial platforms are extensively working on different aspects of the services (i.e. proprietary, open source, private cloud, public cloud, application programming interface, etc [11]. IoTCloud [12], OpenIoT [13], NimBits [14], openPicus [15], Xively [16] are examples of platforms and research projects which are currently working on different areas of IoT and Cloud integration.

However, as mentioned previously, the cloud is not a free resource and its costs need to be managed for customer clarity and confidence.

II. PRICING

Cost saving has been the core reason for cloud adaption across business and IT sectors, since it offers competitive prices and schemes (i.e. pay-as-you-go or prepay). The pricing schemes and services may vary from vendor to vendor and it is important to have good knowledge about the different aspects of pricing and billing costs before IoT applications and Cloud integration can become a reality.

The Service Level Agreement (SLA) comprises of multiple Service Level Objectives (SLOs). It works as a metric to analyze the Quality of Service, i.e. performance, resource availability, cost, reliability etc., being provided by the vendor [17][18]. For Example, an SLO specified as in Table III:

<table>
<thead>
<tr>
<th>Memory (min-max)</th>
<th>16 GB Memory (min) – 32 GB (max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMs (min-max)</td>
<td>4 VMs (min) – 12 VMs (max)</td>
</tr>
<tr>
<td>Storage (min-max)</td>
<td>10 GB Storage</td>
</tr>
<tr>
<td>Ethernet (min-max)</td>
<td>10 Gigabit Ethernet</td>
</tr>
<tr>
<td>Approximate Cost</td>
<td>US$ 0.19 per hour (max)</td>
</tr>
</tbody>
</table>

Now, if the vendor fails to meet the promised service levels as, shown in Table III, based on the SLA the tenant can claim free credit as the vendor has violated the SLA.

A. Pricing Schemes

Each vendor has their own pricing mechanism (i.e. flexible, fair, dynamic, adaptive, subject availability/ automatic) [19][20] but they can be generally categorized into two types: coarse-grained and fine-grained [21].

The majority of the giant vendors implement coarse-grained billing which means the tenant has to pay for an instance-per hour even if the job is completed within 15 minutes, which means paying for 45 minutes extra for unused capacity as well. Vendors justify this by depicting the running cost of a virtual machine instances (i.e. opening, loading, operating and closing the instance) but in the broader picture fine-grained pricing would actual be of more benefit to both tenant and vendor.

In fine-grained pricing, the user can utilize resources (i.e. time, CPU, Memory) as much as is required for the tenants job and pay as per it. The vendor can make more profit by charging more unit price for shorter times to overcome the expenses, in this way the resources that were being wasted for a single tenant can be recovered and the vendor can service more tenants in the time slot that was being provided to a single tenant [21] in the coarse-grain model. The fine-grained pricing scheme has three basic attributes: resource bundle (i.e. VM Instance (Operating
System Usage, CPU, RAM, Data Storage, Bandwidth, etc.), time granularity (duration of the billing cycle) and unit price (per hour/per minute/per month) [21].

B. Virtual Machine

Each virtual machine comprises all the resources an actual physical machine has (i.e. OS, CPU, RAM, Storage, Bandwidth, etc.). Amazon categorizes instances on various aspects [22]:

- Type of OS (i.e. Linux Red Hat Enterprise Linux, Windows, etc.)
- 32 or 64 bit
- CPU
- Compute Unit
- Memory
- Instance Storage
- (S3/EBS/Solid-State Drive/Hard-Disk Drive)
- Region
- Contract type

A virtual machine image (VMI) holds all the required dependencies for running and executing a tenant application in an IaaS cloud [24]. Vendors dedicate storage nodes specifically for storing VMIs because if the disk space is not properly managed it could add extra costs to the Vendor.

The performance of a virtual machine again depends on the type of instance a tenant is using (i.e. general purpose instance, compute optimized instance, memory optimized instance, etc.).

Amazon EC3 offers two different types on VMIs (i.e. Simple Storage Service (S3), Elastic Block Store (EBS) [22]. The mechanism, approach, performance and VM boot up time varies on both of Amazon’s storage offerings [24].

The S3 backed VMIs are always stored in compressed format to save the disk space, they enhance the network performance as well as they need to be transferred to the compute node before starting a virtual machine and decompressed [24].

The EBS backed VMIs are stored on specific compute nodes and are read only when it is required by the virtual machine. This increases the number of I/O operations between the VM and VMIs and despite having a quicker startup time the network performance costs more [24].

A virtual machine boots almost the same way a machine boots after installing a new operating system, the only difference is a VM reads information from a VMI [24]. It has been observed that the VM boot up time may vary based on different versions and types of operating systems [25][26]. This boot time can add further costs to the tenant.

1) Impact of VMs on fine-grained and coarse grained pricing methods

As mentioned before, a VM may contribute a significant amount of extra cost and time at the vendor end, these costs are accrued as a result of VM configuration, startup and maintenance, etc. As per the coarse-grained model a tenant has to pay per hour irrespective of the amount of resources utilized. The fine-grained method is more tenant friendly where the tenant only needs to pay per resources used per minute [21].

As an example, both coarse-grained and fine-grained pricing have been depicted in Fig. 3 and 4 in the context of a virtual machine. From these it can be said that the vendors might save on resources with fine-grained schemes but the VM maintenance overhead will drastically increase.

It takes around 51 seconds for an EBS-backed up VMI to boot up irrespective of the disk-size or content [24]. An average approximate time of 60 seconds has been used considering different types of VMIs being offered by diverse vendors as shown in Fig. 3. The same amount of time (60 seconds) has been considered while shutting down a VM once the tenants job is finished, but in reality each tenant may have more than 50, 60 or 100’s of VMs, and in such situation coarse grained prices may waste a lot of resources but they can save a lot of vendor overhead and maintenance.

![Fig. 3. Impact on resources in a coarse grained pricing scheme](image-url)

![Fig. 4. Impact on resources in a fine grained pricing scheme](image-url)

The resource bundle, time granularity and unit price [21] may vary from vendor to vendor but the situation can get complicated when resources are subcontracted to different third party vendors.
Every few months/quarterly vendors change their pricing scheme offers and it is the tenant’s job to keep an eye for suitable offers. As per law an email or pricing notice is sent out on the latest pricing schemes and changes. If the tenant fails to make the required changes into its cloud setup, by default they will be charged as per the new pricing scheme.

The tenant’s workload or processing can increase anytime but none of the vendors approve scalability and elasticity right away, as they might need to check their available resources for processing, in addition pricing also varies as per off peak/peak timings [27]. Most vendors offer different prices on the basis of: reserved instances, on-demand instances and spot instances [28].

C. Types of Instances

In this section three instance types are outlined and where they would find use in the IoT space.

1) Reserved Instances

Reserved Instances serve as provisioning of resources for future use subject to 100% guaranteed availability [29]. Reserved provisioning offers 20-30% cost benefits over on-demand plans but a tenant is required to pay upfront for a year or years to seek maximum discounts. In the case where a tenant does not wish to pay upfront, they can only attain hourly discounts which are quite smaller in comparison to full or partial payments [22]. These type of instances would be suitable for high volume data gathering IoTs where the tenant would have a good understanding of the data and processing requirements of the IoT network. A smart city IoT application for example.

2) On Demand Instances

On demand instances are charged on hourly basis, with no long term commitments and are suitable when the tenant is unsure about its future computational demands [22]. This type of instance may be suitable for a growing IoT network where a sudden growth in the number of devices and subsequent data is necessary. For example a first responder IoT network where the tenant needs to respond to changing events and storage and processing demands of such events.

3) On Spot Instances

Spot-instances are the most expensive instances; the tenant can place a bid for the unused cloud capacity, where prices for spot instances keep fluctuating. This is quite similar to bidding for products online [22]. This instance type finds little use in the IoT space.

D. Vendors Offered Instances

The types of instances (i.e. General Purpose Instances, Compute Optimized Instances, Graphics Processing Unites (GPU) Instances, Memory Optimized Instances and Storage Optimized Instances) required by the tenant depends on its business applications and usage. For example: tenants who need highest computational power buy, reserve or demand Compute Optimized (C4) instances since they are based on the Intel Xeon E5-2666 v3 processor [22], where C3 instances are designed for running compute-intensive applications based on the Intel Xeon E5-2680 v2 processor.

The above mentioned instances are further sub-categorized based on tenants application, number of virtual CPUs, processor architecture, memory size, operating system platform, instance storage, elastic block storage and network performance [22].

1. General Purpose Instances (T2, M4, M3)
   - T2 (t2.nano, t2.micro, t2.small, t2.medium, t2.large) [23]
   - M4 (m4.large, m4.xlarge, m4.2xlarge, m4.4xlarge, m4.10xlarge) [23]
   - M3 (m3.medium, m3.large, m3.xlarge, m3.2xlarge)

2. Compute Optimized Instances (C4, C3)
   - C4 (c4.large, c4.2xlarge, c4.4xlarge, c4.8xlarge) [23]
   - C3 (c3.large, c3.xlarge, c3.2xlarge, c3.4xlarge, c3.8xlarge) [23]

3. Memory Optimized Instances (R3)
   - R3 (r3.large, r3.xlarge, r3.2xlarge, r3.4xlarge, r3.8xlarge) [23]

4. GPU Instances (G2)
   - G2 (g2.2xlarge, g2.8xlarge) [23]

5. Storage Optimized Instances (I2, D2)
   - I2 (i2.xlarge, i2.2xlarge, i2.4xlarge, i2.8xlarge) [23]
   - D2 (d2.xlarge, d2.2xlarge, d2.4xlarge, d2.8xlarge) [23]

1) Burstable Performance Instances (BPI)

Amazon EC2 permits the tenants to choose between Fixed Performance Instances (e.g. M3, C3, and R3) and Burstable Performance Instances (e.g. T2) [22]. One of the major benefits of using burstable performance instances is it decreases the cost while improving the performance. Compared to other instance types, the boot time for T2 instances is comparatively shorter, in fact configuring or scaling up the instance is also quicker than the rest of the instances [30]. This speed of configuration makes T2 instances particularly suitable for IoT data logging and processing requirements.

Burstable performance instance is designed to perform on a baseline but can burst beyond the baseline occasionally when required, this performance benefit is provided because the T2 instances do not optimize the complete CPU fully and are better suited for workloads which do not fully utilize the CPU cycles. Since the instances are not fully used or when the CPU is idle the credit accumulates and that credit can be used during workload spikes in the future [22] and the most interesting part is each accumulated credit is responsible to deliver performance based of a full core/min.

III. CLOUD AND IOT SERVICE LEVEL AGREEMENTS

Cloud Service Level Agreement is a contractual document comprising the legal and technical details for the promised services to be delivered by the cloud vendor to the tenant. Once it is mutually agreed, both tenant and vendor are liable to follow it [4][31], in the case of any service violations both the tenant and cloud vendor have the right to enter the negotiation phase to pay the penalty or break the contract. It is important to understand the cloud service level agreements before moving IoT applications into the cloud environment, as IoT applications may have a different set of challenges compared to standard computing requirements [5], such as: cloud multi-tenancy, data
governance risk and control, multi-cloud issues, cloud vendor lock-in situations, etc.

A. Instance request process:

Instance requests have a certain response time which may vary from seconds to milliseconds depending on the resource availability. Amazon’s latency in responding to its cloud offerings and services is around 50 milliseconds [4][32].

All conversations for extra services requested or granted between the tenant and vendor are communicated in a Web Service Level Agreement Language and Framework (WSLA) [33]. This also performs the SLA negotiations based on different metrics (i.e. time, price, reliability, etc.) and monitoring the requirement to be met [4][34]. Various research projects for SLA Management (i.e. mOSAIC [18], ETICS [35] etc.) have been done before but none of them have been adapted by the vendors yet.

IV. TECHNOLOGICAL CHALLENGES

As discussed in Section I, with the increasing number of IoT devices and data processing in the coming years may limit the applications of IoT devices in terms of performance, efficiency and support for big data processing. IoT devices integrated with the cloud technology forming Cloud of Things (CoT) may overcome the IoT limitations but at the same time the CoT model may be susceptible to certain technological and security based issues discussed in [36] [37] such as: protocol support, energy efficiency, resource allocation, identity management, IPv6 deployment, data locality, unnecessary communication of data, QoS, etc.

V. RELATED WORK

Many IoT based projects and applications are under way with respect to cloud adoption as mentioned in subsection I.C and I.D (IoT and Cloud Computing Platforms, IoT Applications and Cloud Computing). However all of these projects have yet to focus on the cost aspect of the cloud and how the Cloud Instance costs factor can lead to serious cost disadvantages if the pricing schemes are not fully understood or properly implemented. The research presented in this paper focuses on the cost factor relating specifically to IoT and Cloud fusion. This important analysis has, as of yet, remained an area of limited published research.

IoT projects and applications may differ in terms of adopting suitable cloud virtual resources, i.e. Virtual Machines (VMs), Containers [38]. Containers have demonstrated to be an easier platform for moving applications working on IT legacy systems into the cloud and tend to have better performances in comparison to VMs [39]. However the containers open source platform may lead to security based issues and is limited to Linux based platforms. Cloud vendors have not yet publicised any comparison between VMs and Containers in terms of cost but instead have focused on performance, scalability, elasticity, adaption and customization [39]. This means that cloud VMs still have cost benefits and are not limited to open-source platforms like containers are. The work presented here focused on VMs cost analysis for this reason.

VI. CONCLUSION

The fusion of the Cloud and IoT has a definite future as cloud based services overcome the IoT limitations in terms of security, availability, storage, computation, reliability, scalability and bandwidth. Some industries have already moved their IoT applications into the cloud by using open source toolkits and frameworks which made the adaption and transition easy [11][40]. Each IoT application will vary and as such may require a different cloud type (i.e. public, private, hybrid, community) and service (i.e. IaaS, PaaS and SaaS).

IoT & Cloud fusion still comprises many technical gaps but with the increasing number of IoT devices and cloud computing adaption, IT companies are looking for and proposing solutions to bridge this gap [41][42]. Different cloud vendors offer different pricing schemes based on coarse grained or fine-grained pricing schemes as discussed above. The types of instances, i.e. general purpose, storage optimized, compute optimized, memory optimized, etc. and their availability is subject to whether the tenant reserved the instances or prefers to opt for the more expensive on-demand or spot instances. Each instance differentiates in terms of storage, computation, memory and availability and these factors will determine the best fit for the IoT application and its computation or storage needs.

System cost is always a core factor and the cloud offers both cost savings and convenience of resources for IoT devices and their applications. It will require IoT operators to have a good knowledge about cloud pricing mechanisms, billing cycles, service level agreements and cloud functionality before moving their applications into the cloud, otherwise the decision may not lead to cost-benefits and the tenant may find themselves locked-in to an expensive SLA with a vendor.

REFERENCES


[12] sites.google.com/site/opensourceiotcloud/ [2 April 2017]


[16] sively.com [2 April 2017]


[22] aws.amazon.com [23 March 2017]

[23] docs.aws.amazon.com [23 March 2017]


[25] linux.die.net /man/8/chkconfig [6 April 2016]


