

Design Of Network Infrastructure Of A Cloud Data Center For Use In Health Sector

Chris Talavera

Urb. Campiña Paisajista s/n Barrio
de San Lázaro, Arequipa, Perú
Universidad Católica San Pablo
chris.talavera@ucsp.edu.pe

Julio Santisteban

Urb. Campiña Paisajista s/n Barrio
de San Lázaro, Arequipa, Perú
Universidad Católica San Pablo
jsantisteban@ucsp.edu.pe

Abstract—This article presents the design of the network infrastructure of a Data Center that meets the requirements arising from Cloud Computing, for use in the Health Sector of Arequipa city, focusing on network layer 2 and its dimensionality to meet the requirements of several health service applications. The network infrastructure dimensionality calculation is a complex challenge for an of the ground project , in this article we present a novel approach to solve this challenge.

Index Terms—Data Center, Cloud Computing, Network Design.

I. INTRODUCTION

We live in a connected world. Almost two billion people connect to the Internet and to address this need the community of information technology has created a new service delivery mechanism called "Cloud Computing". In the healthcare industry, Cloud Computing might be a paradigm shift in the use of information technology, among others: transparent management and access to electronic health records of patients, secure and reliable data storage and transmission, automation processes, streamlining workflow and consolidate assets of information technologies for providers of healthcare services; thus leading to obtain a higher quality of service.

Cloud computing especially facilitate the provision of healthcare products and services to patients in remote areas and those who have limited access to quality medical services. For that reason, communication infrastructure has to be powerful and it needs a hardy data center. Having a data center is not a new idea, but they need to make some changes to support the specific characteristics of Cloud Computing in the most optimal way. Therefore, this article shows how to design a network infrastructure using as a stege the MINSA (Ministerio de Salud) namely system of Healthcare in Arequipa, Peru.

II. THEORETICAL FRAMEWORK

The National Institute of Standards and Technology (NIST) define Cloud Computing as a technology model that enables ubiquitous, adapted and demand access network to share a set of configurable computing resources that can be quick provisioned and released with management efforts reduced or minimal interaction of the service provider [2], [8]. The main features of Cloud Computing are self-demand, comprehensive network access, resource pooling, scalability, it is based on the supply of services mainly Software as a Service (SaaS), Platform as Service (PaaS) and Infrastructure as a Service (IaaS) and there are 04 types of Cloud: Public Cloud, Private

Cloud, Community Cloud and Hybrid Cloud which combine two or more forms of clouds (private, community or public) [2], [8], [3], [12].

Cloud infrastructure consists of data centers that hosts servers and using different levels of organization or virtualization techniques it offers cloud services [24]. A logical view of a Cloud Data Center (CDC) shown in 1. This model represents the basic components or building blocks of any CDC. This view introduces encapsulation and insulation layers and impose support system modularity. There are different layers: infrastructure, databases, middleware, applications, management, monitoring and security layer, which one have specific roles and consolidated once formed the Data Center in Cloud.

III. STATE OF THE ART

There are many benefits by incorporating Cloud Computing in the healthcare industry, but to implement that, the design of a Data Center of next generation is necessary, thereby, some services providers have developed a reference architectures, for example Cisco [20], proposes an architecture which consists of three blocks: the first block is composed by network, computing and storage, this layer houses all the services provided to consumer. The second block is security layer, the key point is that security should be end-to-end architecture. The third layer is about infrastructure and services management. This architecture just shows goals to take account on the creation of Cloud Data Center but does not deliver a clear methodology.

Concerning the design of the data center network on [6] can be found the more used topologies types, as a the Fat Tree topology, consisting of two sets of elements, the core and Pods; the Bcube topology that was proposed for Modular Data Center, building to allow installation and procedures simpler physical migration compared with regular Data Centers and DCELL topology defined recursively and uses servers for packet forwarding [25].

Another important issue of Cloud Data Center are the virtualization techniques, respecto that [14] shows evidence that the latest network technologies have not been developed keeping in mind the needs of virtualization, and as a result, the network can become a bottleneck for these implementations. This article, also expose that static topologies require manual intervention to deploy and migrate virtual machines,

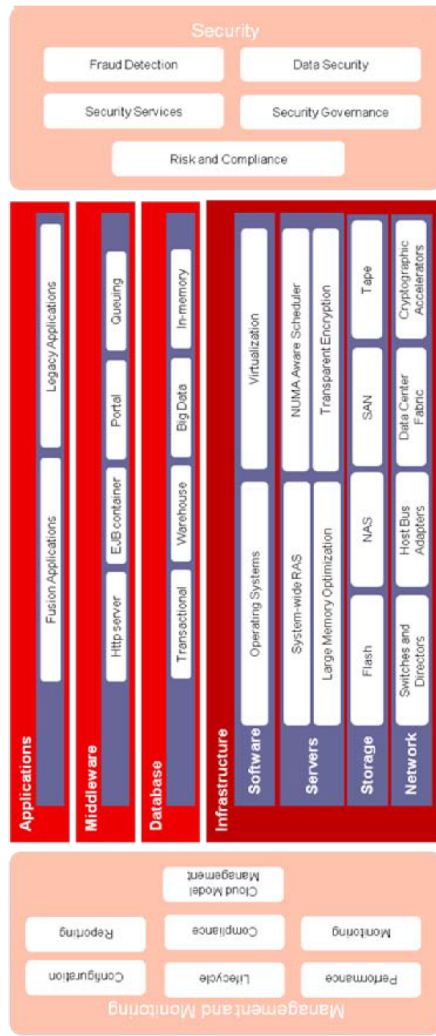


Figure 1. Reference Architecture Cloud Data Center

which adds cost and hinders the ability of the organization to respond quickly to changes in the environment, for that reason OpenFlow is presented as an open source standard designed to address these shortcomings. Based on Ethernet technology, OpenFlow separates the data path and control path by an independent controller. This introduces a new network abstraction layer, analogous to server virtualization, in consequence allows the network to act as a single structure. The benefits are simplicity, being open, scalable and fast.

IV. DESIGN OF CLOUD DATA CENTER

In this section is proposed the solution of Cloud Data Center, the first step is identified the current stage of the healthcare industry specially the main beneficiaries; in the second stage design parameters are defined. The third process is develop the analysis of network traffic; in the fourth step different network topologies are identified and compared with each other in order to choose the best performance. The final step is to perform the dimensioning of links and finally the Data Center interconnect with each of the health centers.

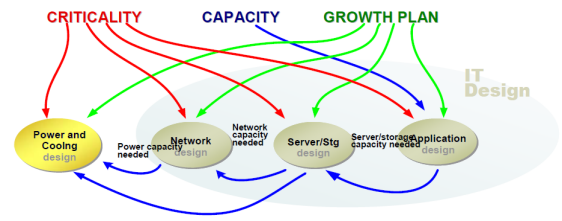


Figure 2. IT Parameters

Year	Attentions
2011	2 770 054
2012	2 920 191
2013	3 078 465
2014	3 245 318
2015	3 421 214
2016	3 606 644
2017	3 802 124
2018	4 008 199

Table I

PROJECTION OF USERS - ANUAL ATTENTIONS

A. Current Situation of MINSA

Overall, the potential beneficiaries in healthcare industry is the staff working in MINSA: health professionals, administrative staff and patients. On [Guías MINSA] it is shown that in Peru there are various categories of establishments which respond to different social and health realities and they are designed to meet demands equivalent. Thus, the level of complexity of the care services is directly related to health service development, specialization and modernization of its resources. There are 111 health facilities located in the province, which are distributed as I-1, I-2, I-3, I-4, II-1, II-2, II-E, III-1, III-2, III-E.

B. Design Parameters

For proper planning process of infrastructure cloud data center, three fundamental IT parameters has to be considered: criticality, capacity and growth or expansion plan. It is shown in the 2 that only criticality and growth plan directly affects the design of the network infrastructure [16].

According to [5]] to choose this parameters there are several methods for example the TIER UPTIME which gives 4 levels of availability. A second method is tied to TIA 942 [1], [17] where the division of 4 levels or Tiers is standard: TIER I for basic infrastructure without redundancies, TIER II for Infrastructure components with redundant capacity, TIER III for redundancy N+1 and TIER IV infrastructure for fault-tolerant 2(N+1). For a healthcare cloud data center it is considered TIER IV.

The first step of design of Cloud Data Center is understanding the needs of the healthcare industry, therefore, the number of network users considering the use of statistical data was projected, as Perú has a constant growth, the average annual growth rates is 5.42% . The projection per each year is shown in I and II.

Year	Medical Staff	Administrative Staff
2011	3 173	851
2012	3 419	973
2013	3 684	1 113
2014	3 969	1 273
2015	4 276	1 456
2016	4 607	1 665
2017	4 964	1 904
2018	5 348	2 178

Table II
PROJECTED NUMBER OF WORKERS IN MINSA

Implicated	Fc (%)	ρ_{hora} (Erl.)	$\rho_{horapico}$ (Erl.)
Patient	80	122	140
	65	99	114
	40	6	7
Medical Staff	80	3	4
	65	2	2
	40	2	2
Administrative Staff	100	2	2

Table III
TRAFFIC INTENSITY BY TYPE OF HEALTHCARE USERS

C. Traffic Analysis

In this section the calculation of minimum, maximum and margin for the network throughput is found, in this way the goal is comply with the parameters of future growth. Using the concurrency factor (CF) which determines the ratio between total simultaneous users and users who use the network in the day, not having accurate statistics, an analysis is made for each involved in the healthcare industry.

Criteria or considerations for the calculation of traffic:

- 1) The peak time is 10 to 20% of daily traffic, so they will take 15% to make the calculations.
- 2) FC for each involved in the health industry was found, IE, patients, health and administrative staff.

The III summarizes the data obtained and thereby the requirement for the network is known.

Respecto to growth parameters defined, the IV shows the minimum, maximum and margin capacity.

D. Definition of Network Servers

Although the determination of servers of CDC is essential, there is no standard way to find the exact number of these devices [18], this reality is that any service provider that offers Cloud Computing had to start him infrastructure from zero, is actually found in the process of adapting their traditional data center to the new trend.

In this paper, the number of servers was calculated based on the modeling of the process to entry to themselves, using the queuing theory and prefixing a parameter of quality of service as: time of service or the CPU usage threshold. This idea borns

Parameter	Quantity (Erl)	Total (Erl)
Max. Throughput	146	161
Min. Throughput	127	142
Margin		15
Total		288

Table IV
GROWTH PLAN PARAMETERS

because Cloud Computing, as part of scalability, automatic resources allocation is performed using mechanisms autoscaling where alarms are configured appropriately to respond in the best way to a requirement, precisely the most used algorithms keep on queuing theory [11], [13], [21], [10], [18].

The queuing model used is denoted as $M/M/c/c$. Where M is a system of arrivals that occurs according to Poison process ratio of λ , where the arrival times are exponentially distributed with mean μ , c represents the number of servers and the maximum number of customers system's allowed (when $c+1$ requests coming into the system, the service is denied for the latter).

In addition, as a parameter of quality of service has decided to consider the total response time of the service(s) for a Cloud Data Center should not be over 450ms [18]. It may have been chosen as a quality parameter the CPU utilization of the server, which according to [13] should be at least 85%.

Thus, following Little relations and queuing theory, the following relationship was obtained (1) [18], [7], [19].

$$s = \frac{\mu}{1 + \frac{\lambda * \mu}{n}} \quad (1)$$

where:

s : Average service time

λ : Arrival rate

μ : Service time

c : Number of servers

n : Number of cores server

The first required parameter is the arrival rate to the system (λ), Number that can be taken as the maximum network throughput, 141 Erlangs plus margin of 15 Erlangs, ie 161 Erlangs. Regarding the length of service, is necessary to know how long a server take to process a request, no doubt this parameter is random, but an approach can be arrive with some tests such as those in [18], where different instances are analyzed in Amazon, so the average value is 178ms (μ). It has also considered a single core server ($n = 1$) And the average service time is 450ms. The number of required servers is 48, which should form clusters or it have to be virtualized.

In order to test these results, real cases have been investigated, in this way it is possible to have a more realistic idea of how many servers would be required in an environment of Healthcare. So, first a survey was conducted to people involved of Information Technology area with goal to know the used way that they use to perform sizing of servers and the most used applications; the survey and its results can be seen in [15]. On the other hand, statistics of the use of networks, servers and applications that run on public institutions was obtained, as well as the number of concurrent users that it houses. The important thing is to know how many cores of 1GHz each institution uses on their network and how many concurrent users are allowed. In theV, the information is shown.

The data presented show a ratio factor equal to 0.049, through it the necessary number of cores is calculated to allow 421 users, which is the number of jobs per peak hour at this stage. The total number of servers to use is 27, according to real statistics, this result shows that the formula previously

Institution	Cores 1GHz	% Use	Real Cores	Concurrent users
Provincial Municipality of Arequipa	11.90	80	9.52	200
District Municipality of Cerro Colorado	34.36	60	20.62	350
Arequipa Judiciary	59.20	70	41.44	800
Catholic San Pablo University	96.00	60	57.60	1300

Table V

STATISTICAL INFORMATION FROM SERVERS AT DIFFERENT INSTITUTIONS

	Fat - Tree	DCell	BCube
Scalability	Good	Excelent	Good
Incremental Scalability	Good	Poor	Poor
Agility	Yes	Yes	Yes
Cabling	Easy	Very Dificult	Dificult
Switch fault tolerance	Poor	Good	Good
Link fault tolerance	Good	Very Poor	Poor
Server fault tolerance	God	Very Poor	Poor
Throughput	Constant	Incremental degradation	Incremental degradation
Cost	Regular	Low	Low
Traffic balance	Yes	No	Yes

Table VI

COMPARATION BETWEEN TOPOLOGIES OF CLOUD DATA CENTER

used to calculate the number of servers, allows us to have a reliability of about 57%, which can be improved if we use another queue.

E. Network Topology

Each topology network has several advantages regarding performance, remember that there are many dimensions to characterize this parameter, such as: latency, bandwidth, cost, resistance to failure, etc.

In VI a summary of the comparison of technologies is presented, considering the above data and some others taken from [6] and its translation to the different dimensions of performance.

By the above comparison, it can be stated that the hierarchical Fat-Tree topology is the best suited for network design Cloud Data Center. Even though Fat-Tree topology is not perfect in fact its biggest problem is the emergence of bottlenecks in the root of the tree, but its advantages and differences with other network topologies make to take the decision to use this design topology of the network architecture.

Considering the traffic analysis and the procedures performed to find the number of network servers, the number of ports required for each server can be calculated, because a fat-tree topology is constructed by k -ports and can support a 100% throughput performance between $\frac{k^3}{4}$ servers, using $\frac{k}{2}$ border switches and $\frac{k}{2}$ aggregation [6], [23].

Therefore, theoretically it has:

- 1) Number of ports: 6
- 2) Number of pods: 6
- 3) Number of core switches: 6
- 4) Number of aggregation switches: 3
- 5) Number of access switches: 3

	All Uplink in active state	Extesión VLAN support	Optimizes the density of Access Switch	Scaling between link Switch
Trinagle Loop	NO	YES	NO	NO
Square Loop	YES	YES	YES	YES
U free Loop	YES	NO	YES	NO
Inverted U free loop	YES	YES	YES	YES

Table VII

COMPARING DESIGNS LAYER 2 NETWORK

The hierarchical model divides networks into modular blocks: access layer, distribution, and core, the next step in design to CDC consist of to select features for each layer in order to improve network performance. Thus, the core layer should be work on Layer 3 of the OSI model to enable the core links to achieve scalability, rapid convergence and to avoid risk of uncontrollable broadcast.

The aggregation layer is very important as this determines the stability and scalability of the entire data center network, as recommended in [22], it is best to model the aggregation layer switches with pairs of interconnected modules that provide services such as content switching, firewall, intrusion detection, and network analysis. Redundancy is important to consider, in this sense, integrated services will be defined in the "active/active" mode.

The access layer works in layer 2 and the model with square loop was chosen, because its resistance to failure is greater compared to model-free loop in addition, the comparison made in the VII shows that this topology provides benefits such as: extension of VLAN, virtual machine mobility, service module redundancy.

In this CDC network design a subnet storage must be considered, specifically a SAN (Storage Area Network) because it is a subnet with high speed storage devices. It is an important part of design therefore it allows a high throughput and lowest latency which creates a high performance across the network.

To find the size of the links in the network, calculate the current and future demand for traffic per user is needed, therefore, an estimated analysis of the various applications and services that use each involved in the industry is made health. But this analysis of traffic must not specifically take each application else must make a distinction made by type of traffic. It is important to note that various services of Cloud Computing (SaaS, PaaS or IaaS), does not introduce a new traffic pattern themselves instead, they should be seen as a new way of consuming different resources [20].

For each applications or services more important the traffic ua calculated, considering in each case the concurrency factor, VIII shows the results.

Then, the analysis establishes that the peak bandwidth required by the network user is 3.16 Gbps. To avoid saturation on network ports, these should be at least twice the calculated capacity, ie. about 6.31 Gbps. Therefore the network ports of access switches must be 10 Gbps.

To calculate the speed of the backbone links distribution Poisson formula is used to find the probability of arrivals to

Applications	Individual Capacity (Kbps)	Total TRtraffic (Mbps)
Telephony over IP	88.8	37.69
Vídeo over IP	2 530.0	1 073.86
Mail	2.58	11.97
Data bases	94.38	893.02
Share files	11.38	59.95
Internet Download	11.38	68.52
Acces to Web Page	56.89	342.50
Complementary services	669.8	691.80
Total (Mbps)	3.39	3 156.46

Table VIII
TOTAL CAPACITY FOR NETWORK SERVICES

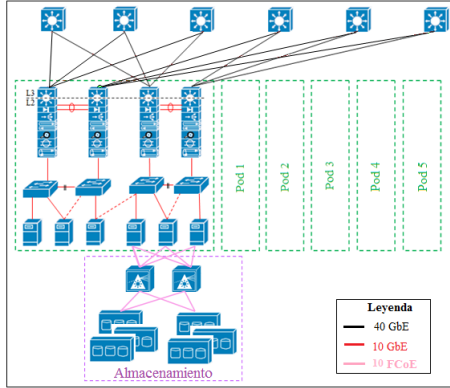


Figure 3. Network Diagram

the up-link ports, based on 2.

$$P(r) = \frac{e^{-\lambda}(\lambda)^r}{r!} \quad (2)$$

Where:

$P(r)$: Probability of arrivals to up-link ports

r : Number of arrivals to up-link port

λ : Average rate of arrivals to up-link port

To calculate, we need the number of ports of each switch, at this case 6 but adding redundancies will take as approx. 12 ports. Thus, assuming that switches of 12 ports is used, the number of simultaneous arrivals is at least 12, the average speed is 12 arrivals per unit time and probability of arrival in the up-link will be 0.11437. The result is used to calculate the speed links up-link Access Switch, by3, proposed by[9], [4].

$$Vel.ptos_{up-link} \geq (Núm.ptos) * (Vel.ptos_{half-plex}) * P(r) \quad (3)$$

Therefore, the above result is determined the speed uplink ports it must be greater than 13.7244Gbps, so the ports should be 40 Gbps or 100 Gbps for the smooth operation of the switch and the entire network is ensured. The network design is shown at 3.

F. WAN Interconnection

To find the speed of the WAN links that reach health facilities traffic demand of each one must be calculate. To achieve this, the first step is to calculate the individual requirements of each person according to the type of traffic and then make a

Traffic Type	Individual Capacity (Kbps)
Telephony over IP	22.58
Video over IP	884.74
Messaging	0.088
Data Bases	94.38
File Sharing	11.38
Internet Download	11.38
Access Web Pages	56.89

Table IX
INDIVIDUAL CAPACITY BY TYPE OF TRAFFIC

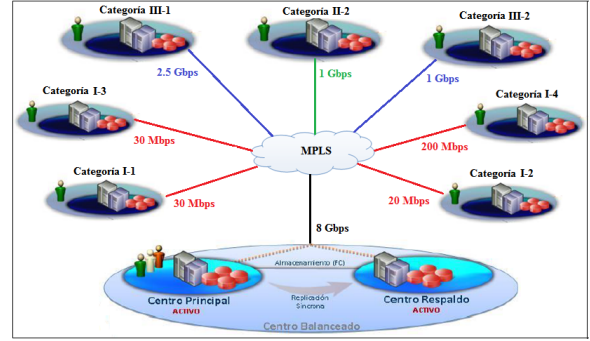


Figure 4. General view of Cloud Data Center

distribution of MINSA patients and staff by level of care and health establishment category, so individual capacity traffic type is shown at IX.

On the other hand, because the information handled in the healthcare industry is very delicate, it is important to consider a backup to the whole network, but for Cloud Computing the current traditional model of active Data Center and passive Data Center, has to be replaced by a new model of extended single data center, in which the different locations DC look as if they were a single seat and the service is actively provided from different physical locations. Therefore the network in general, will be seen as shown in 4.

V. CONCLUSIONS AND FUTURE WORK

- 1) It has been identified the technical mechanisms required for the design of network infrastructure Cloud Data Center, these are: Criticality through which we can choose according to the characteristics of applications available network; Capacity and Growth, these design factors set out to find the maximum and minimum network load and an expansion margin considering it should be a short time because it is active equipment and technology in general.
- 2) This work has completed an estimate of network traffic, based on an analysis of the reality of health facilities and in general of the MINSA (Ministry of Health), it is also thought of short growth of the number of beneficiaries. Thus, it is estimated that the network requires links 10 and 40 GbE. On the other hand, via a mathematical formula validated through statistical defined design that requires about 48 servers of 1 core.
- 3) A data center is a centralized area for storage, handling and distribution of data and information, which consists of several components such as network infrastructure,

services infrastructure, infrastructure management, monitoring, including other. Each has a specific work to be performed optimally allows the entire system to function properly. Indeed, this work has a significant contribution on this point, because although data center is not a new issue, Cloud Data Center is it and to take in account issues performance to take right decisions of design is necessary.

- 4) This article focused on network infrastructure, because this is the main part of a Cloud Data Center because it acts as the heart of communication. Thus, a thorough investigation of the features and functionality changes, new considerations and approaches that should be taken into account in order to design a Cloud Data Center was performed.
- 5) Taking into account the above considerations, the design of the network infrastructure of a Cloud Data Center was proposed. Such design has important features are listed below:

Modular design, with good scalability.

- Allows easily detect network failures and it is a network with redundancy that allows combat failures
- Access quickly to storage devices via the SAN subnet.
- Work with virtualization allowing the use of physical resources effectively..

VI. BIBLIOGRAPHY

REFERENCES

- [1] Norma ansi/tia94.
- [2] "cloud computing synopsis and recommendations", 2012.
- [3] Introduction to cloud computing architecture. White paper, Sum Microsystems, 2012.
- [4] Maroa Aguirre Patiño, Rut Ester España, Ivíçæen Solí Granda, and Alfonso Aranda Segovia. Diseño y simulación de un data center cloud computing que cumpla con la norma pci-dss. 2011.
- [5] Victor Avelar. Guidelines for specifying data center criticality/tier levels. *American Power Conversion (APC)*, pages 2007–0, 2007.
- [6] Rodrigo S Couto, Miguel Elias M Campista, and Luis Henrique MK Costa. A reliability analysis of datacenter topologies. In *Global Communications Conference (GLOBECOM), 2012 IEEE*, pages 1890–1895. IEEE, 2012.
- [7] David de la Fuente García and Raúl Pino Díez. *Teoría de líneas de espera: modelos de colas*. Universidad de Oviedo, 2001.
- [8] "Observatorio Nacional de las Telecomunicaciones y de las TI". Computación en la nube retos y oportunidades, 2012.
- [9] Vanessa Garay Olivo. Estudio y diseño de un centro de asistencia remota para una empresa de soporte de equipos oftalmológicos utilizando voz e imágenes fijas y móviles sobre ip. Master's thesis, Escuela Politécnica Nacional, 2013.
- [10] Daniel Gmach, Jerry Rolia, Ludmila Cherkasova, and Alfons Kemper. Capacity management and demand prediction for next generation data centers. In *Web Services, 2007. ICWS 2007. IEEE International Conference on*, pages 43–50. IEEE, 2007.
- [11] V Goswami, SS Patra, and GB Mund. Performance analysis of cloud with queue-dependent virtual machines. In *Recent Advances in Information Technology (RAIT), 2012 1st International Conference on*, pages 357–362. IEEE, 2012.
- [12] Yashpalsing Jadeja and Kirit Modi. Cloud computing-concepts, architecture and challenges. In *Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on*, pages 877–880. IEEE, 2012.
- [13] Hamzeh Khazaei, Jelena Mistic, and Vojislav B Mistic. Performance analysis of cloud computing centers using m/g/m/m+ r queuing systems. *Parallel and Distributed Systems, IEEE Transactions on*, 23(5):936–943, 2012.
- [14] Riso Mehra. Design and building a datacenter network: An alternative approach with openflow. Technical report, Corporación NEC, 2012.
- [15] Chris Talavera Ormeño. *Diseño de la Infraestructura de Red bajo el modelo de Computación en la Nube para su uso en el Sector Salud de Arequipa*. 2015.
- [16] Neil Rasmussen and Suzanne Niles. Data center projects: System planning. Technical report, American Power Conversion, 2007.
- [17] Reichle Y De-Massari AG (RYM). *RYM Data Center*. 2011.
- [18] Huber Flores Satish Srirama and Michele Mazzucco. Performance testing of cloud applications, interim release. REMICS Consortium 2010-2013, 2012.
- [19] JM Sidi and Asad Khamisy. Single server queueing models for communication systems, 2011.
- [20] Cisco Systems. Cisco computación en la nube - data center strategy, architecture and solutions. Technical report, Cisco Systems, 2009.
- [21] Wenhong Tian. Adaptive dimensioning of cloud data centers. In *Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on*, pages 5–10. IEEE, 2009.
- [22] Jhon Tiso. *Designing Cisco Network Service Architecture*. Cisco Press, 2012.
- [23] Amin Vahdat, Mohammad Al-Fares, Nathan Farrington, Radhika Niranján Mysore, George Porter, and Sivasankar Radhakrishnan. Scale-out networking in the data center. *IEEE micro*, 30(4):29–41, 2010.
- [24] Fabio Luciano Verdi, Christian Esteve Rothenberg, Rafael Pasquini, and M Magalhaes. Novas arquiteturas de data center para cloud computing. *XXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos - Gramado RS*, 2010.
- [25] Kaishun Wu, Jiang Xiao, and Lionel M Ni. Rethinking the architecture design of data center networks. *Frontiers of Computer Science*, 6(5):596–603, 2012.