Proceedings of the Federated Conference on Computer Science and Information Systems pp. 733–742
DOI: 10.15439/2017F111
ISSN 2300-5963 ACSIS, Vol. 11

Techno-economic framework for cloud infrastructure: a cost study of resource disaggregation

Mozhgan Mahloo, João Monteiro Soares and Amir Roozbeh
Cloud Technologies Department, Ericsson Research, Ericsson AB
Stockholm, Sweden
{mozhgan.mahloo, joao.monteiro.soares, amir.roozbeh}@ericsson.com

***Abstract*—The rapid growth of data and high-dependency of industries on using data put lots of focus on the computing facilities. Increasing the efficiency and re-architecting the underlying infrastructure of datacenters, has become a major priority. The total cost of owning and running a datacenter (DC) is affected by many parameters, which until recently were ignored as their impact on the business economy was negligible. However, that is not the case anymore, as in the new era of digital economy every penny counts. The market is too aggressive to ignore anything. Hence, the economic efficiency becomes vital for cloud infrastructure providers despite their size. This article presents a framework to assess cloud infrastructure economic efficiency, taking into account three main aspects: application profiling, hardware dimensioning and total cost of ownership (TCO). Moreover, it presents a cost study of deploying the emerging concept of disaggregated hardware architecture in DCs based on the proposed framework. The study considers all the major cost categories incurred during the DC lifetime in terms of both capital and operational expenditures. A thorough cost comparison between a DC running on a disaggregated hardware architecture with one running on a traditional server-based hardware architecture is presented. The study demonstrates the evolution of the yearly cost over DC lifetime as well as a sensitivity analysis, allowing to understand how to minimize the cost of running cloud. Results show that, lifecycle management cost is one of the main differentiators between two technologies. Moreover, it is shown that in the presence of heterogeneous workloads, having a DC based on a fully disaggregated hardware brings high savings (more than 40% depending on the applications) compared to the traditional hardware architectures independent of the hardware set-up.**

***Index Terms*—datacenter cost, disaggregated hardware, total cost of ownership, hardware pools, reconfigurable hardware**

I. Introduction

The rapid digitalization of industries, combined with the rise of the Internet of Things (IoT) concept, are just a few factors forcing a vast increase of Information Technology (IT) capacity, such as compute, storage and networking in datacenters [1]. In consequence, global spending on datacenter (DC) systems and cloud computing is growing [1]. However, as current ratio between IT capacity and its related cost is already high, it will not be easy to deliver the required capacity in the future using current datacenter technologies and strategies, even by increased spending. Hence, decreasing the total cost of owning and running datacenters is of high interest.

These facts have led the IT community to search for ways to scale DC infrastructures beyond the cost and capacity limitations of today's architecture [2]. This requires technical advancements to be brought to life along with a perception of their financial impact. Any new technology should be financially viable to survive in the competitive markets despite its technical excellence. Vendors must assure business profitability before investing on new technologies. Although we have been witnessing several significant IT's technological advancements in cloud area, there is very little insight on the financial impact of those advancements. In that sense, we argue that a methodology and framework for assessing the cloud infrastructure economic efficiency should be available.

From a technical perspective, we are seeing the DC architecture being fundamentally rethought to become more modular, flexible and smart. In the center of this architecture change is the concept of hardware resource disaggregation [4], whose flexibility not only brings new functional opportunities, but it is also seen as a promising step towards reduced total cost of ownership (TCO). There have been a set of early studies looking to application performance under this new architecture [4][5]. Although these studies showed that migrating certain applications can result in a decrease in performance, they have also pointed out that redesigning of applications with this architecture in mind could boost back application performance. While performance aspects will define/limit to some degree the exact shape of a disaggregated system, the cost will as well. However, there is limited work exploring the cost dimension of this paradigm.

To cover the gap in the current studies, in this paper, we present a methodology and a generic framework to assess cloud infrastructure economic efficiency, considering three main aspects: application profiling, hardware dimensioning and TCO. Moreover, using a simulation tool that implements the proposed methodology and framework, we present a comprehensive cost study of deploying the emerging disaggregated hardware architecture in DCs in comparison with the counterpart alternative of having traditional servers. We analyze the TCO of a DC, considering all the major costs categories incurred during the DC lifetime, both in terms of capital expenditures (CAPEX) and operation expenditures (OPEX). The results of our cost study show considerable cost benefits of deployments of disaggregated hardware architecture compared to the traditional server-based architectures (i.e., more than 40% depending on the applications type).

The remainder of this paper is organized as follows. Section II presents the related work. Section III details the methodology and framework used for the cloud infrastructure economic efficiency assessment. Further, Section IV introduces the different architecture deployment scenarios along with the case studies and assumptions considered in our study. Section

V discusses the cost study results of the different scenarios. Finally, Section VI presents final conclusions and future work.

II. Related Work

The extensive amount of studies addressing cost (in)efficiencies in DCs confirms the high importance of this aspect for DC and cloud providers. [6] evaluates the impact of data-centric workloads on the design of DC. Their observation suggests heterogeneity in the DC, in which running a job on the most cost-efficient server reduces the overall cost. In [7], the cost benefits of software-defined DCs over the traditional hardware dependent design are presented. [8] compares the TCO of a private cloud (based on the dynamic infrastructure) with public cloud alternatives and conventional server models. Their results show that the considered private cloud implementations can be up to 80% less expensive than public cloud options over a five-year period and nearly 90% less than a traditional server approach.

The concept of hardware disaggregation has been increasingly explored in the recent years. The authors of [4] were one of the first to discuss resource disaggregation on a broad perspective. Lately, further work has been done to understand required technical components to realize resource disaggregation, such as [5][10][11]. Today's most tangible realization of a disaggregated system is seen in Intel's rack scale design (RSD) [12] which is part of the foundation of the first disaggregated system available in the market [13]. However, it is important to highlight that today there is not (yet) a complete disaggregated environment, hence it is essential to have a clear and thorough understanding of the ultimate cost and business impacts of this new model to assure vendors of the return on their investment.

Although there is an extensive list of articles analyzing DC TCO considering some specific scenarios, there is lack of a more complete model to assess cost. Moreover, studies on the cost impact of a disaggregated architecture model have been limited. Cost benefits of rightsizing DCs, which is a natural outcome of disaggregated architectures, is shown in [14]. In [15], TCO is analyzed for different processor types confirming benefits of having a new scale out processors. [16] presents the cost benefits of having shared infrastructure in DCs through the comparison of a four-server chassis with shared resources with the single server case showing substantial cost savings even on a small scale.

The work presented in [17] was one of the first to provide initial insights into the cost of disaggregated systems. The authors focused on the impact of memory disaggregation on CAPEX, and ignored the OPEX. [18] goes beyond [17] by providing an overall perspective on the cost impact of full resource disaggregation. However, it does not provide thorough insights on the assumptions nor the models.

None of the aforementioned studies offers a comprehensive framework for estimating cost of ownership of running a datacenter. The available frameworks lack the possibility of comparing TCO of different technologies, architectures, hardware configurations as well as the ability to evaluate the impact of running different application types.

III. Methodology and Framework

To have a comprehensive techno-economic evaluation, a complete framework is required. DC planning consist of several stages that should be considered in a techno-economic model. This section introduces a high-level view of the main modules of the proposed framework, which contains three main modules; application profiling, hardware dimensioning and TCO calculator (see Fig. 1). Table 1 briefly describes each box of the framework shown in Fig. 1.

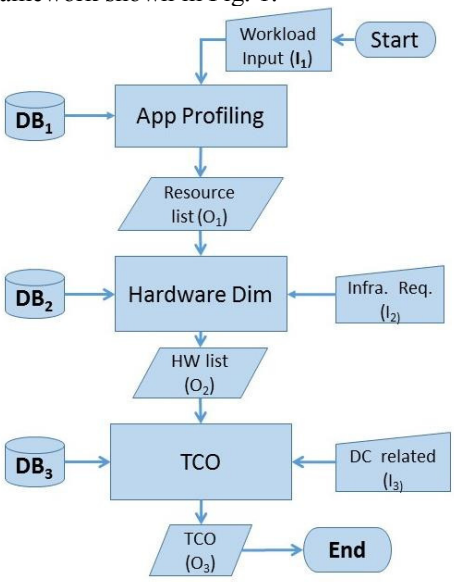

Fig. 1. Flowchart of proposed framework for cost evaluation of DC

A. Application profiling

Knowledge of the types of applications that are planned to be served by the DC and their workloads, as well as their projected yearly growth are essential for dimensioning and defining how much and what type of IT hardware needs to be purchased. For example, some applications are compute-intensive, while others might be considered as memory-intensive or network-intensive. Therefore, this module is responsible for taking the applications and their workload requirements as input and estimate the minimum amounts of IT resources needed to serve those applications using the available information in the database (i.e., DB_1 in Fig. 1). The output of this module is in terms of the unit of various components, such as the number of CPU cores or MIPS, the volume of RAM or storage, and the amount of bandwidth to/from computing nodes. However, due to the diversity of existing applications, proposing a detailed application profiling model is outside of the scope of this paper.

B. Hardware dimensioning module

Hardware dimensioning engine has access to the list of hardware that can be purchased, such as CPU types, RAM volumes, switch models and their specifications (stored in DB_2). It takes the resource requirements generated by the application profiling module as an input to produce the shopping list containing the hardware and software that need to be purchased. It also defines supporting hardware required to run the cloud related equipment such as the number of chassis, racks, power supplies and so on.

For example, if parts of the output of application profiling modules show that 800 CPU cores are needed, then hardware dimensioning module tries to find the best CPU type to cover such requirements considering the cost and other criteria. The answer could be to purchase, 100 CPUs with 8 cores each, or 50 CPUs with 16 cores each, depending on their frequencies, speed, cache size, and so on. The most cost efficient option can be defined through the interaction with the TCO engine.

Table 1. Description of the boxes related to our framework

Box	Description
Workload Input (I_1)	Requirement related to applications to be run on the datacenter, e.g. applications type and load
Database 1 (DB1)	Keeps mapping between application type, load and amount of related hardware resources
Application profiling	Module for estimating amount of hardware resources based on workload input.
Resource list (O_1)	Estimated resource list based on workload input
Infrastructure request (I_2)	Requirement related to DC infrastructure which can affect hardware dimensioning and planning, e.g. power density limit
Database 2 (DB_2)	List of available hardware resources to be purchased such as CPU types, etc.
Hardware dimensioning	This module will calculate the list of hardware resources to be purchased
Hardware list (O_2)	Output calculated by the hardware dimensioning which be used for cost calculation
DC related input (I_3)	DC related input which can impact the cost and should be given by the user to TCO calculator module, e.g. the location or size of DC
Database 3 (DB_3)	Contains hardware related information, such as cost, their power consumption, failure rate, etc.
TCO calculator	Module for estimating TCO for DC
TCO results (O_3)	Ultimate results showing the estimated cost factors and total cost in details

C. TCO module[1]

The results of hardware dimensioning will be sent to the TCO module, to estimate the TCO (i.e., including both CAPEX and OPEX aspects) of the DC for a lifetime of L years. The TCO model includes all the major costs categories incurred during the datacenter lifecycle (i.e., from the deployment phase, when a huge upfront investment is required, up to all cost aspects related to each operational process). Fig. 2 presents the generic TCO cost classification. If there is more than one set of hardware list fulfilling the application requirements, the most cost efficient option can be selected based on the results of TCO module.

1) Pricing model

The price of equipment especially when they are recently introduced to the market is normally decreasing as a result of the increase in the production volume and the market purchase, as well as, matureness of the technology. On the other hand, the expenses related to the human resources such as technician salaries are increasing each year. Therefore, price erosion should be considered while calculating the TCO.

[1] It should be noted that the formulas presented here are simplified version of the versions used and implemented in the simulation tool.

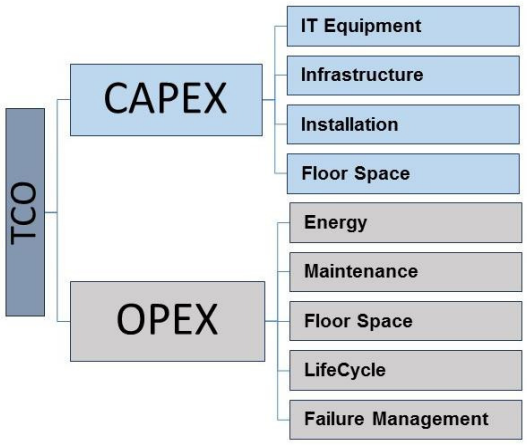

Fig. 2. Cost classification of TCO module

Price erosion during time can be calculated via learning curve used in the industry to predict the reduction/increase of the product cost [19]. However, finding the right learning curve for each product is hard. Hence, in this paper, a simple formula is considered for calculating the cost erosions (Eq. 1).

$$Pr_i = (1 + \alpha)\, Pr_{i-1} \qquad (i \geq 1) \qquad \text{Eq. 1}$$

P_i represents the price in year i of DC lifetime, where P_0 is the price of the component or starting charge in year zero. The coefficient of α denotes the cost change factor in time. This parameter has a negative value when calculating the hardware prices and a positive number for human related resources such as salaries or energy cost. In reality α might vary in time.

2) Capital expenditures (CAPEX)

Th CAPEX covers the initial investment to set-up and run the DC and can be divided into three main parts, i.e., *IT equipment, Infrastructure* and *Installation* cost (see Fig. 2).

a) IT Equipment cost

The IT equipment cost is the sum of all expenses related to purchasing the IT related hardware such as servers, switches, CPUs, chassis, rack. This number can be calculated by multiplying the estimated volume of each hardware (i.e., output of hardware dimensioning module) by their prices (Eq. 2).

$$IT_{eq} = \sum_{k=0}^{n} V_k Pr_k \qquad \text{Eq. 2}$$

Where n denotes the number of component types (e.g., CPUs, disks, servers). V_k and Pr_k represent the volume (i.e., number of units) and the price of equipment k, respectively.

b) Infrastructure (Cooling and Power) cost

A major part of any DC are the cooling and powering related facilities which their capacities need to be estimated based on the required workload in the upcoming years [20]. Some examples of such systems are; chillers, uninterrupted power systems (UPSs), heating, ventilation, air conditioning (HVAC), power distribution units (PDUs).

The infrastructure cost, which is a one-time investment, covers the expenses needed to purchase and install the cooling and powering facilities, and can be estimated by Eq. 3.

$$P\&C_{In} = PUE \times P_{IT} \times Pr \quad \text{Eq. 3}$$

Power utilization efficiency *(PUE)* measures how efficiently energy is used, considering total energy including the power used for IT components, cooling, lighting and other overheads compared to the power consumed by the IT loads. *PUE* varies between 1 and 2, where the ideal value is 1. P_{IT} represents the estimated total energy consumption of IT hardware, and *Pr* is the investment for power and cooling infrastructure for each kilowatt (KW) of consumed power.

c) Installation cost

The purchased IT equipment, needs to be installed in the appropriate location within the DC, with proper connectivity both to the network devices and power distributions units. The installation cost depends on the number of technicians required for the installation, their hourly salary rate, as well as time to install each equipment and can be calculated via Eq. 4.

$$IT_{Ins} = Tech_{sal} \sum_{k=0}^{n} V_k T_k \quad \text{Eq. 4}$$

Where $Tech_{sal}$ reflects the hourly salary of technicians who are installing the components and *n* denotes the number of component types (e.g., CPUs, memory slots, disks, servers) that need to be installed. V_k and T_k represent the volume (i.e., the number of units) and the installation time in hour for equipment type of *k*, respectively.

d) Floor space cost

The real state cost is covered in this category. In some cases, cloud infrastructure provider buys the DC's building or it has to make some initial investment to restructure the building to be suitable for its special purpose. In this case, any related investment is considered as part of the CAPEX, and should be added to the estimated cost of factors mentioned above. However, in the cases which building is rented, floor space cost can be considered zero in the initial year, and be added to OPEX as we discuss later.

3) Operational expenditures (OPEX)

OPEX refers to the expenses occurs during DC operation over a predefined time interval. The main OPEX components considered in this study are indicated in Fig. 2.

a) Energy cost

The energy cost which is one of the major challenges of DC owners can be obtained by summing up the energy cost of all the IT equipment during the project lifetime (*L*). Moreover, an estimation of the overhead power consumption (energy consumed by the lighting and cooling facilities) should be included in the calculation of energy cost using *PUE* coefficient, as shown in Eq. 5.

$$E_n = H_{year} \times PUE \sum_{i=0}^{L} Pr_i \sum_{k=0}^{n} V_{ik} P_{ik} \quad \text{Eq. 5}$$

H_{year} and Pr_i denote number of hours per year and energy price per kilowatt-hour in year *i*, respectively. Number of component types, the volume of each type and their power consumption in kilowatt in year *i*, are shown by n, V_{ik} and P_{ik}.

b) Hardware lifecycle management cost

IT equipment needs to be replaced as their performance degrades with time, or new generations of same hardware comes to the market with better performance. These expenses are considered in this cost category and reflect the investment required to procure and install new equipment during DC lifetime. The number of equipment to be replaced is calculated based on their current volume as well as their lifetime. Lifecycle management cost can be calculated via Eq. 6.

$$\mathrm{LC} = \sum_{i=0}^{L} \sum_{k=0}^{n} r_{ik} V_{ik} \left(Pr_{ik} + Tech_i^{sal} T_k\right) \quad \text{Eq. 6}$$

$$r_{ik} = \begin{cases} 1 & if\ i = x \times lc_{ik} \\ 0 & otherwise \end{cases}$$

Where *L* and *n* denote DC's lifetime and the number of component types (e.g., CPUs, disks, servers), respectively. r_{ik} is the coefficient defining if a component of the type *k* reached the end of its lifecycle in year *i* and needed to be replaced in the current year of DC lifetime (*i*). If *i* is equal to a factor of component *k* lifecycle (lc_{ik}), r_{ik} *is* equal to one, and zero otherwise. V_{ik} and Pr_{ik} represent the volume (i.e., the number of units) and the price of equipment *k*, in year *i*. $Tech_i^{sal}$ reflects the technician salary in year *i*, and T_k presents the number of hours needed to install equipment *k*.

c) Maintenance cost

A regular maintenance routine is needed to keep the DC's equipment and infrastructure up and running. This includes monitoring and testing the equipment, updating the software (including renewing licenses when needed), and the renewal of supporting components such as batteries. Maintenance cost consists of the human resource expenses as well as cost of supporting components. However, as it is hard to estimate this expenses with such a fine grain approach, we have considered a linear relation between cost of maintenance and CAPEX.

d) Floor space

As discussed, the cloud providers have two options to secure their floor space, i.e., buy/build the building or lease one. In the later case, the floor space cost is a yearly rental fee paid by DC owner to house its equipment[2]. It also includes the area required for placing the infrastructures. In this study, we first calculate the required area by estimating the total number of racks needed to serve the defined workloads, in addition to the space for placing cooling and power facilities. Then, this number is multiplied by an average rental fee per year to estimate floor space cost (see Eq. 7).

$$FS = \sum_{i=0}^{L} Pr_i \left(\alpha A_{rack} N_i^{rack} + A_{of}\right) \quad \text{Eq. 7}$$

Where Pr_i denotes the yearly rental fee per square meter of DC in year *i.* Parameter α reflects the working area for technicians or corridors in front of racks. Moreover, A_{rack} and N_i^{rack} denote area needed for a rack in a DC and number of racks in year *i*, respectively. Finally, an extra area for placing the infrastructures, control systems and offices are also considered by adding A_{of} to the equation.

[2] In this article, the DC owner is considered to be the entity owning all the IT related equipment. The facility/building where the data center is hosted is owned by a separate entity that charges a certain fee for the rental of the space.

e) Failure management cost

The cost of fixing the failures, such as replacing faulty components, or repairing them when possible is also part of the OPEX. However, estimating failure management cost is a very complex task and deserves a separate study.

IV. Deployment Scenarios

In this section, the two DC architecture scenarios considered in this paper are presented; the traditional server-based model, and the disaggregated-based model. Moreover, we present the DC workload case studies considered and detail the assumptions and parameters used in our study.

A. Server-based model (Hardware-Defined Infrastructure)

Traditional DC architectures follow server-oriented model, composed of pools of servers with fixed configuration. The fixed configuration offers limited sharing capabilities among resources, preventing them from being able to adapt to different workloads. Hence, DCs are usually planned to serve the peak demand. DC providers employ server virtualization technologies to implement resource sharing and improve utilization, while reducing their costs. However, still DCs operate at very low utilization rate [21] that means the resources paid for are not being utilized to their full capacity.

In this model, DC's lifecycle management becomes tightly bound to the lifecycle of a server. This causes problems (e.g. high cost) for providers who wish to upgrade part of their infrastructure for higher performance as the resources composing a server have different lifecycles. For example, if a DC provider wants to upgrade or increase capacity by using new memory type or CPU technology, in most cases, it ends up with replacement of the entire server, even though not all components need to be upgraded.

B. Disaggregated-based model

The hardware disaggregation principle breaks traditional physical server boundaries and considers resources as individual and modular components. Resources tend to be organized in a pool-based way, i.e. pool(s) of compute units, memory units, storage units, network interfaces, and other resources like accelerators. This brings greater modularity to a DC's lifecycle, which in turn allows the operators to optimize their resources in a more efficient way. In such environment, hosts are logically composed on top of hardware pools. Each resource pool can serve multiple hosts, and a single host can consume resources from multiple resource pools. This approach is allowing to maximize resource utilization by increasing the degree of resource sharing [5].

C. Case studies

We have considered three different type of applications, namely: systems applications and products (SAP) HANA, video on demand (VoD), and Mesos. These were chosen due to their different requirements, in terms of CPU and memory resources [22][23][24]. Each application is considered to have a different amount of load during day and night (See Table 2). The load variations of applications are adjusted in a way that total CPU and RAM requirement during day and night are nearly the same aiming to maximize the resource utilization at all the time.

We define three different scenarios based on the hardware architecture and technology used in the DC related to IT equipment: fully disaggregated architecture (DisAgg), server-based architecture with homogeneous set of hardware (Agg_1Pod), and server-based architecture considering (three) different and specialized hardware silos, one per application (Agg_3Pod). In the first two scenarios, the same type of IT equipment is dimensioned for all the applications meaning that resources can be shared among applications during different time of the day/night, while in the third scenario, each application has its own server type based on its needs.

Table 2. Application load profiles during day and night.

Application	Load unit	Day load	Night load
SAP	Server	42	30
VoD	Streams	1000000	400000
Mesos	Jobs	8000	12000

D. Input parameters and assumptions

1) Application profiling

Table 3 presents the maximum amount of required CPU cores and memory volume (GB) per scenario for each application using the following methods. SAP HANA standard specification consists of one or more very large servers, where individual server configuration is equal to four CPUs (minimum 15 cores) and 1.5 TB of Memory. So, the hardware for the required workload can be calculated using Eq. 8 and 9 [22].

$$CPU_{core} = N\_s \times 4 \times N_c \quad \text{Eq. 8}$$

$$RAM_{volume} = N\_s \times 1500 \quad \text{Eq. 9}$$

Where, N_s represents the number of running servers (42 servers during day time and 30 servers during night hours, according to Table 2), and N_c is the number of cores per CPU (15 in this example). VOD requirements are calculated based on Eq. 10 and 11 [23], where S represents the number of simultaneous streams (see Table 2).

$$CPU_{core} = S \times 0.013 \quad \text{Eq. 10}$$

$$RAM_{volume} = S \times 64 \quad \text{Eq. 11}$$

In case of Mesos, there is a 1 to 8 relation between CPU core and RAM volume, meaning that for each CPU core, 8 GB of memory are required. Mesos needs at least three servers (or VMs) as follows; one bootstrap node (2 cores and 16 GB RAM), one master node (4 cores and 32 GB RAM) and one agent node (2 cores and 16 GB RAM). However, the recommended configuration is to have three master nodes which can support many agent nodes [24]. The number of agent nodes grows with the amount of jobs planned to be executed. We consider one job per agent node at each point of time, where a job can have many tasks [24].

Table 3. CPU, memory and storage requirement per scenario

Scenario		CPU cores	RAM (GB)	Storage (GB)
DisAgg		31534	262712	300000
Agg_1Pod				
Agg_3Pod	Pod 1	2520	63000	120000
	Pod 2	13000	64000	140000
	Pod 3	24014	192112	40000

2) Hardware dimensioning

Table 4 presents the results of hardware dimensioning for each scenario in first year based on the application loads in Table 2 and application's requirements in terms of hardware resources [22][23][24]. A ten percent increase in the load per year is also considered, which means new hardware needs to be purchased to accommodate the growth each year.

In the case of fully disaggregated architecture, compute, memory, network and storage sleds are used to accommodate the components such as CPU, memory (e.g. RAM), NIC cards and storage disks (e.g. HDD, SSD). Server-based scenarios are dimensioned based on commercially available servers ([25][26][27]) which can fulfill applications requirement with the lowest amount of wasted resources. For example, as the minimum requirement for SAP server is 4 CPUs and 1.5 TB of memory, a server with 4 CPU sockets and a large amount of memory slot should be selected (in this case [25]).

Table 4. Hardware dimensioning results

Component/Item	Lifecycle (years)	Volume in number		
		DisAgg	Agg_1Pod	Agg_3Pod
Rack	7	32	50	56
Compute sled (4 socket)	5	359	0	0
Memory sled (48 DIMM)	5	86	0	0
Network sled (4 NICs)	5	359	0	0
Storage sled (20 SSD)	5	10	10	10
Server (4 socket-48 DIMM)	3	0	500	44
Server (2 socket-24 DIMM)	3	0	0	670
Server (2 socket-12 DIMM)	3	0	0	325
CPU (16 cores)	3	0	2000	174
CPU (18 cores)	3	0	0	1339
CPU (20 cores)	3	0	0	650
CPU (22 cores)	3	1436	0	0
RAM (64 GB)	4	4128	0	0
RAM (32 GB)	4	0	14300	6012
RAM (16 GB)	4	0	0	16080
SSD (960 GB)	5	200	200	200
NICs (2*25 GB ports)	4	1436	2000	2163

In the case of Agg_1Pod scenario, since workloads can share servers, all applications should be dimensioned based on highest requirements, meaning that all applications will use the model of [25]. While, in the case of Agg_3Pod, servers with 2 CPU sockets are enough for serving VoD and Mesos workloads. However, due to their different CPU core to memory proportion, different servers with 24 and 12 DIMMs are selected for them. The storage is considered the same for all scenarios because it is already separated from servers even in today's DCs. Except one of the server models from [25] which needs two rack units (RUs), the rest of the servers/sleds fit in one RU of a two RUs chassis inside rack.

The number of racks are calculated based on the number of required chassis to accommodate servers/sleds. In many cases, DCs are limited in the amount of watt per square meters they can offer, due to a variety of reasons such as safety or existence of power infrastructure facilities. This is reflected in our assumptions by filling up only half of the racks (42 RUs).

The networking equipment, e.g. top of rack switches, aggregation switches, etc. are not considered in this study. However, high capacity connectivity requirement between compute and memories in disaggregated scenario, are added to the price of compute sleds.

Components lifecycles are calculated based on the architecture types, i.e. in case of server-based scenario, replacement window of a server is equal to the lifetime of the server's component with the shortest lifecycle, while in the case of disaggregated architecture each component has its own independent replacement window. Furthermore, the coefficient of cost change factor (i.e., α in Eq. 1) is considered to be constant (3 percent) for the whole DC lifetime. The H_{year} is equal to 8760 (i.e., hours per year) and Pr_0 in Eq. 5 assumed to be 0.2 $ in this study. A predefined lifetime of 3 to 5 years are considered for various components based on component warranty (CPU, SSD [28], and NIC [29][30]), known refreshment lifecycles (four years for RAM [31] and hard disk) (see Eq. 6). Moreover, in Eq. 7, Pr_0 (i.e., yearly rental fee per square meters of DC) assumed to be 500 $ and the coefficient α (working area for technicians) is equal to 2. The failure management cost is excluded and not addressed here and we assumed that the maintenance cost per year is equal to 5 percent of the CAPEX.

Component prices used for cost calculations are selected and/or estimated based on the values in [25][26][27][32][33][34][35]. Since the hardware related to disaggregated scenarios is not commercially available, we derived the prices based on equations 12, 13 and 14.

$$P_{ComSl} = \alpha P_{ser} \quad \text{Eq. 12}$$

$$P_{MemSl} = \beta P_{ser} \quad \text{Eq. 13}$$

$$P_{NetSl} = \delta P_{ser} \quad \text{Eq. 14}$$

Where P_{ComSl}, P_{MemSl} and P_{NetSl} represent price of compute, memory and networking sleds excluding the CPU, RAM or NIC, and P_{ser} reflects price of conventional server with similar configuration (e.g. same number of CPU sockets, RAM slots, etc.). α, β and δ reflect the relation between the price of compute, memory and networking sleds, with the price of the server with the same capacity, respectively. Due to the need of high-speed networking in the disaggregated architecture, the prices are derived based on the cost of current servers plus added value of new boards and high performance networking (i.e. $\alpha+\beta+\delta>1$). Due to the high demanding communication requirements between CPUs, α has a relatively large value (1.3), while β and δ are 0.3 and 0.2, respectively. This means that the price of a disaggregated setup is 1.8 higher than a server with the same capacity.

V. Cost Analysis

We have developed a tool implementing the proposed framework based on the Java language, to be able to study and understand the cost impact of various technologies, infrastructures, and architectures while planning a DC. This tool is used to present some case studies, comparing TCO of new disaggregated hardware architectures and the conventional server-based hardware model for a DC. This section details the cost study results based on the assumptions discussed in the previous section.

A. Total cost of ownership (TCO)

Fig. 3 illustrates the accumulative TCO for the three scenarios for a DC lifetime of ten years. The disaggregated scenario offers much lower TCO compare to two other

scenarios. The cost difference grows over time due to the impact of OPEX reduction in the disaggregated scenario.

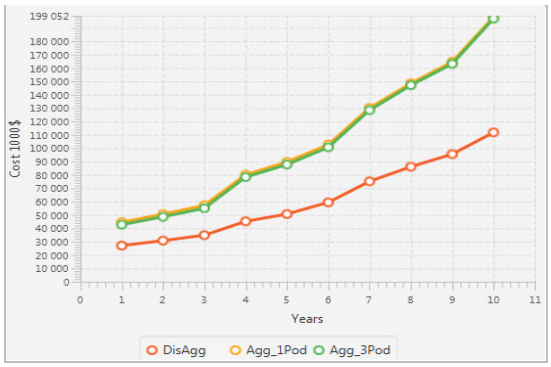

Fig. 3. Accumulative TCO per year for all scenarios

Fig. 4 shows the TCO for the three scenarios for ten years of lifetime. It can be seen that using disaggregated hardware is possible to save around 40 percent in cost after ten years. The figure also highlights the importance of OPEX, as it is twice as big as the initial investment (CAPEX).

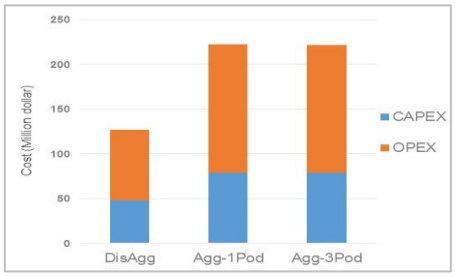

Fig. 4. Total cost for 10 years

B. *Cost breakdown and value argumentation*

Fig. 5 presents the cost breakdown to assess the impact of each cost category described in Section III on the total cost for each scenario. These numbers allow identifying the main contributors of DC's TCO, which is essential for understanding where the reductions presented above come from. It becomes evident that lifecycle management (around 35%), IT equipment (around 30%) and energy cost (around 20%) are the most expensive elements of TCO. This means that reducing any of this cost factors can lead to a considerable saving in TCO for DC owners, while focusing on improving in other categories such as having less number of technicians, has a more negligible impact on the total cost reduction.

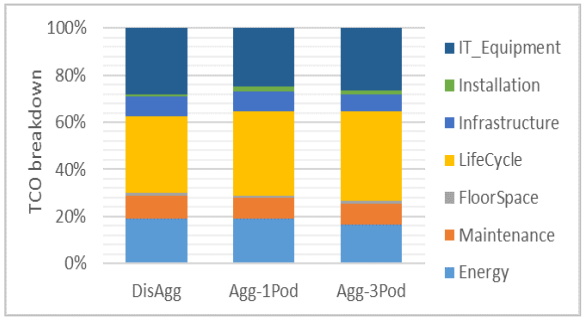

Fig. 5. Normalized TCO breakdown

Fig. 6 illustrates the expenses per cost category for three scenarios. The IT equipment cost is around 35 to 40 percent lower for disaggregated architecture. This is due to the lower amount of IT equipment purchased (32 racks compared to 50 and 56 in other two cases). A large reduction of the amount of required hardware comes from the increased hardware utilization of resource pooling (above 90% for both CPU and memory) shown in Fig. 7 and Fig. 8.

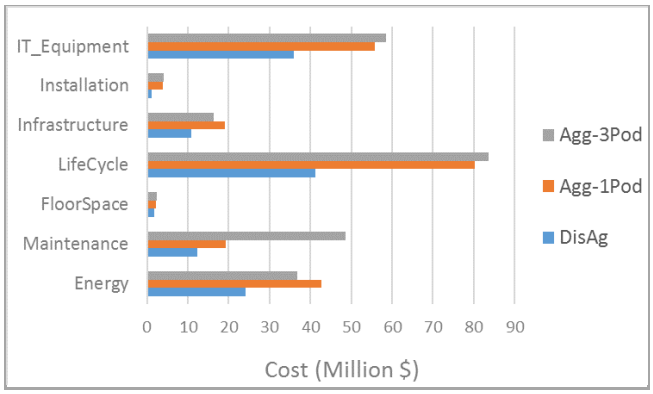

Fig. 6. Expenses per cost element for all scenarios

Though amount of assigned CPU cores is nearly the same in all scenarios, around 20 percent of CPU cores are wasted in the Agg_3Pod scenario. This is caused by the overprovisioning of resources to accommodate peaks when the sharing of resources is not possible.

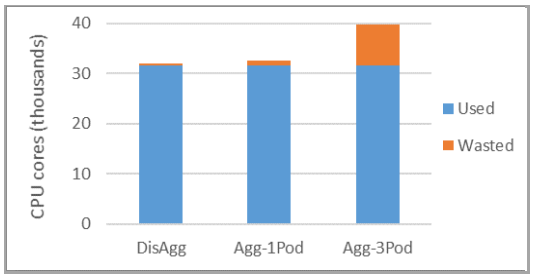

Fig. 7. Amount of allocated and wasted CPU cores during peak

Under-utilization percentage for memory increases to around 40 and 35 percent for Agg-1Pod and Agg-3Pod scenarios, respectively. This is both because servers are dimensioned for highest CPU utilization instead of memory, as well as the coarse granularity in the server's configurations and the limited boundaries for sharing the resources. This means that, when all CPUs are used in a server, residue memory is wasted and cannot be used by neighboring servers.

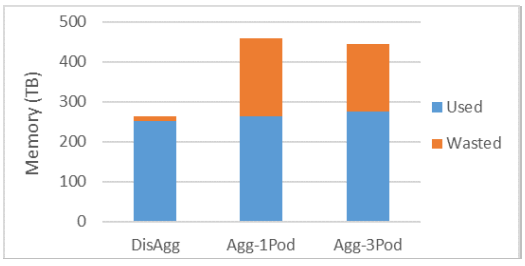

Fig. 8. Amount of allocated and wasted memory during peak

Fig. 9 shows an example of VM/server assignment to physical resources aiming to clarify increased utilization and fewer resource requirements of DisAgg scenario compared to Agg-1Pod. Two types of VM are considered, with 8 and 4 CPU cores as well as 32GB and 48GB of memory, respectively. Considering a homogeneous set of hardware, the minimum amount of resources to serve 2 VMs of each type is shown in Fig. 9. As shown, 4 RAM slot (8GB each) and 8 CPU cores are wasted in the Agg_1Pod, while in DisAgg case, resources are fully utilized and the demand could be satisfied with less hardware (25% fewer cores and 16% less memory).

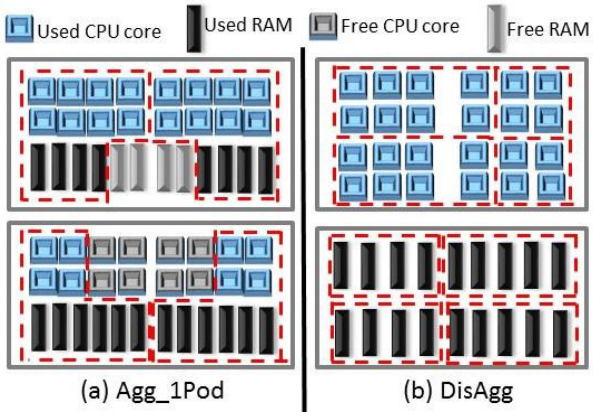

Fig. 9. VM allocation example for two scenarios

Lower amounts of components lead to the reduction of installation cost, as less time is needed to install, test and operate the resources. Increased utilization and less amount of IT resources can be translated to lower power consumption (around 35% to 40% based on Fig. 6). This leads to a lower power and cooling capacity, which reduces datacenter infrastructure cost in case of disaggregated architecture. Lower number of racks and hardware, as well as infrastructures brings around 20% and 40% of the reduction in the floor space and maintenance expenses for ten years of operation.

As shown in Fig. 5 and Fig. 6, one of the main contributors of cost reduction in case of disaggregated architecture is lifecycle management cost. Currently the lifetime of a traditional server is equal to the lifetime of the component with shortest life (i.e. CPUs with 3). This leads to more frequent and unnecessary replacements of hardware for the rest of components with longer life. However, while managing independent pools of resources, hardware refreshment process is more efficient, as each part will be replaced at the end of its own lifetime. This means that, in the two server-based scenarios, motherboard, memories, NIC cards and CPUs need to be replace every X_1 years (CPU lifetime), while in case of disaggregation, CPUs are replaced after X_1 years and memories after X_2, and NICs after X_3 years (where $X_1 <= X_2, X_3$). Therefore, 50% reduction in lifecycle cost of DisAgg scenario comes from both having lower amount of hardware to replace, and more efficient replacement process.

The same argumentation is valid for hardware failure management, meaning that in case of failure of one component (e.g. CPU), the entire server needs to be replaced and will not be operable in case of server-based scenarios, while in disaggregated case, only the failed component need to be replaced, and the rest of hardware remains operational. Note that due to complexity and tight relation of failure management cost with software and platform layers, it is not assessed as part of the TCO in this article.

Fig. 10 illustrates the TCO evolution for all the scenarios showing the cost in a given year. The amount of yearly investment varies a lot from year to year, which is mostly due to the hardware refreshment windows. There is a jump in OPEX every three years when the CPUs (entire server in Agg-1Pod and Agg_3Pod) need to be replaced.

Operational cost of datacenter is always lower in case of disaggregated scenario, though the difference varies year by year. The other two scenarios are very similar both in terms of variation trend and exact cost values.

C. Sensitivity analysis

The impact of variations in some input parameters and assumptions such as datacenter size and lifetime on the TCO fluctuation and savings are analyzed in this section. Fig. 11 shows the impact of increase in the price of disaggregated hardware, such as compute, memory and networking sled on the total cost of ownership of DC compared to the server-based scenario (see Eq. 12,13,14).

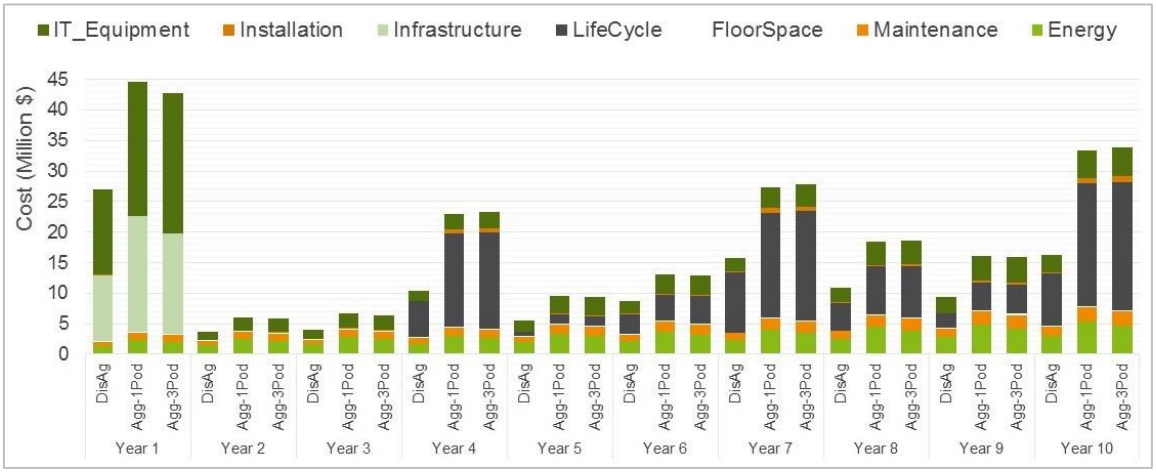

Fig. 10. TCO evolution per year for three scenarios

Red (DisAgg1) and yellow (Agg_1Pod) lines are considered as the baseline of the comparison, as they depict the results presented earlier in Fig. 3. The green and blue lines depict the yearly TCO considering a 5 and a 10 time increase in the price of compute, memory and networking sled (α,β and δ in Eq. 12, 13, and 14) in case of disaggregated hardware architecture. As it can be seen, with up to 10 times increase in the hardware cost the TCO can still be compensated with the improvement in the utilization rate of server systems.

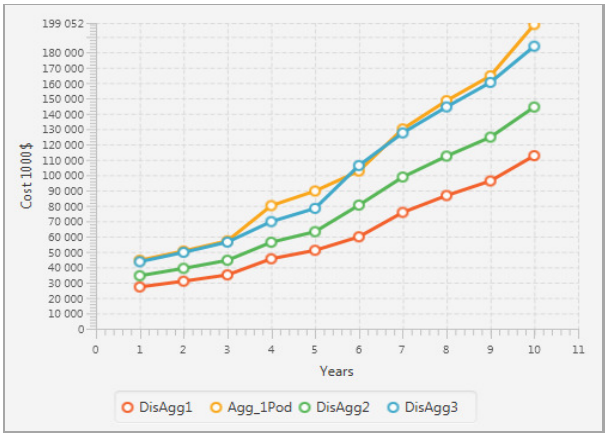

Fig. 11. Accumulative TCO for various disaggregated hardware prices

Fig. 12 shows the average yearly investment on datacenter for disaggregated and server-based scenario with homogeneous hardware (Agg_1Pod) considering variation in datacenter lifetime. As shown, the average spending per year decreases by spanning the datacenter lifetime. This is caused by the fact that the initial onetime investment (CAPEX), which represents a large part of TCO, is spanned over a larger time period.

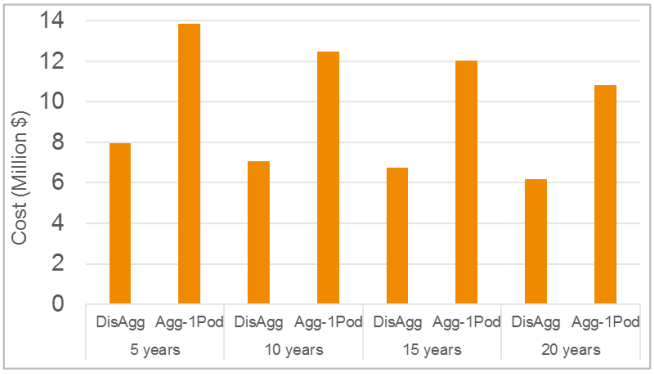

Fig. 12. Average yearly cost for different datacenter lifetime

Fig. 13 represents the cost savings in TCO by using disaggregation architecture compare to two other scenarios for a lifetime of 10 years for 3 different datacenter sizes; Small, Medium and Large. The large case is as the same size as the scenario presented in the previous section, while the workload requirement and IT equipment's are downsized by a factor of 5 and 10 for Medium and Small scenarios, respectively. It is evident that the larger the datacenter the higher the savings, due to better utilization which is the result of the increased level of resource sharing and economy of scale.

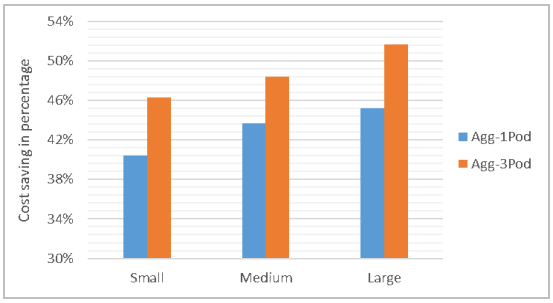

Fig. 13. Saving percentage when having disaggregation

The last part of our sensitivity analysis addresses the impact of having a single application on TCO (see Fig. 14). The TCO difference is much smaller when running only one application in a datacenter, as the hardware configuration can be optimized in both scenarios, and there is no benefit of sharing where disaggregated architecture has the most leverage. This can be translated as the benefits of multi tenancy in datacenters. The saving for single application scenarios is around 16% compared to 40% when having 3 distinct types of applications.

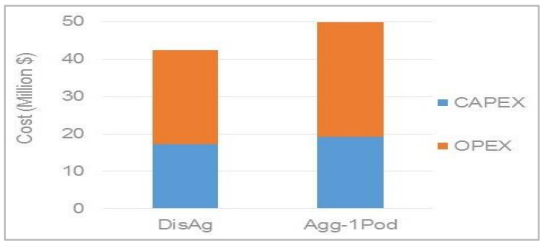

Fig. 14. Datacenter TCO for single application type

VI. Conclusions

This paper presents a comprehensive techno-economic framework for estimating the cost of ownership of running a datacenter. The proposed framework supports a detailed cost comparison of different technologies, architectures, and hardware configurations. Moreover, it also allows to evaluate the impact of running different application types.

The first part of the paper focuses on detailing the framework and presenting the rationale behind it. The second part takes as a case study the prominent case of disaggregated hardware datacenter architectures and evaluates it towards the proposed framework. The study comprises a comparison of having separate pools of resources in a datacenter (as in a disaggregated architecture) towards having currently available server-based architectures with homogeneous and heterogeneous hardware configurations. The results show that cost savings of around 40% are achieved using disaggregated hardware for a 10 years' period of datacenter operation. The savings are a result of better utilization in disaggregated architecture as well as independent lifecycle management of components (due to components being arranged by pools instead of mixed as in traditional servers). Moreover, a detailed TCO breakdown is presented which allows datacenter operators to have a better understanding of their TCO dynamics to act upon minimization of CAPEX and OPEX during deployment and operational phases.

Finally, the impact of uncertainty in some input parameters and assumptions on the cost results is evaluated. It was shown

that the longer the datacenter lifetime, the lower the investment per year. Moreover, amounts of cost savings increases slightly by expanding the size of datacenters. Our results also show the importance of sharing the resources among multiple applications or tenants to maximize the benefit from disaggregated hardware architectures.

TCO assessments can give an idea of the cost associated to deploying and operating a datacenter. However, a thorough business viability assessment is needed to understand the return on investment and revenue stream. Therefore, a possible future direction for this study would be to include cash flow analysis considering various business models to guide the operators on how much investment should be done at which time for each technology to have the greatest profit

REFERENCES

[1] GigPeak, "GigPeak Announces Record Quarter Shipment of ICs for Data Center Applications, and Sampling of SR and LR PAM4 IC Chipsets", September 2016 [Online]. Available: http://www.businesswire.com/news/home/20160915005508/en/GigPeak-Announces-Record-Quarter-Shipment-ICs-Data. [Accessed 21st November 2016]

[2] Gartner, "Gartner Says Worldwide IT Spending Is Forecast to be Flat in 2016", July 2016 [Online]. Available: http://www.gartner.com/newsroom/id/3368517. [Accessed 21st November 2016]

[3] Ericsson, "Hyperscale Cloud: Reimagining Datacenters from Hardware to Applications", White Paper, May 2016.

[4] S. Han, et.al., "Network Support for Resource Disaggregation in Next Generation Datacenters," in Proc. of the 12th ACM Workshop on Hot Topics in Networks, 2013, pp. 10:1–10:7. http://doi.acm.org/10.1145/2535771.2535778

[5] P. X. Gao, et. al., "Network Requirements for Resource Disaggregation," in Proc. of the 12th USENIX Conference on Operating Systems Design and Implementation (OSDI), 2016, pp. 249–264. doi:10.1145/2535771.2535778

[6] C. S. Li, et. Al., "Composable Architecture for Rack Scale Big data Computing," Future Generation Computer Systems (2017), pp. 180–193, https://doi.org/ 10.1016/j.future.2016.07.014

[7] S. Polfliet, F. Ryckbosch, and L. Eeckhout, "Optimizing the Datacenter for Data-Centric Workloads," International Conference on Supercomputing. June 2011, doi:10.1145/1995896.1995926

[8] Taneja Group Market Analysts, "For Lowest Cost and Greatest Agility, Choose Software-Defined Data Center Architectures Over Traditional Hardware-Dependent Designs," Technology Brief, August 2017

[9] S. A. Bain, I. Read, J. J. Thomas, and F. Merchant, "Advantages of a Dynamic Infrastructure: A Closer Look at Private Cloud TCO," IBM White Paper, 2009

[10] K. Lim, et. al., "System-level Implications of Disaggregated Memory," in High Performance Computer Architecture (HPCA), 2012, pp. 1–12, doi: 10.1109/HPCA.2012.6168955

[11] P. Costa, H. Ballani, K. Razavi, and I. Kash, "R2C2: A Network Stack for Rack-scale Computers," in Proc. of the ACM Conference on Data Communication (SIGCOMM), 2015, doi:10.1145/2785956.2787492

[12] J. Weiss, et. al., "Optical Interconnects for Disaggregated Resources in Future Datacenters," in European Conference on Optical Communication (ECOC), 2014, doi: 10.1109/ECOC.2014.6964255

[13] Intel, Intel Rack Scale Design (RSD), http://www.intel.com/content/www/us/en/architecture-and-technology/rack-scale-design-overview.html

[14] Ericsson, Hyperscale Data System 8000 (HDS 8000), http://www.ericsson.com/hyperscale/cloud-infrastructure/hyperscale-datacenter-system

[15] APC, "Determining Total Cost of ownership for Datacenter and Network Room Infrastructure," White Paper; http://www.apc.com/salestools/cmrp-5t9pqg/cmrp-5t9pqg_r4_en.pdf

[16] B. Grot, et. Al., "Optimizing Datacenter TCO with Scale-out Processors," IEEE Computer Society, doi: 10.1109/MM.2012.71

[17] Dell, "Shared Infrastructure: Scale-out Advantages and Effects on Tco," White Paper; http://www.dell.com/downloads/global/products/pedge/en/shared_infrastructure_scale_out_advantages_and_effects_on_tco.pdf

[18] B. Abali, R. J. Eickemeyer, H. Franke, C. S. Li, and M. A. Taubenblatt, "Disaggregated and Optically Interconnected Memory: When Will it be Cost Effective?," arXiv:1503.01416, 2015

[19] Mainstay, "An Economic Study of the Hyperscale Data Center," White Paper, January 2016

[20] S. Verbrugge, K. Casier, J. V. Oteghem, and B. Lannoo, "white paper: Practical steps in techno-economic evaluation of network deployment planning," 2009

[21] Data Center Power and Cooling, White Paper, August, 2011, http://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/unified-computing/white_paper_c11-680202.pdf

[22] C. Delimitrou, and C. Kozyrakis, "Quasar: Resource Efficient and QoS Aware Cluster Management", in Proc. of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), 2014, pp. 127-144, doi:10.1145/2541940.2541941

[23] SAP System Requirement; https://dcos.io/docs/1.7/administration/installing/custom/system-requirements/

[24] Video on Demand Application Example; http://www.unified-streaming.com/cases/performance-vod-and-live-use-cases-customer-examples

[25] B. Hindman, et Al., "Mesos: A Platform for Fine-grained Resource Sharing in the Datacenter," in Proc. of the 8th USENIX Conference on Networked Systems Design and Implementation, 2011, pp. 22–22.

[26] Dell, PowerEdge R830 Rack Server, http://www.dell.com/us/business/p/poweredge-r830/fs

[27] Dell, PowerEdge R630 Rack Server, http://www.dell.com/us/business/p/poweredge-r630/pd?ref=PD_OC

[28] Dell, PowerEdge R430 Rack Server, http://www.dell.com/us/business/p/poweredge-r430/pd?ref=PD_OC

[29] Intel, Data Center Blocks Warranty and Support, http://www.intel.com/content/dam/support/us/en/documents/server-products/server-boards/DCB_Warranty_Brief_Sept_2016.pdf

[30] Atto, Technical Specifications Fast Frame NIC, https://www.atto.com/software/files/techpdfs/TechnicalSpecifications_FastFrameNIC.pdf

[31] Qlogic, Overlapping Protection Domains, http://www.qlogic.com/Resources/Documents/TechnologyBriefs/Adapters/OverlappingProtectionDomains.pdf

[32] J. Meza, Q. Wu, S. Kumar, and O. Mutlu, "Revisiting Memory Errors in Large-Scale Production Data Centers: Analysis and Modeling of New Trends from the Field," in Proc. of 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 2015, pp. 415-426, doi=http://dx.doi.org/10.1109/DSN.2015.57

[33] Intel® Xeon® Processor E5-2600 v4 Product Family, http://ark.intel.com/products/series/91286/Intel-Xeon-Processor-E5-2600-v4-Product-Family#@All

[34] Dell, RAM 64 GB, http://www.dell.com/en-us/shop/accessories/apd/a8451131?c=us&l=en&s=dhs&cs=19&sku=A8451131

[35] Dell, RAM 32 GB, http://accessories.us.dell.com/sna/category.aspx?c=us&l=en&s=biz&cs=555&mfgpid=239010&category_id=4325&~ck=bt

[36] Sandisk, SSD drive 960GB, http://shop.sandisk.com/store/sdiskus/en_US/DisplayProductDetailsPage/productID.304914300