Resource Management Challenges in the Next Generation Cloud based Systems: A Survey and Research Directions

Ragaa A. Shehab
Computer and Systems Engineering
Faculty of Engineering, ASU
Cairo, Egypt
13805@eng.asu.edu.eg

Mohamed Taher
Computer and Systems Engineering
Faculty of Engineering, ASU
Cairo, Egypt
mohamed.taher@eng.asu.edu.eg

Hoda K. Mohamed
Computer and Systems Engineering
Faculty of Engineering, ASU
Cairo, Egypt
hoda.korashy@eng.asu.edu.eg

***Abstract*— Traditional resource management techniques for the conventional cloud computing systems will not be able to meet the QoS required for the IoT connectivity applications, big data analytics applications, and cognitive computing applications. Toward the next generation edge clouds, this work analyzes the dissimilarities of resource management issues between conventional cloud computing and edge cloud computing systems. In addition, we provide a classification and a summary for the resource management challenges of the next generation cloud based systems. Finally, the future research directions are highlighted.**

***Keywords*— Resource Management , Edge Cloud, Edge Computing, Fog Computing, Cloudlets, MEC.**

I. Introduction

Resource management in the conventional cloud computing systems is a hard problem, due to the scale of the datacenters, resources heterogeneity and independence, as well as the wide range of management objectives [1].

Recently, research focuses on the limitation of the conventional centralized cloud computing systems to meet the required QoS for the innovated applications. This is due to the increasing number of the connected sensors and devices, with large volume of emitted data. As well as the remoteness and centralization of the conventional clouds that suspect to bottleneck, congestion, and latency [2].

The next generation cloud computing systems [3],[4],[5] aim to tackle the issue by supporting multiple providers in a decentralized computing paradigm closer to the end users, instead of using a remote centralized datacenters from a single provider.

In this work, we aim to study the dissimilarities in resource management issues between conventional centralized cloud computing and distributed edge clouds and edge computing. An analysis to the resource management challenges for the next generation cloud based systems was provided. In addition, the future research directions in the field are highlighted.

Paper organization is as follow: Section 2 analyzes the disparities in resource management issues for both centralized conventional clouds and distributed edge clouds. Section 3 classifies the resource management challenges for the next generation cloud based systems: challenges due to innovated infrastructure, computing paradigm and impacted application areas are discussed. Section 4 summarizes resource management challenges and open research topics for the next generation cloud based systems. Section 5 concludes the paper.

II. Dissimilarities in Resource Management Issues in Conventional Cloud Computing and Edge Cloud Computing

Resource management is the process of allocating computing, storage, networking and (indirectly) energy resources to a set of applications. The aim is to jointly meet the performance objectives of the applications, the infrastructure (e.g., data center) providers, and the users. The enabling technologies are virtualization and containers. Virtualization facilitates the statistical multiplexing of resources across customers and applications. Container is an alternative to the virtualization technology, except that it shares the same operating system kernel for all applications.

In this section, we aim to analyze and highlight the dissimilarities in resource management issues between conventional cloud computing toward the next generation edge clouds and edge computing. Dissimilarities in resources' nature, applications' nature, and in management techniques are provided in table1, table2, and table3 respectively.

A. Dissimilarities in Resources' Nature[5],[6],[7].

TABLE I. DISSIMILARITIES IN RESOURCES' NATURE

Resources	Conventional Cloud Computing	Edge Cloud and Edge Computing
Model	• Centralized shared pool of configurable computing resources. pooled to serve multiple consumers using	• Distributed, federated and hierarchical computing resources. • Complement the cloud computing model, extra resources could be

Resources	Conventional Cloud Computing	Edge Cloud and Edge Computing
	multi- tenant model • Physical and virtual resources • Computing, storage, and networking resources	rented from the cloud. • Physical and virtual resources • Computing, storage, and networking resources
Size and computatio nal power	• Very large in size, each typically could contain tens of thousands of servers with high computation capacity.	• Resources at each location may differ in size and capacity, spanning from a light weight node to a medium computational capacity node according to the requirements.
Availabilit y	• Small number of large size datacenter	• Large number of small size datacenter
Scalability	• Central scalability	• Central and edge scalability
Belongs to	• Powerful servers	• Servers, routers, switches, access points, gateway, on box road side unit,..
Location	• Centralized in limited numbers of big datacenters • Far from the user. • Accessed by internet connection.	• Distributed in large number of geo-graphical locations and hierarchally distributed anywhere along the continuum between Cloud and Things. • At the network edge closer to end user. • Accessed by local connections.
Operation	• Large companies of cloud operators and technical expert teams	• Customers on their premises, or small to large companies, depending on size.
Market condition	• Monopoly	• Monopoly break, and fairer competition among all parities.

B. *Dissimilarities in Applications' Nature.[3],[6]*

TABLE II. DISSIMILARITIES IN APPLICATIONS' NATURE.

Application	Conventional Cloud Computing	Edge Cloud and Edge Computing
Examples	• Mainstream cloud applications.	• smart vehicles, e-healthcare, WSN, smart grid, smart IP cameras, smart cities, smart factory, smart building, smart supply chain,…
Types	• delay tolerant application • computationally intensive application	• Real time and interactive applications. • Predictable and low latency applications • Contextual location awareness applications. • Large scale, • Geographically distributed applications • Seamless mobility and • wide coverage support applications
Bandwidth consumption	• Consume huge backbone network bandwidth	• Low backbone bandwidth consumptions as data are processed and filtered locally; insights only need to be transferred to the cloud. • No network congestion.
Application placement	• Application modules are centralized at remote datacenters	• Distributed anywhere along the continuum between cloud and things.
Application management operator	• cloud operator, experts team	• applications placement, deployment and control is under research

C. *Dissimilarities in Resource Management Techniques [1],[4].*

TABLE III. DISSIMILARITIES IN RESOURCE MANAGEMENT TECHNIQUES

RM	Conventional Cloud Computing	Edge Cloud and Edge Computing
Mobility support	• Does not support user's mobility	• Wide coverage with user mobility support
Partners	• Cloud provider • Cloud user (broker) • End user	• Cloud provider • Cloud user/Edge provider • End user/ IoT device
Operator/ business model	• Owned by single operator. i.e. Amazon, Google, azure,…	• Private companies should cooperate for federated deployment, and operations. • organizational standards for this cooperation is (under research)
Managemen t Model	• client / server model	• client / server model • peer to peer communication • distributed model • cluster based model
Objectives	• Satisfy customer's SLA (service availability, performance, …..), • Satisfy infrastructure management objectives • Load balancing • Fault tolerance • power consumption	• Provide stringent latency requirement. • Control Workload size at edge nodes • Control data transmission size and frequency to the cloud • Monetary cost control for rented cloud infrastructure, • Power consumption • Workload balance…(open research)
Research directions	• Focus on relation between cloud provider and cloud user for IaaS, PaaS, and SaaS models.	• Focus on interplay and interaction between cloud and edge nodes • Focus on cooperation between neighbor edge nodes …(open research)
Decision maker	• Centralized cloud management system	• Cloud management system offload workloads to edge nodes in centralized schemes • Edge node offload its workload to cloud or to neighbor edge nodes for better QoS support • Broker in between offload to cloud or edge nodes according to predefined objectives. (open research)

III. Resource Management Challenges In The Next Generation Cloud Based Systems

According to [8], evolved infrastructure, computing paradigms and impacted applictions are the main factors contributing to the next generation cloud based systems, see figure[1]. In this section, we analyze from a resource management perspectives the challenges facing those paradigms.

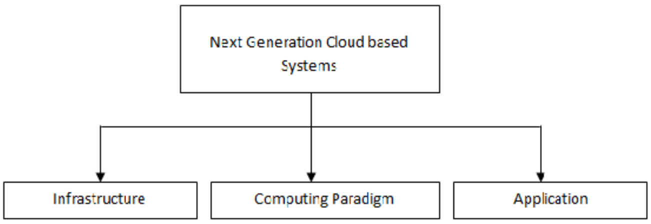

Fig. 1. Main Factors Contributing to The Next Generation Cloud Based Systems

A. Resource Management Challenges due to the new Emerging Cloud Computing Infrastructure

Figure[2] provides alternative infrastructure models to the centralized conventional cloud computing systems.

1) **Multi cloud:** owned by a single provider or multiple providers. Multi cloud could be hybrid or federated cloud. Hybrid cloud is a combination of public and private clouds. Federated clouds brings together different cloud providers under a single umbrella. Resource management challenges for this model include:

- Pricing and billing model need to account for different types of resources offered by multiple providers.
- Resources are dynamically added or removed with no unified catalogue that able to report the available resources on the cloud.
- No unifying environment, resource management need to be programmed manually.
- awareness of the bandwidth, latency and network topology when accessing public cloud from private cloud is a must.
- application and data migration across multiple vendors.

2) **Micro Cloud and Cloudlet:** extending cloud infrastructure towards the edge of the network. It deploy small sized, low cost and low power computing processors co-located with routers and switches or located in dedicated spaces closer to user devices for general purpose computing. Cloudlet generally used in literature in the context of mobile computing. Resource management challenges include:

- Edge nodes networking and management over multiple sites.
- Application partitioning and scheduling utilizing close low power micro-clouds along with the remote powerful datacenters.
- Workload offloading criteria and management, considering user to edge offload, cloud to edge offload and edge to edge offload.
- Resource availability and certainty using only network management abstraction.

3) **Adhoc Cloud:** in this paradigm, underutilized resources, such as smart phones or servers owned by organizations can be harnessed to create an elastic infrastructure. Which can support low latency computing for unprivate non-critical applications. Resource management challenges are:

- Available resources are unknown beforehand.
- Reliable resource management mechanisms to alleviate the malicious activities.
- Power aware resource management for the restricted resources devices contributing in the adhoc clouds

4) **Heterogeneous Cloud:** Heterogeneity arises from multiple cloud providers using hypervisors and software suites from multiple vendors. In addition, low-level heterogeneity at the infrastructure level arise from different types of processors combined to offer virtual machines (VM) with heterogeneous compute resources. Resource management challenges are:

- Management of the heterogeneous resources needs high level abstraction.
- Application scheduling and workload deployment across heterogeneous resources.

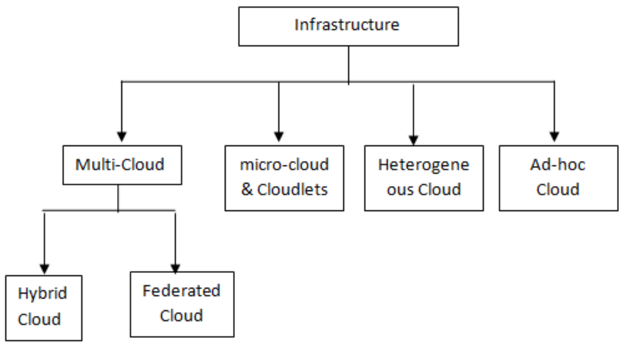

Fig. 2. *Alternative infrastructure models to the centralized conventional cloud computing systems.*

B. Resource Management Challenges due to The Emerging Computing Paradigms

Figure[3] provides alternate compuitng paradigms to the centralized conventional cloud computing systems.

1) **Volunteer Computing:** also known as ad-hoc computing, and peer to peer computing. Resource management challenges are:

- Unguaranteed availability of resources result in challenges in resource reservation and pricing.
- Management complexity of the heterogeneous underlying infrastructures.
- Security procedures have to be imbedded in the resource management to encourage volunteers to participate.

2) **Serverless Computing:** conventional cloud requires application to be hosted in a virtual machine (VM), and the basic cost model is per VM per hour, and does not consider the idle time of the VM. Serverless computing paradigm aim to bill the user per application's actual processing and memory cost. Resource management challenges for this paradigm are:

- Based on modules execution, and for a function as a service, new resource management and pricing model have to be considered. No over or under provisioning of resources have to be considered, instead developers have to consider control, cost and flexibility of the applications.
- Implementing serverless computing at the edge, will facilitate and cope with the increasing number of IoT devices connected to the network, however studying the tradeoff between using traditional services and serverless services have to be considered.

3) **Software Defined Computing:** this approach isolate the underlying network hardware from the data control.

- Software defined could be applied to networking, storage and computing for alleviating resource management complexity in configuring and operating the underlying physical resources.

4) **Fog Computing:** fog utilizes hierarchal networking, computing all over the path from end user to the cloud datacenters. Fog applications can be vertically and horizontally partitioned. Vertically partitioned at various levels of the hierarchal network and horizontally partioned across neighbour peer nodes. In the same time workload could be offloaded in two directions. First, from cloud datacenters to fog nodes, second, from end devices to fog nodes. Fog node may be a micro cloud and cloudlet, thus thier challenges are included, other resource management challenges are:

- Distribution of resources over multiple tiers of hierarchal networking adds extra degree of difficulty to its management
- Fog edge nodes have limited resources (switch, router, base station) unless it will be a general purpose computing or cloudlet.
- Intelligent management techniques should be used to cope with the heterogeneity in edge nodes, enable better edge scalability, and to avoid overwhelming the cloud management system.
- Vertical and horizontal resource provisioning for application's scaling up.
- Forward and backward workload offloading criteria.

5) **Mobile Edge Computing** (MEC): also known as multi access edge computing for cellular networks, exist in the context of 5G systems. May or may not be connected to the conventional cloud datacenters. Resource management challenges:

- Main effort are concentrated over the radio access network (RAN) resources as the most valuable resource.
- Cloud may have a limited role in MEC resource management.

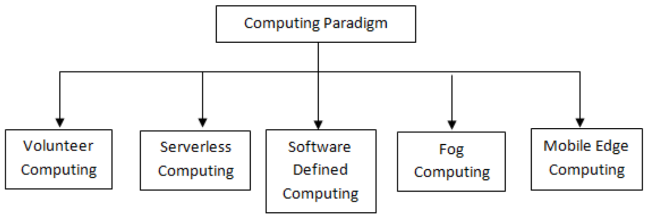

Fig. 3. Alternative compuitng paradigms to the conventional cloud computing systems.

C. *Resource Management Challenges due to the Emerging Applications*

Figure[4] provide an alternate applications to the conventional cloud computing systems.

1) **IoT Connectivity:** it implies embedding sensors into infrastructures (transportation, building healthcare, communication,…) and into people devices (wearable , smart phones,…). These sensors emit variety of data. Resource management challenges are:

- Secure resource management technique across sensor network , wireless network and data centers networks
- Innovated resource management technique that capable of dealing with structured, semi structured and unstructured data emitted by the huge number of sensors.

2) **Big Data Computing:** emerging applications and computing models generate huge amount of data referred to as big data. Storing these data at the conventional cloud for analytics is an expensive and unpractical solution. The trend is to process data at edge clouds close to its generating sources and to store only the resultant knowledge. Resource management challenges are:

- Resource management techniques that are able to deal with the geographically distributed cloud nodes, that may be ad- hoc clouds, heterogeneous clouds with various processing power and infrastructures, while minimizing the developers' programming effort.
- Resource management framework for data analytics, this framework should be able to scale horizontally across multiple hosts of the same datacenter and vertically across low power end processors to high processing capabilities data centers.

3) **Cognitive Computing:** currently social media users generate huge amount of images, audio and videos. Cognitive computing relaies on machine learning algorithms for video analysis and image recognition which consumes high processing power. Resource management challenges:

- Resource management techniques should be able to deal with hardware accelerators for reducing computing time of machine learning algorithms over large volume of data.

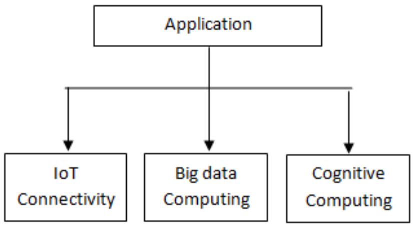

Fig. 4. *alternate application areas to the conventional cloud computing systems.*

IV. Summary of Resource Management Challenges in the Next Generation Cloud Computing Systems

Figure[5] summarizes the resoruce mangement challenges for the next generation cloud computing systems.

1) Heterogeneity [9] that arises from:

a) Computing power (hardware, and software) ranging from a resource constrained access point to powerful datacenter. But generally edge clouds have a medium to weak computing and storage capabilities.

b) Connected devices ranging from a resource constrained sensor or dump object and smart phones and up to a powerful processor laptop.

c) Business model, owners, brokers, resource management has to deal with Vendor lock in problem given different abstractions, resource types, networks and images as well as variable vendor specific costs for migrating large volumes of data.

d) Service Level Agreements (SLA) across multiple owners and brokers.

e) Accounting and Pricing models.

f) Applications: delay tolerant to real time applications

g) Netwrok topology, latency and bandwidth

h) Processed data: structured , semi structured , or unstructured data

2) Distribution: resources may be horizontally and vertically distributed, horizontal resource distribution e.g: volunteer clouds, MEC . horizontal and vertical distributed resoruces e.g: fog computing. Where in fog computing resources are distributed anywhere in the continuum from cloud to the thing at the network edge.

3) Standarization: lack of standards and suitable interfaces between conventional cloud, edge clouds, and connected devices.

4) Mobility: to provide any time, any where and any thing connectivity, end device movement may change its point of attachment, and dynamically alter various connectivity based metrics such as: bandwidth, computation, storage, latency. To meet the required applications QoS, resource management efforts have to be done including: resource reservation and provisioning to prepare the destination node. Mobility prediction could assist in resource management by predicting the end device destination. Extra overhead has to be done to support the service of device mobility, where mobility speed and pattern has to be considered.

5) Reliability: due to the distributed nature of resources, computational services across multiple hierarchical levels have to be validated. certainty have to be considered in the management process.

6) Elasticity: the dynamic nature of connected devices poses challenges to deal with unpredictabillty and variablity of the load. In adition to the possibility of resource failure and shortage. Flexibility of resource management is a must to ensure resource availability and to avoid downtime.

7) Scalability: for application and workload deployment, vertical and horizontal resources provisioning are needed.

8) Virtualization: although adds a level of flexibility in resource management it adds extra management overhead and susceptible to latency especially in its migration.

9) Power consumption: resource management has to monitor the power status of each device and allocate resources based on the power status and expected response and service type. Computational task offloading may be affected by the power status of the IoT device, to this end resources would be dynamically assigned at real time.

10) Data: adds another level of complexity to resource estimation, as multimedia data consume more resources than text data, and data security is a must in cases such as healthcare data

11) Security: due to the heterogeneity, resources is vulnerable toward security attacks

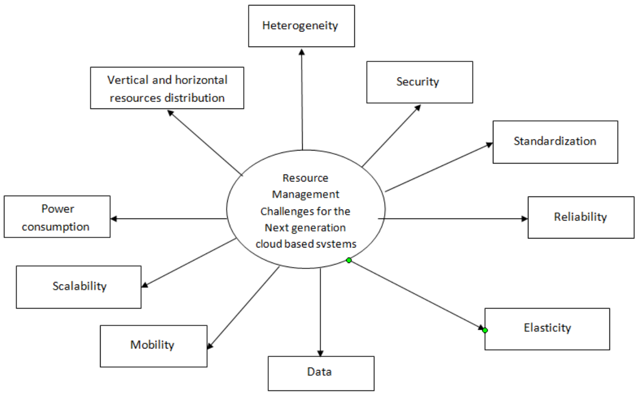

Fig. 5. Summary of the resource mangement challenges for the next generation cloud based systems

V. Trends And Future Resource Management Research Directions (Gaps In The Research Area)

1) Secure resource management techniques to avoid malicious usage of hybrid , adhoc and heterogeneous distributed clouds. As well as being able to deal with vendor lock in problem for the coordination in application and workload deployment specially for horizontal and vertical application scale up.

2) Standardization of:

a) Multi parity service level agreement (SLA) ,

b) Cloud interworking and interfaces between centralized cloud, edge clouds, end devices, under various network topologies.

c) Accounting and pricing models that cope with the clouds heterogeneity

d) Billing model for serverless computing.

3) Energy aware resource management techniques for the heterogeneous distributed clouds

4) Location awareness resource management technique with mobility support to assign a deployed application certain affinity over a specified edge cloud

5) Further resource management research should be directed to fully utilize micro clouds and cloudlets computing capabilities' to serve the local traffic for reducing traffic load and latency toward the remote cloud

6) Adaptive resource management framework and techniques that are able to deal with wide variety of applications while maintaining large number of heterogeneous dynamically connected end devices

7) Software defined network aided resource management framework.

8) Resource demand profiling, for the connected end devices that aim to facilitate the process of resource reservation and management.

9) The study of resource management performance under virtualization versus containerization techniques.

10) Study and optimization of resource management footprint in distributed clouds in respect to centralized resource management techniques

11) Efficient resource monitoring system across multiple heterogenous distributed edge clouds.

VI. Conclusions

Resource management in the next generation cloud based systems is a hot topic for research. Resource management realizes the deployment of a class of highly demanded applications with better quality of service while optimizing the resources utilization. Due to the innovated clouds' infrastructures, computing paradigm, and emerging applications, resource management faces considerable challenges. In this work, a survey of resource management challenges for the next generation cloud based systems was proposed, along with the future research directions.

References

[1] Brendan Jennings and Rolf Stadler. Resource management in clouds: Survey and research challenges. Journal of Network and Systems Management, 23(3):567–619, Jul 2015. ISSN 1573-7705. doi: 10.1007/s10922-014-9307-7. URL https://doi.org/10.1007/s10922-014-9307-7.

[2] Luiz F. Bittencourt, Omer Rana, and Ioan Petri. Cloud computing at the edges. In Markus Helfert, V´ıctor M´endez Mu˜noz, and Donald Ferguson, editors, Cloud Computing and Services Science, pages 3–12, Cham, 2016. Springer International Publishing. ISBN 978-3-319-29582-4.

[3] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu. Edge computing: Vision and challenges IEEE Internet of Things Journal, 3(5):637–646, Oct 2016. ISSN 2327-4662. doi:10.1109/JIOT.2016.2579198.

[4] J. Pan and J. McElhannon. Future edge cloud and edge computing for internet of things applications. IEEE Internet of Things Journal, 5(1):439–449, Feb 2018. doi:10.1109/JIOT.2017.2767608

[5] M. Chiang and T. Zhang. Fog and iot: An overview of research opportunities. IEEE Internet of Things Journal, 3(6):854–864, Dec 2016. ISSN 2327-4662. doi: 10.1109/JIOT.2016.2584538.

[6] Robert Barton (Cisco) Michael Martin (IBM Canada) Nedim Goren (NIST) Charif Mahmoudi (NIST) Michaela Iorga (NIST), Larry Feldman (G2). The nist definition of fog computing: Fog computing conceptual model, march 2018. URL https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.500325.pdf).

[7] Timothy Grance Peter Mell. The nist definition of cloud computing, September 2011. URL https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.

[8] Blesson Varghese and Rajkumar Buyya. Next generation cloud computing: New trends and research directions. Future Generation Computer Systems, 79:849 – 861, 2018. ISSN 0167-739X. doi: https://doi.org/10.1016/j.future.2017.09.020. URL http://www.sciencedirect.com/science/article/pii/S0167739X17302224

[9] Behailu Negash, Amir M. Rahmani, Pasi Liljeberg, and Axel Jantsch. Fog Computing Fundamentals in the Internet-of-Things: Intelligenceat the Edge, pages 1–100. Springer International Publishing, Cham, 2018. ISBN 978-3-319- 57639-8. doi: 10.1007/978-3-319-57639-8 1. URL https://doi.org/10.1007/978-3-319-57639-8_1