BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **23**, No 3
Sofia • 2023 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2023-0024

Analysis of Energy and Network Cost Effectiveness of Scheduling Strategies in Datacentre

Afia Bhutto[1], *Aftab Ahmed Chandio*[1+], *Kirshan Kumar Luhano*[2], *Imtiaz Ali Korejo*[1]

[1]*Institute of Mathematics and Computer Science, University of Sindh, Jamshoro, Pakistan*
[2]*Sindh University Laar Campus at Badin. University of Sindh, Jamshoro, Pakistan*
[+]*Corresponding author*
E-mails: afiabhutto1@gmail.com chandio.aftab@usindh.edu.pk kirshan.luhano@usindh.edu.pk imtiaz@usindh.edu.pk

Abstract**:** *In parallel and distributed computing, cloud computing is progressively replacing the traditional computing paradigm. The cloud is made up of a set of virtualized resources in a data center that can be configured according to users' needs. In other words, cloud computing faces the problem of a huge number of users requesting unlimited jobs for execution on a limited number of resources, which increases energy consumption and the network cost of the system. This study provides a complete analysis of classic scheduling techniques specifically for handling data-intensive workloads to see the effectiveness of the energy and network costs of the system. The workload is selected from a real-world data center. Moreover, this study offers the pros and cons of several classical heuristics-based job scheduling techniques that take into account the time and cost of transferring data from multiple sources. This study is useful for selecting appropriate scheduling techniques for appropriate environments.*

Keywords**:** *network-intensive, energy cost, data center, workload, cloud computing.*

1. Introduction

Cloud computing is a framework that controls a huge number of data, and records to contribute to the intensive service. Mainly, the cloud environment is operated on a data center which is a network consisting of a huge number of physical machines whether they are heterogeneous or homogenous. In a data center, the system owner has the primary goal to reduce the overall system cost such as energy consumption of the system. Furthermore, in this platform physical machines are virtualized. These virtualized machines are combined with network virtualization. So, users can access many Virtual Machines (VM) [1]. Today, cloud computing has extended its service to data-intensive, computing-intensive, and network-intensive on distributed platforms, such as, Hadoop Distributed File System (HDFS) and Map-Reduce (MR) paradigm. Cloud computing is an independent system to control the big load of an application data, which may be data-intensive or computing-intensive as well as network-intensive applications through the proceeding of scheduling. These intensive functions can also be expanded by customer demands [2]. Network-intensive application is used to display a large and increasing number of functions in cloud framework. The functions are intermittently subjected to bottlenecks in communication speed and relationship among machines on which they are displayed. Network-intensive has the idea of executing numerous categories of services (i.e., LAN, SAN, or IPC) on an exclusive Ethernet-based network. Ethernet is low-cost accessible, and comparatively easy to use [3].

In the current age, large investments have been made in enormous data centers supporting cloud services, by companies [5] such as Facebook, eBay, Microsoft, Yahoo!, and Google. Further, [5] considers different components to quantify data center housing costs, e.g., infrastructures (Power, Distribution, and cooling), servers (CPU, memory, storage systems), network links, transit, equipment, and Power draw (Electrical utility costs). In recent years, networks in cloud computing have also been a source of study for the marketplace [6]. The core design concern for data center operators is network infrastructure. It represents a significant portion of the initial investment and does not directly contribute to future earnings. Therefore, a key driver for maximizing data center profits by reducing network infrastructure costs.

Cloud computing is a recently popular and helpful field in distributed computing. It helps in scheduling users' requests, i.e., jobs in a good way. The scheduler is required to manage equivalence between Quality of Services (QoS) and jobs in order to achieve high performance of a system [4]. Scheduling is a highest problem in cloud computing because cloud server has to facilitate more than one customer in cloud computing system and provide a service for the system owner to achieve the goal of energy efficiency. The major concept of scheduling is to increase system needs and decrease alter time of tasks. Most of the existing job scheduling techniques do not meet the required standards and requirements. Therefore, an efficient scheduling technique has become an important problem to be solved in cloud computing. So, this study provides a complete analysis of classic scheduling techniques, specifically for handling data-intensive workloads to see the effectiveness of energy and network cost of the system. The workload is selected from a real-world data center. Moreover, this study provides advantages and disadvantages of several classical heuristics-based job-scheduling techniques that consider the time and cost of data transfer from multiple sources, as this study is useful for selecting appropriate scheduling techniques for an appropriate environment. For the evaluation of scheduling strategies, we consider the following performance metrics, (a) network cost in cluster (NCC), (b) Network Cost in Racks (NCR), (c) Overall energy consumption (mJoul), (d) Overall running times, (e) Cluster-based Running Time (RTC) and (f) cluster-based energy consumption (mJoul)

In the organization of this paper, Section 2 shows a literature review; Section 3 explains the data center information; Section 4 shows research methodology, and Section 5 explains the results and discussion.

2. Literature review

Cloud data centers have geographically end-users and are distributed across the globe in a large-scale cloud computing infrastructure. The main difficulty for cloud data centers is how to efficiently and correctly process service in the millions of requests that come in from end-users on a regular basis, these apps and services are accessible via the internet. The related research is reviewed and depicted in Table I.

Table 1. Summary of state-of-the-art of Scheduling techniques extracted from literature

Sr. No	Literature Review		
	Research Applications	Research Techniques	References
1	Scientific Application, Workloads: CIWs, DIWs	Improved Particle Swarm Optimization	[10]
2	Scientific Application, Workloads: CIWs	Resource allocation, load balancing	[11]
3	Scientific Application, Workloads: CIWs	Classical-Job Scheduling-(FCFS, SJF, LJF)	[8]
4	Scientific Application, Workloads: CIWs	Linear Programming, Combinatorial and Stochastic	[12]
5	Scientific Application, Workloads: CIWs, DIWs	Resource Scheduling Algorithm (RSAs)	[13]
6	Scientific Application, Workloads: CIWs	Classical-Job Scheduling-(FCFS, SJF-FF, LJF-FF,MinET-FF,MaxET-FF)	[8] [9]
7	Scientific Application, Workloads: CIWs, DIWs	Goal Programming, Game Theory	[14]
8	Scientific Application, Workloads: CIWs	Thermal Power Technique	[15]

The author in [7] focuses on the characterization of datacenter workload for optimization. The same study determined the overall data center load efficiency and energy efficiency. DVFS and DPM techniques have been used for energy efficiency. Simple statistic methods are used to characterize the workload. However, the behavior of jobs and physical machines i.e., nodes have been investigated before and after the scheduling process [8]. Moreover, the same research study discovers distinctive features in the workload as a result of the characterization as follows: (i) the majority of jobs require a single CPU for execution; (ii) the leftover jobs exact an even number of CPUs for execution; (iii) half of all jobs run for less than an hour; (iv) half run for more than an hour.

In the same study, authors [8] present the issues scheduling of Virtual Machine (VM) in an Infrastructure as a Service (IaaS) cloud environment to decrease operating cost and to focus on the system meets the Quality of Service (QoS) factors. The research studied traditional scheduling schemes combined with power management technology namely DVFS, i.e., dynamic voltage and frequency scaling. To explore the environment a strategy based on the FCFS two policies based on Scale namely Minimum Work First (SMJF) and SHortest Job First (SHJF) and spot replacement AGgressive BackFill priority (AGBF) strategy have been considered. The study has identified the advantages and disadvantages of the investigated scheduling strategies for virtual machines and provided suggestions to choose the solution that best suits the environment. Moreover, researchers have used a real-world HPC workload collected from a production data center to run simulations.

Studies [9, 24] provide a comparison of job scheduling in large-scale parallel systems as follows: (i) reduce queue time, response time, and energy consumption; (ii) maximize the overall use of the system. Additionally, the study examines the behavior of thirteen different work scheduling policies, i.e., priority-based first-fit backfilling and window-based policies. All of the policies have been extensively simulated and their performance metrics used in comparisons. For the simulation, an actual data center workload consisting of 22,385 jobs has been employed in the experiment depending on their performances. In addition, this paper also gives a detailed workload characterization to optimize system performance and the design of the scheduler. The most important aspects of the workload is (a) wide (b) narrow (c) long and (d) short jobs are characterized by in-depth investigation of the situation performance of the scheduler. This research focuses on the advantages considering the benefits and drawbacks of various job scheduling policies in order to select a suitable employment scheduling policy in a given situation scenario.

The data-intensive applications involving the analysis of large data sets are becoming increasingly important as many fields of science and business face thousands of data growth. The explosive growth of data is mainly due to the internet, smart cities, and social networks. Terabytes to petabytes of data are stored in data-intensive systems. In order to conduct sophisticated queries and deliver fast results such systems demand a lot of storage as well as a lot of computing power. Furthermore, the velocity at which this data is generated creates significant storage connecting and processing issues. Users take benefit from the abstraction of high availability, usability, and efficiency provided by a data-intensive cloud.

On the other hand, data-intensive workloads impose essentially no burden on the computer servers but necessitate large data transfer. Data-intensive workloads are designed to represent applications such as video file sharing in which a simple user request transforms into a video streaming process. As a result, the data center interconnection network rather than processing capability becomes a bottleneck for data-intensive workloads. For any working environment, the primary task is scheduling and its actions are ordered by the processor. As there is no way to properly designate resources to ensure maximum skills in a strict and practical way and since, cloud services require a high level of control and resource management, effective scheduling is important for managing jobs and tasks scale and because the management system executes an essential role. Scheduling is used in cloud computing in order to achieve high performance and optimal system throughput. The speed, efficiency, and optimal use of resources are largely determined by the type of schedule selected for the cloud computing environments. The various scheduling criteria are maximum CPU usage and minimum throughput [22, 23].

3. Datacentre and research methodology

Cloud data centers are highly multiplexed shared environments that allow several tenant's VMs and processes to co-exist in the same cluster to achieve cost efficiencies in on-demand scaling. These applications are largely disorganized and mutually trustworthy because they come from unrelated clients [16]. In this paper, we use the important notations, and their descriptions are shown in Table II.

Table 2. Nomenclature

Notation	Description
VM	Virtual Memory
QoS	Quality of Services
NCC	Network Cost of Clusters
NCR	Network Cost of Racks
RTC	Overall Running Time of Cluster
OEC	Overall Energy Consumtions
ORT	Overall Running Time
DVFS	Dynamic Voltage and Frequency Scale
DPM	Dynamic Power Management
DCN	DataCenter Networks
FCFS	First Come First Serve
SJF	Shortage Job First
LJF	Largest Job First
Min-Min (MinMI)	Minimum Job with Minimum Execution / Millions of Instructions
ABF	Aggressive Backfilling
FF	First Fit
NC	Network Cost
ECC	Energy Consumption of Cluster

3.1. Datacenter architecture

Switching infrastructure of two or three-tier data centers is generally used in traditional Data Center Networks (DCNs). The core, aggregation, and access layers are the three most common layers of the data center. The three-tiered data center network DCN architecture is well-defined in [17].

3.2. Job scheduling algorithms

The job scheduling algorithm can be preemptive or non-preemptive. In the non-preemptive scheduling algorithm, no force can prevent the execution of jobs on the other hand due to many factors [18]. Job execution may stop in the preemptive scheduling algorithm. There are many popular scheduling algorithms in the cloud and the focus of this study is on a set of five job scheduling algorithms incorporating with First Fit (FF) strategy that is discussed in this section.

3.2.1. First Come First Serve (FCFS)

FCFS algorithm is very simple and fast. The main purpose of this algorithm is to execute a job that is placed in a queue. FCFS works like the first come name indicates the first job to be served and first fed and executed in the CPU. The drawback of FCFS is one of the non-preemptive and slow scheduling schemes [19], however, the fairness in job placement is its advantage.

3.2.2. Shortest Job First (SJF)

SJF does the job with the shortest job first manner. In addition, the job with the largest size waits for execution. The queue takes a long time to complete the process. This technique of scheduling can be either preemptive or non-preemptive [20].

3.2.3. Largest Job First (LJF)

LJF is a scheduling algorithm based on the size of the job. The processes are sorted into the ready queue by their job size, which is listed in descending order. This algorithm is based on the fact that the job with the largest size is handled first, as the name implies [15].

3.2.4. Min-Min Algorithm

Min-Min starts with a set of unassigned tasks at first; it calculates the minimum execution time for all tasks on all resources and then selects the minimum value among these minimum durations of all tasks on a resource. Then, schedule the task on the resources spend as little time as possible on the resources, and update the time available for the resources for all other tasks. In other words, this algorithm is based on the fact that the job with the minimum millions of instructions is handled first [17].

3.2.5. Aggressive backfilling

Aggressive Backfilling is a scheduling improvement that enables a scheduler to utilize resources more effectively by running operations out of sequence. This algorithm is an extended version of FCFS, in which fairness is not affected. It uses a variety of criteria to prioritize the jobs in the queue before sorting them into a list with the highest priority listed first. Until it reaches a job it cannot start, it steps through the priority list starting each job one at a time. Because every work and reservation has a start time and a wall clock limit, it is possible to estimate when every job in the queue will be finished [18].

3.3 Cloud-based simulation framework

In the cloud-based framework, there are PCs from PC1 to PCn as illustrated in Fig. 1. The request for VM initializes the process, and then the request is forwarded into task queue, the list of requests stored separately. The task sent to the task manager for the processing of CPU intensive, memory intensive, I/O intensive, communication intensive. There are several VM machines, and ten groups of machines have assigned a cluster and multiple clusters have been created for task distribution by which the minimum energy will be consumed and produce the efficiency of the system. Moreover, the each VM processed to the host manager, which forwards to the suitable host.

Table 3 indicates the simulation configuration of network cost effective job scheduling strategies. An event-based simulator is designed to simulate the study of energy cost and network cost effective job scheduling strategies of cloud-virtualized environment. Moreover, the study tested job-scheduling strategies with the real-world workload in simulation. On the dedicated HPC cloud, CPU is divided into classes based on network cost computing. These classes are categorized into three groups as in previous studies [12, 24]. We further design these groups of clusters via different racks in a cluster. As each cluster has nine racks, holding machines belong to different classes.

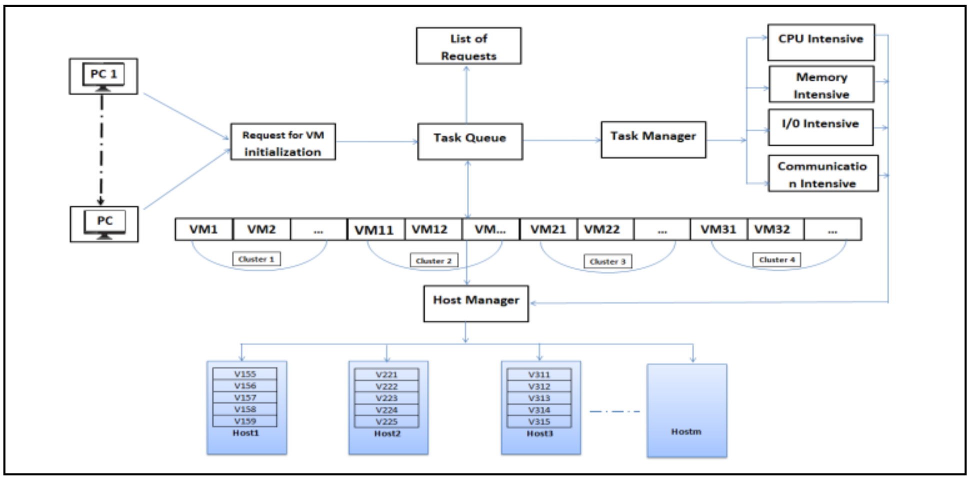

Fig. 1. Cloud-based framework

Table 3. Simulation configuration

Class	Cluster-I			Cluster-II			Cluster-III			
	Racks	CPU	Total CPU	Racks	CPU	Total CPU	Racks	CPU	Total CPU	Total CPU
Class-I (3.0 GHz)	4	40	160	4	40	160	4	40	160	480
Class-II (3.3 GHz)	3	80	240	3	80	249	2	80	200	680
							1	40		
Class-III (3.3 GHz)	1	160	240	1	160	240	1	160	240	720
	1	80		1	80		1	80		
			640			640			600	1880

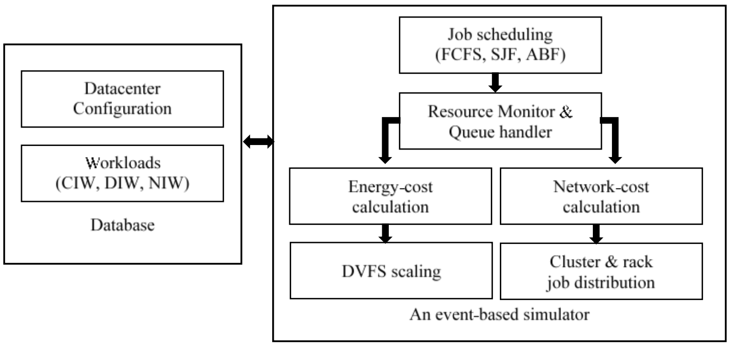

Fig. 2. Complete block-wise steps of the studied simulation process

Furthermore, we provide the complete steps of our simulation in Fig. 2. Database holds the dataset of datacenter system and workloads. The workload may be compute-intensive, data-intensive, or network-intensive workload as an input. An event-based simulator connects with database for resource monitoring and queuing the satisfied resources and workloads by applying the particular job scheduling strategy, such as FCFS, SJF, ABF, etc., DVFS scaling process for minimizing energy-cost and job distribution on cluster and racks for minimizing network-cost are eventually done by the simulator. The interval (i.e., 10 minutes) is set to eventually execute the job scheduling strategies for updating the datasets, monitoring resources and handling the queue.

3.4. Performance metrics

In this study, for performance evaluation, we select the following six performance metrics in order to understand the energy and network cost-effectiveness of job scheduling strategies in especially heterogeneous environment. (a) Network Cost in Cluster (NCC) is considered as primary performance metrics, in which, the cost is calculated if the tasks in a job run on more than one cluster. For example, a job has two tasks and both tasks get chance to execute on two different clusters (i.e., Cluster-A and Cluster-B) a specific cost value would be added as network cost. In this study, we assume 0.1 value as network cost [25]. (b) Network Cost in Racks (NCR) is the similar to NCC performance metric, however NCR is calculating the cost value (i.e., 0.1) if tasks in a job get chance to execute on more than one racks on a cluster. (c) Overall Energy Consumption (OEC) is considered as a second important performance factor of the job scheduling strategies, in which, we calculate the energy consumed by each task in a job then it is accumulated from all jobs executed on the system, as defined in [15, 24]. (d) Overall Running Time (ORT) is used to calculate the running time of each task in a job then it is accumulated as a total running time from all jobs running on the system [9]. (e) Cluster based Energy Consumption (ECC) is used to calculate the energy consumption of each cluster. In other words, the energy consumption of each task in all jobs executing on a cluster is calculated to see the consumption of each cluster. (f) Cluster based Running Time (RTC) is similarly calculating the running time of each task in all jobs being executed on a cluster.

4. Results and discussions

This part describes the findings and explanations of the energy cost and network cost effective job scheduling strategies for cloud-virtualized environment. As the primary goal of this study is to reduce network costs and energy cost, we first assess how well the different network cost effective job scheduling strategies perform in terms of Network Cost of Clusters (NCC). The *y*-axis of the following graph (i.e., Fig. 3) shows the NCC in aspect to cluster system, while the *x*-axis represents the network cost effective job scheduling strategies (ABF-NC, LJF-FF-NC, FCFS-FF-NC, SJF-FF-NC, and MinMI-FF-NC) to reduce network cost by using clusters with on-demand power management scheme. The figure clearly demonstrates that the FCFS-FF-NC and MinMI-FF-NC network cost effective job scheduling strategies give the best performance, while the LJF-FF-NC network cost effective job scheduling strategy gives the worst performance in aspect to Network Cost of Clusters (NCC).

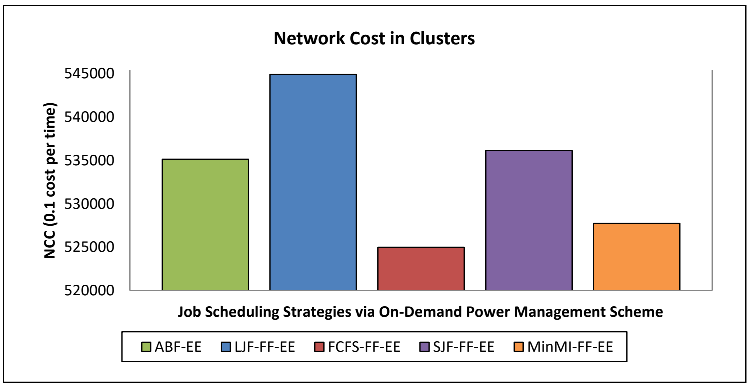

Fig. 3. Job scheduling strategies of Network Cost in Cluster (NCC)

Secondly, we evaluate how well the different network cost effective job scheduling strategies perform in terms of Network Cost of Racks (NCR). The *y*-axis of the following graph (i.e., Fig. 4) shows the NCR, while the *x*-axis represents the network cost effective job scheduling strategies (ABF-NC, LJF-FF-NC, FCFS-FF-NC, SJF-FF-NC, and MinMI-FF-NC) to reduce network cost by using racks with on-demand power management scheme. The figure clearly demonstrates that the FCFS-FF-NC and ABF-NC network cost effective job scheduling strategies give the best performance, while the LJF-FF-NC and SJF-FF-NC network cost effective job scheduling strategies give the worst performance in aspect to Network Cost of Racks (NCR).

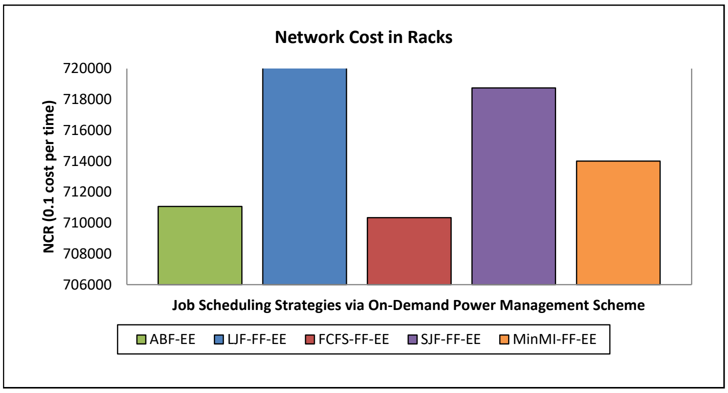

Fig. 4. Job scheduling strategies of Network Cost in Racks (NCR)

Thirdly, we evaluate how well the different network cost effective job scheduling strategies perform in terms of Overall Energy Consumption (OEC). The *y*-axis of the following graph (i.e., Fig. 5) shows the OEC, while the *x*-axis represents the network cost effective job scheduling strategies (ABF-NC, LJF-FF-NC, FCFS-FF-NC, SJF-FF-NC, and MinMI-FF-NC) to reduce energy cost by using clusters and racks with on-demand power management scheme. The figure clearly demonstrates that the FCFS-FF-NC and SJF-FF-NC network cost effective job scheduling strategies give the best performance, while the ABF-NC network cost effective job scheduling strategy gives the worst performance in aspect to Overall Energy Consumption (OEC).

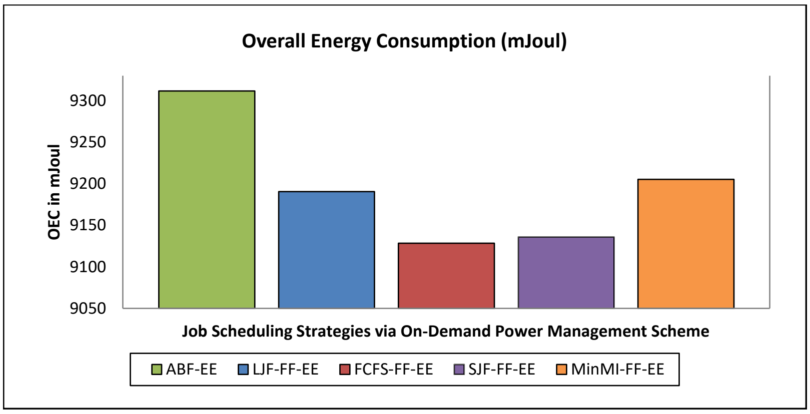

Fig. 5. Overall Energy Consumption (OEC) (mjoul) by job scheduling strategies

Fourthly, we evaluate how well the different network cost effective job scheduling strategies perform in terms of Overall Running Time (ORT). The *y*-axis of the following graph (i.e., Fig. 6) shows the ORT in particular seconds time unit, while the *x*-axis represents the network cost effective job scheduling strategies (ABF-NC, LJF-FF-NC, FCFS-FF-NC, SJF-FF-NC, and MinMI-FF-NC) to reduce running time cost by using clusters and racks with on-demand power management scheme. The figure clearly demonstrates that the ABF-NC and LJF-FF-NC network cost effective job scheduling strategies give the best performance, while the SJF-FF-NC and FCFS-FF-NC network cost effective job scheduling strategy gives the worst performance in aspect to Overall Running Time (ORT).

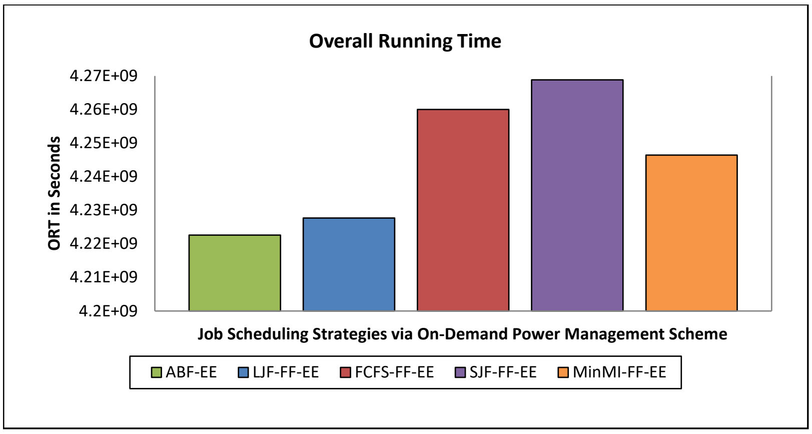

Fig. 6. Overall Running Times (ORT) of job scheduling strategies

Additionally, we evaluate how well the different network cost effective job scheduling strategies perform in terms of Energy Consumption in each Cluster (ECC). The *y*-axis of the following graph (i.e., Fig. 7) shows the ECC, while the *x*-axis represents the network cost effective job scheduling strategies (ABF-NC, LJF-FF-NC, FCFS-FF-NC, SJF-FF-NC, and MinMI-FF-NC) to reduce energy consumption cost by using clusters and racks with on-demand power management scheme. The figure clearly demonstrates that the Cluster-C gives the best performance, while the Cluster-A and Cluster-B give the worst performance in aspect to Energy Consumption in each Cluster (ECC).

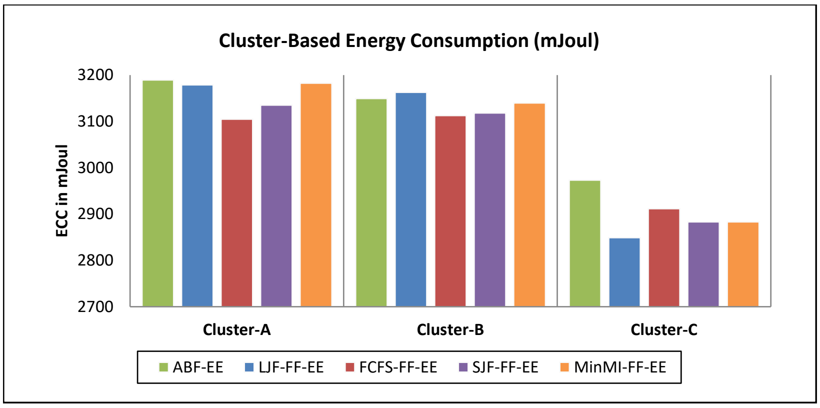

Fig. 7. Job scheduling strategies of Cluster-based Energy Consumption (mjoul) (ECC)

Furthermore, we provide the analysis of cluster-based energy consumption in aspect to percentage of energy consumed by each cluster shown as in the Fig 8 The *x*-axis shows the percentage of energy consumption in mJoul, while *y*-axis shows the studied job scheduling strategies placing the jobs on three clusters such as Cluster-A, Cluster-B, and Cluster-C. From the figures, we can easily see that all of the studied job scheduling strategies similarly are generating energy on the clusters. Such as Cluster-C in all job-scheduling strategies (i.e., lightest bar in each color group in the figure of all job scheduling strategies) consumed less energy compared against other clusters.

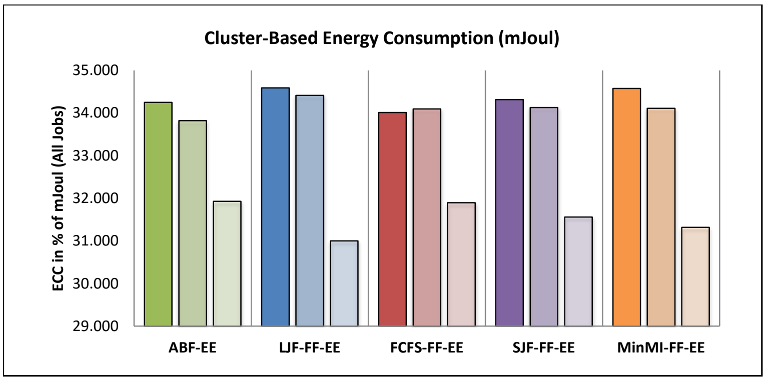

Fig. 8. Job scheduling strategies of Cluster-based Energy Consumption (mjoul) (ECC) in percentage

Additionally, we evaluate how well the different network cost effective job scheduling strategies perform in terms of Running Time in each Cluster (RTC). The *y*-axis of the following graph (i.e., Fig. 9) shows the RTC in particular seconds time unit, while the *x*-axis represents the network cost effective job scheduling strategies (ABF-NC, LJF-FF-NC, FCFS-FF-NC, SJF-FF-NC, and MinMI-FF-NC) to reduce energy consumption cost by using clusters and racks with on-demand power management scheme. The figure clearly demonstrates that the Cluster-C gives the best performance, while the Cluster-A and Cluster-B give the worst performance in aspect to Running Time in each Cluster (RTC).

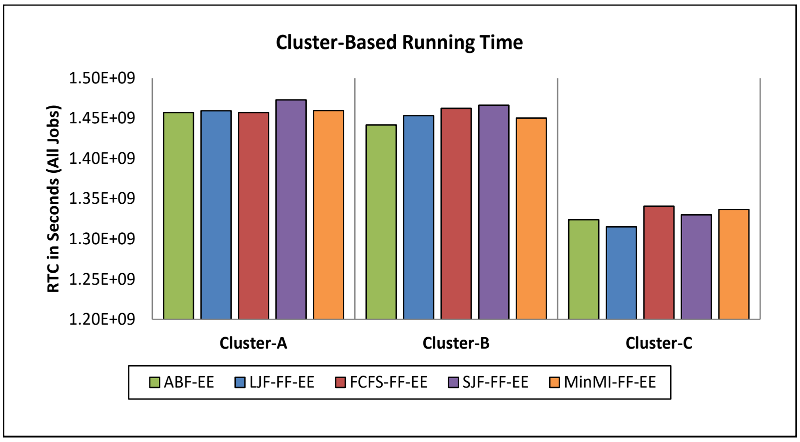

Fig. 9. Job scheduling strategies of Cluster-based Running Time (RTC)

Moreover, similar to Fig. 8, we also provide the analysis of cluster-based overall running time in aspect to percentage of running time by each cluster shown as in the Fig 10. The *x*-axis shows the percentage of running time, while *y*-axis shows the studied job scheduling strategies placing the jobs on three clusters such as Cluster-A, Cluster-B, and Cluster-C. From the figures, we can easily see that all of the studied job scheduling strategies similarly tacking time of the jobs on the clusters. Such as Cluster-C in all job-scheduling strategies (i.e., lightest bar in each color group in the figure of all job scheduling strategies) took less running time compared against other clusters.

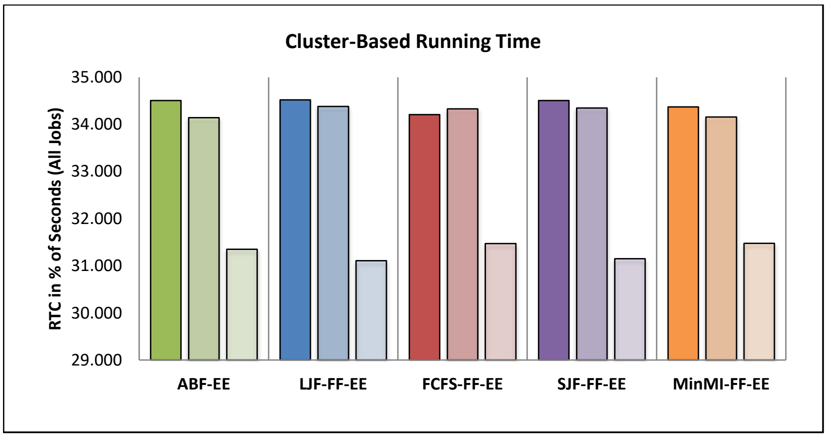

Fig. 10. Job scheduling strategies of Cluster-based Running Time (RTC) in percentage

Conclusion

This study represent a comparative analysis of energy-efficient VM job scheduling strategies for HPC workloads in a virtualized environment. In order to reduce the network cost effective job scheduling strategies in cloud virtualized environment, this study looks at VM job scheduling strategies for HPC workloads in a cloud virtualized environment and examines a five common timeliness: SHJF, LJF, FCFS, AgBF and MinMI. Our investigation shows that a single job scheduling strategy is insufficient for managing resources in virtualized systems in a network cost and energy-efficient manner.

***Acknowledgment*:** Afia's work is partly supported by her MPhil degree program in University of Sindh Jamshoro.

References

1. Thirumala Rao, B. Scheduling Data Intensive Workloads through Virtualization on MapReduce Based Clouds. – Int. J. Distrib. Parallel Syst., Vol. **3**, 2012, No 4, pp. 99-110. DOI: 10.5121/ijdps.2012.3411.
2. Mansouri, N. A New Job Scheduling in Data Grid Environment Based on Data and Computational Resource Availability. – AmirKabir Int. Sci. Res., Vol. **47**, 2015, No 1, pp. 41-53.
3. Kliazovich, D., P. Bouvry, S. U. Khan. DENS: Data Center Energy-Efficient Network-Aware Scheduling. – In: Proc. of 2010 IEEE/ACM Int. Conf. Green Comput. Commun. GreenCom 2010, No August 2010, pp. 69-75. DOI: 10.1109/GreenCom-CPSCom.2010.31.
4. McClatchey, R., A. Anjum, H. Stockinger, A. Ali, I. Willers, M. Thomas. Scheduling in Data Intensive and Network Aware (DIANA) Grid Environments. – J. Grid Comput., Vol. **5**, 2007, No 1, pp. 43-64.
5. Greenberg, A., J. Hamilton, D. A. Maltz, P. Patel. The Cost of a Cloud. – ACM SIGCOMM Comput. Commun. Rev., Vol. **39**, 2008, No 1, pp. 68-73. DOI: 10.1145/1496091.1496103.
6. Li, X., Y. Li, T. Liu, J. Qiu, F. Wang. The Method and Tool of Cost Analysis for Cloud Computing. – In: Proc. of CLOUD 2009 – 2009 IEEE Int. Conf. Cloud Comput., 2009, pp. 93-100. DOI: 10.1109/CLOUD.2009.84.
7. Mayyad Jaber. Architecture de Système d'information distribué pour la gestion de la chaîne logistique : Une approche orientée servic. Phd Thesys. Lion, France, 2009 (In French).
8. Chandio, A. A., M. S. Korejo, I. A. Korejo, M. S. Chandio. To Investigate Classical Scheduling Schemes with Power Management in IaaS Cloud Environment for HPC Workloads. – In: Proc. of IEEE Student Conf. Res. Dev. Inspiring Technol. Humanit. SCOReD 2017, Vol. **2018-January**, 2018, pp. 121-126. DOI: 10.1109/SCORED.2017.8305409.
9. Chandio, A. A., et al. A Comparative Study on Resource Allocation and Energy Efficient Job Scheduling Strategies in Large-Scale Parallel Computing Systems. – Cluster Comput., Vol. **17**, 2014, No 4, pp. 1349-1367. DOI: 10.1007/s10586-014-0384-x.
10. Sana, M. U., Z. Li. Efficiency Aware Scheduling Techniques in Cloud Computing: A Descriptive Literature Review. – PeerJ Comput. Sci., Vol. **7**, 2021, pp. 1-37. DOI: 10.7717/PEERJ-CS.509.
11. Leontiou, N., D. Dechouniotis, S. Denazis, S. Papavassiliou. A Hierarchical Control Framework of Load Balancing and Resource Allocation of Cloud Computing Services. – Comput. Electr. Eng., Vol. **67**, 2018, No March, pp. 235-251. DOI: 10.1016/j.compeleceng.2018.03.035.
12. Kliazovich, D., J. E. Pecero, A. Tchernykh, P. Bouvry, S. U. Khan, A. Y. Zomaya. CA-DAG: Modeling Communication-Aware Applications for Scheduling in Cloud Computing. – Journal of Grid Computing, Vol. **14**, 2016, pp. 23-39.

13. S i n g h, S., I. C h a n a. Resource Provisioning and Scheduling in Clouds: QoS Perspective. – Springer US, Vol. **72**, 2016, No 3.
14. S. U. K., D. K l i a z o v i c h, P. B o u v r y. Simulation and Performance Analysis of Data Intensive and Workload Intensive Cloud Computing Data Centers. – Opt. Interconnects Futur. Data Cent. Net., 2013.
15. W a n g, L., S. U. K h a n, J. D a y a l. Thermal Aware Workload Placement with Task-Temperature Profiles in a Data Center. – The Journal of Supercomputing, Vol. **61**, 2012, pp. 780-803.
16. H e a d q u a r t e r s, A. Cisco Data Center Infrastructure 2.5 Design Guide. – Cisco, No 6387, 2007.
17. K a s h i f, B., S. U. K h a n, J. K o l o d z i e j, L. Z h a n g, K. H a y a t, S. A h m a d M a d a n i, N. M i n-A l l a h, L. W a n g, D. C h e n. A Comparative Study Of Data Center Network Architectures. – In ECMS, 2012, pp. 526-532.
18. S u t h a, K., G. M. K. N a w a z. Research Perspective of Job Scheduling in Cloud Computing. – In: Proc. of 8th International Conference on Advanced Computing (ICoAC'2016), 2017, pp. 61-66.
19. A b d u l l a h i, I. Process Scheduling In Longest Job First (LJF) Algorithm. A Proposed Framework for Starvation Problem.
20. S i n d h u, S. Task Scheduling in Cloud Computing. – Comp. Eng. Techn, 2015.
21. E t m i n a n i, K. A Min-Min Max-Min Selective Algorihtm for Grid Task Scheduling. 2007.
22. I l a n k u m a r a n, A., S. J. N a r a y a n a n. An Energy-Aware QoS Load Balance Scheduling Using Hybrid GAACO Algorithm for Cloud. – Cybernetics and Information Technologies, Vol. **23**, 2023, No 1, pp. 161-177.
23. B h a r g a v i, K., S. G. S h i v a. Uncertainty Aware T2SS Based Dyna-Q-Learning Framework for Task Scheduling in Grid Computing. – Cybernetics and Information Technologies, Vol. **22**, 2022, No 3, pp. 48-67.
24. C h a n d i o, A. A., N. T z i r i t a s, M. S. C h a n d i o, C.-Z. X u. Energy Efficient VM Scheduling Strategies for HPC Workloads in Cloud Data Centers. - Sustainable Computing: Informatics and Systems, Vol. **24**, 2019, 100352.
25. T z i r i t a s, N., C.-Z. X u, T. L o u k o p o u l o s, S. U. K h a n, Z. Y u. Application-Aware Workload Consolidation to Minimize Both Energy Consumption and Network Load in Cloud Environments. – In: Proc. of 2013 42nd IEEE International Conference on Parallel Processing, 2013, pp. 449-457.

Received: 05.06.2023; Second Version: 09.07.2023; Accepted: 24.07.2023 (fast track)