Science World Journal Vol. 16(No 3) 2021
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

PROPOSED APPROACH FOR RESOURCE ALLOCATION MANAGEMENT IN SERVICE ORIENTED ARCHITECTURE (SOA) ENVIRONMENT

Salome Danjuma, Gabriel Lazarus Dams, Ashiru Simon

Department of Computer Science, Kaduna State University, Tafawa Balewa Way, – Kaduna, Nigeria

*Corresponding Author Email Address: damsgabe@kasu.edu.ng Phone: +2348060137413

ABSTRACT
Service Oriented Architecture (SOA) uses the concept of distributed systems. It is an underlying framework that manages different services using component-based software engineering that invoke interfaces in collaboration with web services management. SOA is known as the backend behind all service oriented systems like cloud computing, grid computing, utility computing and web services. In this paper SOA covers the gap between implementations of web services and consuming applications giving a logical view of all the available resources.
This paper focuses on proposing an approach to resource allocation management across SOA environment to deal with how these resources are managed and allocated to these services. The approach is known as Active/Standby Resource Management Allocation (ASRMA). The main concern for the proper management of these resources is to minimize waiting time of request to be allocated resources thereby meeting an aspect of the Quality of Service (QoS) to multiple providers and requesters as well. Resource allocation basically deals with assigning resources to each incoming service request, these allocations deal with time and limited resources. The challenge of ever increasing needs of users or consumers with limited service resources such as CPU time, memory capacity and network bandwidth prompted the need for quality of service thereby improving proper allocation of these resources. This approach can be achieved by having more resources on standby in case of over utilization or underutilization of the active resources.

Keywords: Active Resource Pool, Standby Resource Pool, Service-Based System (SBS), Resource allocation, Service Request, Quality of Service (QoS)

INTRODUCTION

Computer application development is the ever fastest growing field in computer world, changing and shaping lives for the greater good. Many of our computing resources and infrastructures today are being managed by service-based systems (SBS) such as cloud computing, grid computing and utility computing. Service-orientation has become the main architectural pattern for distributed embedded systems which has made all applications to be virtually service based. Businesses and government rent services from different providers to carter for their computing needs and manage their resources. Therefore, service can be seen as a self-contained software component with the ability to perform a specific or more tasks within a given period of time (Yau and An, 2009,Wei et al., 2013).
Service Oriented Architecture (SOA) is an architecture that is dynamic and loosely coupled in nature with each service rendering its function independent of the other. It is the framework behind the orchestration to design applications that uses such services. According to Yau and An (2011), it is an architectural style required for application development. World Wide Web Consortium (W3C) described SOA as a component-based architecture where interfaces can be invoked and discovered; as such, these services come into play with the help of web service interface. Web services uses Simple Object Access Protocol (SOAP) and Extensible Markup Languages (XML) among other technologies for communication over HTTP (Huang et al., 2019).
This architecture makes it possible for network resources to be consumed as autonomous software services that can be tapped into without the full knowledge of the underlying technologies. This architecture operates independent of these technologies with well-defined interfaces. The unique thing about this architecture is the way interfaces are invoked in a standard way without the knowledge of the client on how it is being done (Huang et al., 2019). SOA comprises of two entities which are program and request; the program integrates the services together thereby providing the request with desired throughput. Yau and An (2009),Yau and An (2011),Wei et al. (2013), viewed SOA as an architecture that facilitates many service-based systems.
Resource allocation recognizes and assigns resources for a specific period to various activities within the component service resource pool (Dongre and Ingle, 2019). This process is challenging the impact of Quality of Service (QoS) parameters such as accessibility, timeliness and reliability (Alshinina and Elleithy, 2017,Singh et al., 2021). There is a need for effective selection of resources to meet clients requirements or requests (Barenji and Barenji, 2017, Wu et al., 2020). Resource Allocation Management allocates resources on the basis of unutilized resources in a pool of virtual machines. It also analyzes shared resources and its value within time and space for better decision making on resource utilization (Huang et al., 2019).
This paper focuses on resource allocation management in Service-Based System (SBS) with regards to QoS requirements for performance such as response time. The target is to offer better utilization of resources within the stipulated time with limited resources, thereby increasing the efficiency of the flow of services within the SOA environment.

Related Works and Challenges

Most of the related works under resource allocation across SOA environment are becoming more popular to improving the QoS throughput. Many researchers used different models and algorithms to give a view on how performance will be increased when resources are properly managed. Hussain et al. (2013) for instance, used the scheduling algorithm which uses the high

Science World Journal Vol. 16(No 3) 2021
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

performance computing (HPC) systems within clusters, grid and cloud systems environment for the analysis. The technique optimizes the resources in ideal situation taking note of the QoS requirements. The scheduling process organizes the allocations of the resources into centralized, decentralized and hierarchical. In Hossain et al. (2012), a video surveillance as a service technique (VSaaS) allocates its resources through the use of computational resources and virtual machine resources; therefore the virtual machine (VM) resource allocation model was proposed to meet the QoS requirement within SOA environment.

Adaptive resource management algorithm was proposed by Wei et al. (2013) to predict future resource requirement of a service. It is an agent-based framework that utilizes different types of agents to coordinate the allocation, distribution and termination of virtual resources.

An adaptive resource allocation for Service-based System (SBS) to achieve QoS feature known as throughput was proposed by Yau and An (2009). A resource allocation throughput model was developed for both atomic and composite services to analyze the relationship between resource allocation and throughput of services within the SBS.

Alshinina and Elleithy (2017), integrated SOA and Wireless Sensor Network (WSN) to address challenges like communication, QoS and heterogeneitics of sensor hardwares. It is based on providing low power communications in memory usage and transmission.

The work in Dongre and Ingle (2019) avoid service delay to improve QoS in terms of accessibility, reliability and timeliness. It also uses the prioritization algorithm to prioritize request based on service level objectives.

Linear scheduling for task and resources (LSTR) is a technique by Abirami and Ramanathan (2012) to perform tasks and schedule resources accordingly; thereby maximizing the system throughput and resource utilization process.

All the related works mentioned, emphasized more on the resource allocation management and quality of service (QoS) optimization. Meanwhile, this paper investigates reduction of the response time in a situation of service queuing and delay in allocation of resources based on Active/Standby Resource Management Allocation (ASRMA) Approach.

Resource Allocations within SOA Environment

Resources are the most important assets in any organization. These organizations develop and manage services based on Service-Level Agreement (SLA) thereby satisfying QoS requirements (Mahendran and Mekala, 2018). In Yau and An (2011), Service Oriented Architecture (SOA) is an architectural model that integrate services with invoked standard interface, whereas in García-Valls et al. (2013), SOA has to do with building distributed applications in a decoupled way where services reside in remote nodes in the network and communicate via messages or events. These interfaces hide the complexity underlying the interactions among various services. Resource allocation deals with allocation of resources among various services which are independent of platforms and programming language due to its loosely coupled structure (Huang et al., 2019, Chandel and Poreddy, 2019).

Services under SOA have unique characteristics, these are:

- **Loosely coupled**: there are no direct dependencies among individual services.
- **Service Abstraction**: beyond the SLA description, a service hides its logic from the outside world.
- **Service Reusability**: services aim to support potential reuse.
- **Service Composability**: a service can comprise other services, and developers can coordinate and assemble services to form a composite;
- **Service Stateless**: to remain loosely coupled, services do not maintain state information specific to an activity, such as a service request.
- **Service Discoverability**: services allow service consumer use mechanisms to discover and understand their descriptions.

The Existing System

In Hussain et al. (2013), service-based systems (SBS) are large ultimate scale system that requires an appropriate architectural model that allows efficient management of geographically distributed resources over multiple administrative domains. These systems can be divided into computational, data and service grid in regard to the system at hand. Scheduling goes hand in hand with resource allocation because it defines ways in which resources are allocated. On the other hand request scheduler in Ravulakollu and Chejara (2013), discussed the scheduling process based on the service layer to provide better services from the user point of view. The service layer is categorized into User web Interface, Request Scheduler, Request Dispatcher. The overview of the system workflow is shown in figure 1.

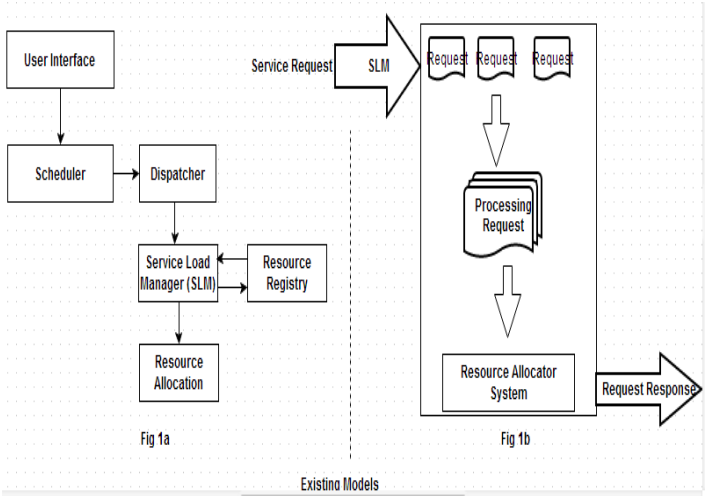

Figure 1: Resource Allocation Process for the Existing System

User web interface (from figure 1) is the user interaction platform within Service-Based System (SBS) where new accounts can be created and managed. New applications can be uploaded from this interface; user can also start/stop a service from running. Request scheduler takes care of different service requests arriving simultaneously, this number can be in thousands for a huge SBS. The scheduler handles each request one after the other and then buffer them before queuing them to the next step. The dispatcher takes the scheduled incoming request from the scheduler and arranges them according to their requested resources; the dispatcher then passes them to the service load manager then down to resource manager. The frequency of picking scheduled request for resource manager decides the scalability of the architecture.

Science World Journal Vol. 16(No 3) 2021
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Challenges of the Existing System

There are so many challenges and issues arising from the resource allocation system. More so, several proposed solutions have been done by researchers to curb the problem(Gawali and Shinde, 2018).

Some of these challenges can be seen as follows:

The time taken by each service request to respond to available resources. Figure 2 below shows a typical resource allocation flowchart where a service request is send to the service load manager (SLM) for a resource, and the service load manager (SLM) checks for the availability of the resource requested from the resource registry. If the resource is available, then the service will be allocated the resource. However, if the resource is not available, the service request falls on the queue which increases the response time of the request thereby affecting its throughput.

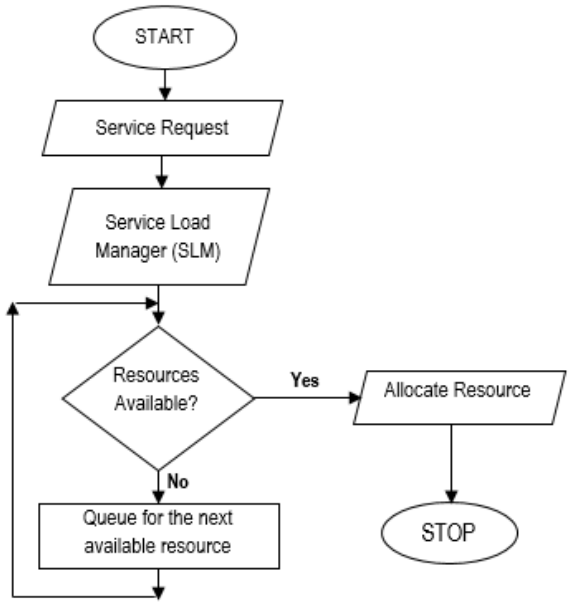

Figure 2: Typical Resource Allocation Flowchart

- Quality of Service (QoS) monitoring: There are so many providers having different polices and techniques working towards achieving that QoS features; features like throughput and service delay which rely on system resource allocation at the applications' runtime. These services compete for resources like CPU time, memory and network bandwidth which influence the dynamic change at runtime for resources status, workflow priorities and QoS requirements. Meeting up to the requirements of QoS, an effective technique need to allocate these system resources to each service effectively to maximize the response time (Yau and An, 2011).

Hence, there is need for a new approach to resource allocation system for SBS so as to address the challenges mentioned.

Overview of the Proposed System Approach

This approach makes effort to improve the resource allocation management showing how resources can be adequately allocated to a service request in runtime from a standby resource pool which will minimize the waiting time for available resource. An inactive standby resource pool is introduced in the approach which will only require activation before allocation. With this approach, the throughput requirements of the multiple workflows in Service-Based System (SBS) will be satisfied. More so, the approach addresses the challenges and issues of the existing Resource Allocation where service request needs to fall on queue for the next available resource.

The approach is named as Active/Standby Resource Management Allocation (ASRMA) which deals with service request workflow within SBS environment. The approach also analyzes the relationship between the resource allocation and throughputs of multiple workflows in SBS (Singh et al., 2021). It is good to know that it addresses the issues like:

- Optimal resource allocation providing the resources for service request workflow in SBS until service request are exhausted.
- The issue of priority of workflows when resource allocation is over utilized and is at critical state.

The major difference between the proposed approach and any other approach is the availability of system resources which are kept on standby based on the architecture of the SBS (Wei et al., 2013). The allocation of resources from the standby pool is done using the priority based allocation method (Nirav and Buchade, 2014).

Below is the algorithm for ASRMA system approach when a service request is sent for resource allocation:

Start
Step 1*: Request for service*
Step 2*: Check for incoming request requirements*
Step 3*: Group requests based on resource(s) required*
Step 4*: Check for resource availability*
Step 5*: **IF** resource is available in the active pool, assign resource based on first-come-first-serve (FCFS)*
***ELSE** check the standby resource pool, activate and allocate the resource based on priority using analytical hierarchy process (AHP) methodology.*
Step 6*: Notify the service load manager (SLM)*
Step 7*: Update the request registry*
Stop

As the request for service is received by the request registry with its request requirements, the service load manager (SLM) will check the incoming request from the request registry with its respective requirements. The SLM will group the requests based on the resource(s) required and notify the resource allocator to assign resource(s) to the request based on the availability of the resources. The resource allocator will check the active resource pool for resource availability and allocate the resource to each request group. Where there is group of requests waiting for allocation during runtime due to unavailability of resources in the active resource pool, the resource allocator will check the standby resource pool and activate the found resource to allocate. The methodology of allocation from the standby resource pool will be different from the one in the active resource pool. The analytical hierarchy process (AHP) methodology will be used for allocation of resources in this pool because of its significance in helping out with prioritization (Nirav and Buchade, 2014). Since minimizing waiting time of service request thereby meeting the quality of service (QoS) requirements on such request is the focus of this approach, there is need to for standby resource pool and prioritization of the request for allocation of resources using AHP methodology. The standby resource pool process is introduced in the flowchart (figure 3)

Science World Journal Vol. 16(No 3) 2021
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

where the service load manager (SLM) will check for the availability of the resource in the standby pool. The resources in this standby pool are inactive and awaiting activation. Once the required resource is found by the SLM, it will be activated and allocated to the service request. When allocation is done, the SLM is updated as well as the request registry.
The flowchart and the architectural view for the proposed resource allocation approach are shown in figures 3 and 4 respectively.

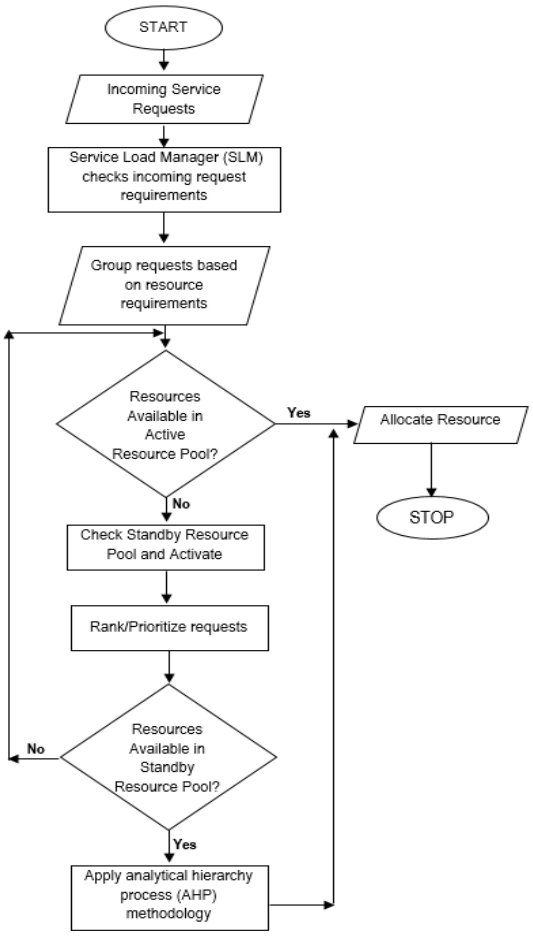

Figure 3: Flowchart for Proposed ASRMA Approach

Service request and its resource allocations is the subject of concern in this paper; Yau and An (2009) sees services as atomic and composite while García-Valls et al. (2013) described service as companion and as a graph. Service as a companion comprise of many service implementations while service as a graph comprise of nodes representing service request and the connecting lines as the message exchange.

Architectural View of Active/Standby Resource Management Allocation (ASRMA) Approach
The Active/Standby Resource Management Allocation (ASRMA) approach comprises of five (5) components namely: request/resource registry, service load manager (SLM), resource allocator, active resource pool and standby resource pool (figure 4). The components are explained and the relationship amongst them are described below.

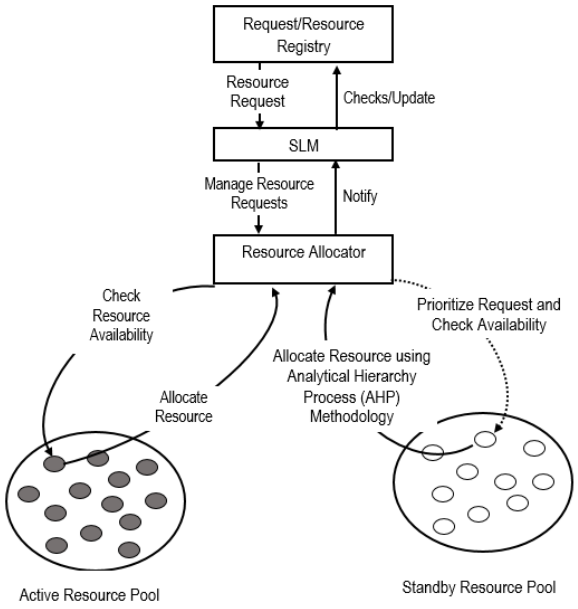

Figure 4: Architectural View for Proposed ASRMA Approach

- **Request/Resource Registry:** it manages the available resources such as hardware, software and virtual machines. More so, it stores the status of each service request that has been assigned a resource. To every service request assigned to a resource, there should be an update in the request/resource registry. The database server manages this registry which also stores some vital information about the machine like the IP address and MAC address.

- **Service Load Manager (SLM):** this checks the incoming service request from the request/resource registry first, to know if the available resource is ready for use. The request/resource feedback is either blocked, free or standby, based on the current state of the service-based system (SBS). The SLM is also responsible for grouping incoming service request based on the resource(s) required such as processor, memory, software, time for execution etc.

- **Resource Allocator:** this is in charge of allocating resources across the waiting queue. After the allocation, the request/resource registry is updated based on the allocation of resource to a service request.

- **Active Resources:** these are resources allocated to every service request within the workflow of SBS. These resources can be switched to standby when the need arises particularly when incoming requests are queuing up resources and some of these requests are high priority request. This is done by sending a message to the specific machine keeping these resources to activate the standby resources.

- **Standby Resources:** these resources are not active on the system, but can be activated by the resource allocation manager when they are needed. The mode of allocation of resources uses the priority algorithm and analytical hierarchy

process (AHP) methodology. AHP methodology is a technique that uses many criteria and attributes for decision making in a complex environment such as the Service-Based System (SBS). As the request for different resources increases which will create queue of requests, the AHP in the standby resource pool will help in deciding the prioritization and allocation of the resources based on what type of resources required so as to minimize the queue of requests that will be waiting for resources in the active resource pool.

Summary of Proposed ASRMA Resource Allocation Strategy

There are several resource allocation strategies for cloud computing. Every strategy proposed by a researcher is to address an existing resource allocation challenge within a SBS (Pradhan et al., 2016, Walia and Kaur, 2016). The ASRMA resource allocation strategy is based on the following process of which part of the concept is adapted from (Singh et al., 2021).

1. Identify request requirements
2. Check on resources availability
3. Allocate resource base on first-come-first-serve (FCFS)
4. Activate resource from standby pool where necessary
5. Allocate resource base on priority

As service request continuously come in, to compete for resources within the Service-Based System (SBS), the service load manager (SLM) identifies each request requirement and manages these requests by allocating resources through the resource allocator to each service, based on the availability of resources from the active resource pool and service load. SLM is needed to provide a balance distribution of load within the SBS environment. This is to ensure effective utilization for load balancing in terms of the number of requests handled by each machine, number of active requests and total CPU utilization. Once the active resource pool becomes exhausted and the resources within the pool are unavailable for allocation, the resource allocator will fall back to the standby resource pool to move and activate found resource(s) to the active resource pool and allocate it to the queuing requests.

It is worth noting that the SLM determines which resources to be placed on active or standby based on the incoming service request priority. The SLM needs to only send a message to the machine, to either switch to standby or activate a resource for service request. This process helps the SLM to maximize the limited resources within the SBS to fulfill the large number of service request. The request/resource registry stores the status of every resource and service request that has been assigned a resource. These status are being updated under a specific time interval.

Conclusion and Future Work

Resource allocation and management across service oriented architecture (SOA) environment includes resource discovery, allocation and monitoring process. The management of these resources involves physical resources such as CPU cores, disk space and network bandwidth. These resources are shared among virtual machines running heterogeneous workloads

In this paper, the resource allocation management with Quality of Service (QoS) requirements was reviewed and discussed with the aim of proposing an approach that will improve the service request response time within Service-Based System (SBS) environments. The Active/Standby Resource Management Allocation (ASRMA) approach was presented with the view to give a better performance of resource allocation and also reduce the response time with the available resources by introducing the standby resource pool.

This approach enables the service load manager (SLM) to effectively handle resource distribution among different user requests thereby increasing the performance of service management within the SBS. This approach will also aid in the reduction of the waiting time for the next available resource with the help of standby resources whose allocation is based on priority using the analytical hierarchy process (AHP) methodology which thus should improve the performance of Resource Allocation.

The effectiveness of this approach hasn't been evaluated but can be evaluated on the Google Cloud Platform using different micro service applications. More so, details of the proposed resource allocation approach in terms of modeling and the algorithms behind it can be seen as a future work.

REFERENCES

ABIRAMI, S. P. & RAMANATHAN, S. (2012). Linear Scheduling Strategy for Resource Allocation in Cloud Environment. International Journal on Cloud Computing: Services and Architecture, Vol. 2, pp. 93-17.doi: 10.5121/ijccsa.2012.2102

ALSHININA, R. & ELLEITHY, K. (2017). Performance and Challenges of Service-Oriented Architecture for Wireless Sensor Networks. Sensors (Basel, Switzerland), Vol. 17, pp. 1-39.doi: 10.3390/s17030536

BARENJI, A. V. & BARENJI, R. V. (2017). Efficient Resource Allocation in Mass Customization based on Service Oriented Architecture. ArXiv, abs/1702.03289

CHANDEL, V. & POREDDY, J. (2019). Resource Management in Mobile Cloud Computing: MSaaS & MPaaS with Femtocell and Wi-Fi. International Research Journal of Engineering and Technology (IRJET), Vol. 11, pp. 2286-2291

DONGRE, Y. V. & INGLE, R. B. QoS Based Optimal Resource Allocation in Service Composition for Heterogeneous Devices. 2019 International Conference on Communication and Electronics Systems (ICCES), 17-19 July 2019 2019. pp. 763-767.

GARCÍA-VALLS, M., BASANTA-VAL, P., MARCOS, M. & ESTEVEZ, E. (2013). A bi-dimensional QoS model for SOA and real-time middleware. International Journal of Computer Systems Science and Engineering,

GAWALI, M. B. & SHINDE, S. K. (2018). Task scheduling and resource allocation in cloud computing using a heuristic approach. Journal of Cloud Computing: Advances, Systems and Applications, Vol. 7(4), pp. 1-16.doi: https://doi.org/10.1186/s13677-018-0105-8

HOSSAIN, M. S., HASSAN, M. M., QURISHI, M. A. & ALGHAMDI, A. Resource Allocation for Service Composition in Cloud-based Video Surveillance Platform. 2012 IEEE International Conference on Multimedia and Expo Workshops, 9-13 July 2012 2012. IEEE, pp. 408-412.

HUANG, J., HE, J., YAN, K., LI, W., YANG, J., XIA, T., YE, Y. & TU, L. (2019). Research on Smart Grid Planning and Construction Based on GIS Resource Allocation Technology. IOP Conference Series: Earth and Environmental Science, 300, 042076.doi: 10.1088/1755-1315/300/4/042076

HUSSAIN, H., MALIK, S. U. R., HAMEED, A., KHAN, S. U., BICKLER, G., MIN-ALLAH, N., QURESHI, M. B., ZHANG, L., YONGJI, W., GHANI, N., KOLODZIEJ, J., ZOMAYA, A. Y., XU, C.-Z., BALAJI, P., VISHNU, A., PINEL, F., PECERO, J. E., KLIAZOVICH, D., BOUVRY, P., LI, H., WANG, L., CHEN, D. & RAYES, A. (2013). A survey on resource allocation in high performance distributed computing systems. Parallel

Computing, Vol. 39, pp. 709-736.doi: https://doi.org/10.1016/j.parco.2013.09.009

MAHENDRAN, N. & MEKALA, T. (2018). Improving Energy Efficiency of Virtual Resource Allocation in Cloud Datacenter. Indian Journal of Science and Technology, Vol. 11, pp. 1-8.doi: https://doi.org/10.17485/ijst%2F2018%2Fv11i19%2F174531

NIRAV, S. M. & BUCHADE, A. (2014). Priority Based Resource Allocation in Cloud Computing. International Journal of Engineering Research & Technology (IJERT), Vol. 3(5), pp. 855-857

PRADHAN, P., BEHERA, P. K. & RAY, B. N. B. (2016). Modified Round Robin Algorithm for Resource Allocation in Cloud Computing. In: Science, P. C. (ed.) International Conference on Computational Modeling and Security (CMS 2016). Elsevier B.V.

RAVULAKOLLU, K. & CHEJARA, P. (2013). Service-Oriented Cloud Architecture Schema To Bridge Gap Between Student, Staff And Academia. Indian Journal of Computer Science and Engineering, Vol. 4(5), pp. 324-337.

SINGH, H., BHASIN, A. & KAVERI, P. R. (2021). QRAS: efficient resource allocation for task scheduling in cloud computing. SN Applied Sciences, Vol. 3, pp. 1-7.doi: 10.1007/s42452-021-04489-5

WALIA, N. K. & KAUR, N. Resource Allocation Techniques in Cloud Computing: A Review. Proceedings of 4th International Conference on Advancements in Engineering & Technology (ICAET-2016), 2016. pp. 638-641.

WEI, Y., BLAKE, M. B. & SALEH, I. (2013). Adaptive Resource Management for Service Workflows in Cloud Environments. Proceedings of the 2013 IEEE 27th International Symposium on Parallel and Distributed Processing Workshops and PhD Forum. IEEE Computer Society.

WU, D., YANG, Z., WANG, H., YANG, B. & WANG, R. UCRA: A User-Centric Context-Aware Resource Allocation for Network Slicing. 2020 International Conference on Computing, Networking and Communications (ICNC), 17-20 Feb. 2020 2020. pp. 808-814.

YAU, S. & AN, H. (2011). Software Engineering Meets Services and Cloud Computing. Computer, Vol. 44 (10), pp. 47-53.doi: 10.1109/MC.2011.267

YAU, S. S. & AN, H. G. (2009). Adaptive resource allocation for service-based systems. Proceedings of the First Asia-Pacific Symposium on Internetware. Beijing, China: Association for Computing Machinery.