iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XII **Month of publication:** December 2021

DOI: https://doi.org/10.22214/ijraset.2021.39394

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XII Dec 2021- Available at www.ijraset.com

A Survey on Resource Management Techniques in Cloud Computing

Kapil Tarey[1], Vivek Shrivastava[2]
[1, 2]*International Institute of Professional Studies, DAVV, Indore (MP)*

Abstract: ***Cloud computing refers to a computer environment in which traditional software systems, installations, and licensing concerns are replaced with comprehensive on demand," pay as you need" internet based services. In this scenario, many cloud customers can request multiple cloud resources at the same time. As a result, there should be a plan in place to ensure that resources must be prepared for the needy customer in proficient way in order complete their needs. In cloud computing systems, resource management is a critical and difficult issue. It must meet numerous service quality requirements and, as a result, reduce SLA violations. This paper survey different resource management technique for cloud infrastructures.***
Keywords: Cloud, Resource management and techniques.

I. INTRODUCTION

Cloud computing is a technology in which numerous data applications, services, and infrastructures are maintained over the internet and central distant datacenters in order to supply necessities to customers. It enables users to utilize and deploy services and resources without the need for installation, and users may subscribe to different resources or access their files from any device with internet connection [2]. This model offers excellent computing facilities as well as ubiquitous, simple, on-demand resource access to its members. It configures a common set of computer resources, and these resources, including as services, networks, storage, applications, bandwidth, and infrastructures, may be allocated fast with right management techniques [5, 6]. The most fundamental element of cloud computing is that, rather of having personnel physical facilities, its subscribers and users may hire or rent resources from any cloud service provider [8]. Users must therefore pay for the services or resources that they have consumed.

Cloud computing enables customers to get resources in a flexible and dynamic manner. The most key challenge in resource management is defining the proper number of needed resources for allocation in order to meet users' demand while simultaneously reducing costs from the users' perspective and making the optimum use of resource usage from the service providers' viewpoint. Resource management is the collective effort of interprocess communication such as resource collection, distribution, and on-demand administration to ensure the system's functionality. Consideration of Service Level Agreement (SLA) during resource management is critical. SLAs are the elementary deal among cloud users and cloud service providers [14]. As Quality of Service, SLA also ensures several aspects such as performance, accessibility, trustworthiness, response time, security, and energy savings (QoS). Allocation of resources may be done in both a static and dynamic way, allowing resources to be used more efficiently. This allocation should not breach any SLAs and should also adhere to the QoS requirements. High energy use, excess or insufficient resource supply must not disrupt resource management [15]. The primary goal of resource management may differ for customers and service providers [17, 19]. The goal for users is to reduce operating and maintenance costs without owning physical systems, whereas the goal for cloud service providers is to earn profits through resource allocation and management. To achieve these goals, cloud customers must advise cloud service providers about whether resources will be assigned statically or dynamically [20]. This information will be intake for CSPs to assure the quantity and availability of resources. Because resource allocation for various applications may change, the cloud service provider will also be able to calculate or manage resource requirements based on application.

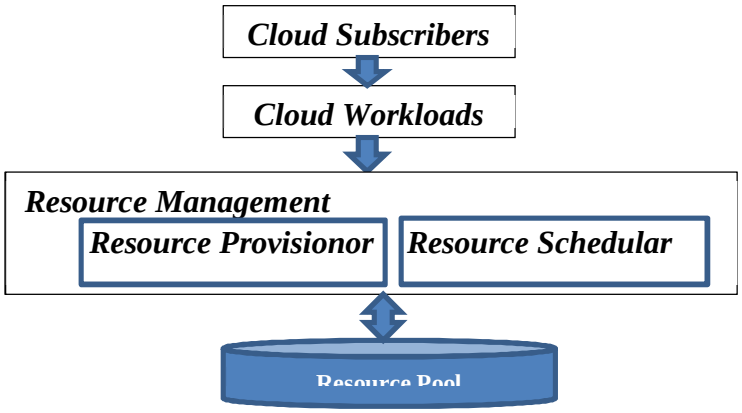


Fig. 1 Resource Management in Cloud

©IJRASET: All Rights are Reserved

Resource management incudes transfer of jobs from their start to their execution and end. In the cloud, resource management works in steps. In first step, resource allocation identifies adequate resources for precise jobs on the basis of QoS requirements delivered by consumers, while resource planning plots and implements consumer jobs based on designated resources. Fig. 1 describes this process.

Remaining structure of paper contains: Section II that deals with resource allocation. Section III discusses resource allocation parameters. Section IV offers several resource management options. Section V compares several resource management approaches. Section VI contains the conclusion.

II. VARIOUS RESOURCE ALLOCATION MECHANISMS

Resource allocation may be categorized as follows based on the demands of the users and the type of requests/claims:

1) *Fixed/Static Allocation:* Fixed/static allocation is best suited for effective allocation of requests/claims with predictable and typically static requests/workloads. The consumer anticipates their resource requirements and may enter into agreements with cloud service providers [22]. The cloud service provider then prepares the necessary resources ahead of the start of the service. Billing or charges for services are due in the form of set fees or on a monthly basis.
2) *Dynamic/Active Allocation:* Dynamic/active allocation is useful when requests/claims are unexpected and may be altered, or when requests/workloads vary. In general, virtual devices may be quickly transferred by constructing new nodes in the cloud [23]. This cloud service provider may give more resources as needed and withdraw them when they are no longer required. For billing purposes, it mostly employs a "pay-as-you-go" approach.
3) *User self-allocation:* It is similar to self-service in that consumers obtain resources from any cloud service provider via any web-based service/website. For this, the consumer must sign in with an account on the website. They can order or reserve desired resources/services for the appropriate period of time after entering into the website [24]. The consumer will be charged for that resource/service.

III. PARAMETERS FOR RESOURCE ALLOCATION

A. In order to provide benefits to cloud users, costs must be kept to a minimum.
B. Profit growth should be achieved by delivering benefits to the cloud service provider.
C. In addition to low costs and large profits, the SLA parameter should be examined in such a way that its violation is minimized.
D. Power or energy consumption may be excessive during resource management and provisioning according to VM placement and migration. As a result, guarantee of decreased power consumption or low energy absorption must be taken into account.
E. If a VM or other service fails, the allocation should not be affected and the service should continue to be delivered.
F. When completing a job, the resource provisioning approaches mentioned fundamentally take little time to reply.

IV. RESOURCE MANAGEMENT APPROACHES

Cloud resources may be used and managed more efficiently by using resource management and allocation. Datacenters employ a variety of resource management methodologies, including static and dynamic management. Each has its own set of advantages and disadvantages. Management techniques are used to enhance QoS parameters [1,3,5,13], reduce costs and optimise profits [17,21,24], improve response time [2], offer services even when there is a failure [11,14,15], and improve performance [21,24,25]. reduces SLA violations [20], effectively uses cloud assets [4,6,9,10,16,19,20,22,24], and reduces power consumption [7,21].

The overall process of resource management uses elementary taxonomy of resources in cloud as shown in Fig. 2.

According to previous research, minimizing execution time is a hot topic in cloud resource management. Management of appropriate resources to jobs is always a challenging task, and based on QoS requirements, recognizing the finest job and its resource combination is an significant research in the cloud.

When a workload is supplied to a resource management algorithm, it accesses the resource information, which comprises information related to all resources available at datacenters, and gathers scheduling result based on the job needs[16]. The resource management output is returned to the cloud customer via the algorithm. If the required resources are not available in line with the QoS requirement, the algorithm wishes to resubmit the workload as a SLA document with amended QoS criteria [18].

Various issues like resource distribution, ambiguity, and heterogeneity, are not addressed by classic resource management techniques. As a result, we can increase the effectiveness of services and applications of cloud by addressing these cloud environment aspects.

©IJRASET: All Rights are Reserved

During management of resource their plotting, implementation and their observation can be performed. After resource management, comes resource management. To begin, the cloud consumer sends the task to be executed. Following that, jobs are mapped to appropriate resources depending on the quality of service criteria stated by the cloud customer. Quality of Service may include throughput, CPU, memory use, and so on are often evaluated for every customer in the cloud.

V. SURVEY AND COMPARATIVE STUDY OF RESOURCE MANAGEMENT TECHNIQUES (RMT)

Survey and comparison between RMTs is quite tough task due to multiple resource scheduling techniques and also there is lack of standards. Hence, evaluation and study of RMTs is noteworthy to discover the cost and performance effective resource scheduling algorithms. Following table shows the comparison of different RMTs found in literature studied for this paper.

Resource Management Technique	Application Used	Specific purpose of mechanism	Management Criteria	Tool Used	Challenges
Bargaining Based RMT	Blend of workflows	To increase user satisfaction	Bid Density	Green Cloud Simulator	Execution time is high
Compromised cost and Time based RMT	Identical Workloads	To forecast future expense	Budget and deadline	Cloudsim	Not considered heterogeneous workloads
Dynamics and Adaptive based RMT	Computing of intensive workloads	To eliminate running time and price	Revenue maximization	Amazon EC2	Lack of user satisfaction
Resource Management Technique	Application Used	Specific purpose of RMT	Management Criteria	Tool Used	Challenges
Energy based RMT	Identical Workloads	To reduce execution time	Execution time and Power consumption	Cloudsim	Performance degradation
Hybrid based RMT	Fork and Join	Minimize cost and meet deadline	Execution time	Cloud based simulator	Budget increased
Nature inspired and Bio Inspired RMT	Computation workload	Minimize makespan	Cost, makespan and degree of imbalance	Cloudsim	Time complexity increased
Deadline-driven RMT	Scientific workloads	Efficiently allocation of resources	Reduces execution time.	CloudAnalyst	Not for high data demanding applications
Active provisioning in many tenant RMT	Dynamic workloads	Service includes multiple tenants	Cloud based simulation	Aneka	No real world testing
Lightweight Approach RMT	Composite or random workloads	Flexibility and resource efficiency.	Elastic Application Container	Test bed	Not suitable for web applications
Failure-aware RMT	Hybrid workloads	Quality of service improved	Able to improve the users' QoS	Java based simulator	Not able to run real experiments
Resource Management Technique	Application Used	Specific purpose of RMT	Management Criteria	Tool Used	Challenges
Profit based RMT	Composite workloads	To fulfill SLA	Processing time and communication cost	Monte Carlo Simulator	Only support for single tier applications
Priority based RMT	Composite service applications	To reduce processing time and improved revenue	Priority and processing time	Java based simulator	Cost/SLA is not considered
Time based RMT	Deadline constrained workload	To meet deadline	Improve data transfer, computational cost and network bandwidth	Java based discrete time simulation	Performance degrade

©IJRASET: All Rights are Reserved

VI. CONCLUSION

The results of this research study have been examined in different of methods, including resource categorization and resource scheduling evolution. It is also difficult to determine the appropriate workload-to-resource mapping without an effective resource provisioning strategy.

The following facts can be deduced:

1) If resources are booked in advance, the cost of the given cloud service may be decreased.
2) A comparison and valuation of RMTs in the cloud can help to preference the scheduling algorithm based on the job's QoS essentials.
3) Allocating resources depending on workload type (homogeneous or heterogeneous) might increase resource usage.

The appropriate matching between workload and resource can greatly boost performance. It is difficult for providers to precisely determine the quantity of resources necessary for a particular task. We think that this survey will be useful to researchers of cloud resources scheduling.

REFERENCES

[1] Almuqati, M. T. (2021). Challenges in the cloud computing model of resource management issues such as service level agreements and services models. International Journal of Software Innovation, 9(3), 42–51. https://doi.org/10.4018/ijsi.2021070103
[2] Bittencourt, L. F., Senna, C. R., & Madeira, E. R. (2018). Scheduling service workflows for cost optimization in hybrid clouds. 2018 International Conference on Network and Service Management. https://doi.org/10.1109/cnsm.2010.5691241
[3] Budgen, D., & Brereton, P. (2020). Performing systematic literature reviews in software engineering. Proceedings of the 28th International Conference on Software Engineering. https://doi.org/10.1145/1134285.1134500
[4] Calheiros, R. N., & Buyya, R. (2018). Cost-effective provisioning and scheduling of deadline-constrained applications in hybrid clouds. Web Information Systems Engineering - WISE 2012, 171–184. https://doi.org/10.1007/978-3-642-35063-4_13
[5] Govindarajan, K., Kumar, V. S., & Somasundaram, T. S. (2017). A distributed cloud resource management framework for high-performance computing (HPC) applications. 2016 Eighth International Conference on Advanced Computing (ICoAC). https://doi.org/10.1109/icoac.2017.7951735
[6] Guan Le, Ke Xu, & Junde Song. (2017). Dynamic Resource Provisioning and scheduling with deadline constraint in Elastic Cloud. 2017 International Conference on Service Sciences (ICSS). https://doi.org/10.1109/icss.2017.18
[7] Hwang, J., & Wood, T. (2018). Adaptive Dynamic Priority Scheduling for virtual desktop infrastructures. 2012 IEEE 20th International Workshop on Quality of Service. https://doi.org/10.1109/iwqos.2012.6245988
[8] Lakkadwala, P., & Kanungo, P. (2018). Memory utilization techniques for cloud resource management in cloud computing environment: A survey. 2018 4th International Conference on Computing Communication and Automation (ICCCA). https://doi.org/10.1109/ccaa.2018.8777457
[9] Netjinda, N., Sirinaovakul, B., & Achalakul, T. (2016). Cost optimization in cloud provisioning using particle swarm optimization. 2016 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. https://doi.org/10.1109/ecticon.2016.6254298
[10] Nishant, K., Sharma, P., Krishna, V., Gupta, C., Singh, K. P., Nitin, & Rastogi, R. (2017). Load balancing of nodes in cloud using ant colony optimization. 2017 UKSim 14th International Conference on Computer Modelling and Simulation. https://doi.org/10.1109/uksim.2017.11
[11] V. A. (2013). A survey on various resource allocation policies in cloud computing environment. International Journal of Research in Engineering and Technology, 02(05), 760–763. https://doi.org/10.15623/ijret.2013.0205005
[12] Pandey, S., Wu, L., Guru, S. M., & Buyya, R. (2010). A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. 2010 24th IEEE International Conference on Advanced Information Networking and Applications. https://doi.org/10.1109/aina.2010.31
[13] Pawar, C. S., & Wagh, R. B. (2017). Priority based dynamic resource allocation in cloud computing with modified waiting queue. 2017 International Conference on Intelligent Systems and Signal Processing (ISSP). https://doi.org/10.1109/issp.2017.6526925
[14] Prodan, R., Wieczorek, M., & Fard, H. M. (2019). Double auction-based scheduling of scientific applications in distributed grid and Cloud Environments. Journal of Grid Computing, 9(4), 531–548. https://doi.org/10.1007/s10723-011-9196-x
[15] Rahman, M., Hassan, R., Ranjan, R., & Buyya, R. (2019). Adaptive workflow scheduling for Dynamic Grid and cloud computing environment. Concurrency and Computation: Practice and Experience, 25(13), 1816–1842. https://doi.org/10.1002/cpe.3003
[16] Raju, R., Babukarthik, R. G., Chandramohan, D., Dhavachelvan, P., & Vengattaraman, T. (2018). Minimizing the makespan using hybrid algorithm for cloud computing. 2018 3rd IEEE International Advance Computing Conference (IACC). https://doi.org/10.1109/iadcc.2018.6514356
[17] Salehi, M. A., Javadi, B., & Buyya, R. (2020). Performance analysis of preemption-aware scheduling in multi-cluster grid environments. Algorithms and Architectures for Parallel Processing, 419–432. https://doi.org/10.1007/978-3-642-24650-0_36
[18] Singh, S., & Chana, I. (2016). A survey on resource scheduling in Cloud computing: Issues and challenges. Journal of Grid Computing, 14(2), 217–264. https://doi.org/10.1007/s10723-015-9359-2
[19] Singh, S., & Chana, I. (2016). Energy Based Efficient Resource Scheduling: A step towards green computing. International Journal of Energy, Information and Communications, 5(2), 35–52. https://doi.org/10.14257/ijeic.2016.5.2.03
[20] Van den Bossche, R., Vanmechelen, K., & Broeckhove, J. (2018). Cost-efficient scheduling heuristics for deadline constrained workloads on hybrid clouds. 2018 IEEE Third International Conference on Cloud Computing Technology and Science. https://doi.org/10.1109/cloudcom.2011.50
[21] Verma, A., & Kaushal, S. (2019). Budget constrained priority based genetic algorithm for workflow scheduling in cloud. Fifth International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom 2013). https://doi.org/10.1049/cp.2019.2206

©IJRASET: All Rights are Reserved

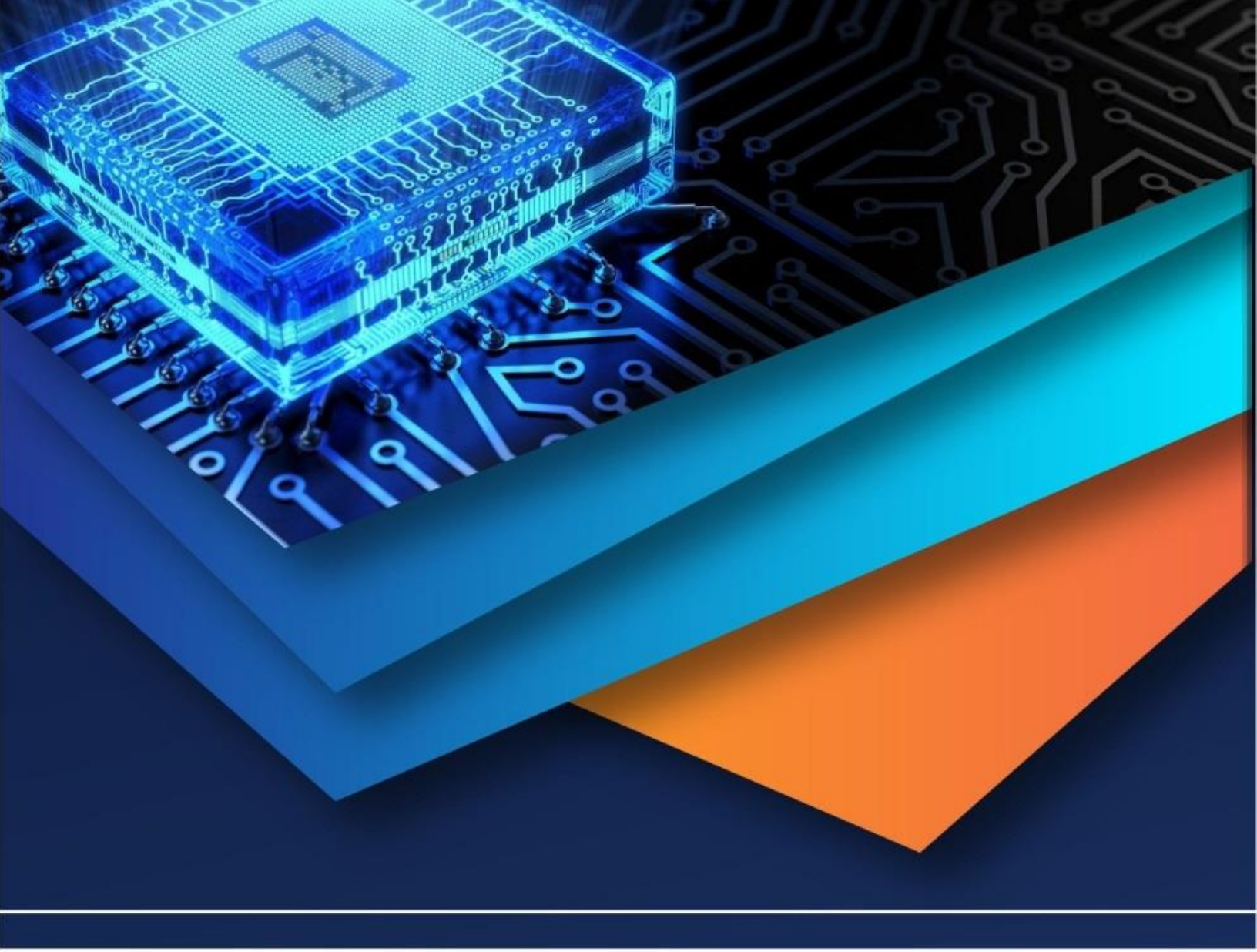
[22] Wu, L., Garg, S. K., & Buyya, R. (2011). SLA-based resource allocation for software as a service provider (SAAS) in cloud computing environments. 2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. https://doi.org/10.1109/ccgrid.2011.51

[23] Yang, Z., Yin, C., & Liu, Y. (2011). A cost-based resource scheduling paradigm in cloud computing. 2011 12th International Conference on Parallel and Distributed Computing, Applications and Technologies. https://doi.org/10.1109/pdcat.2011.1

[24] Yassa, S., Chelouah, R., Kadima, H., & Granado, B. (2017). Multi-objective approach for energy-aware workflow scheduling in cloud computing environments. The Scientific World Journal, 2013, 1–13. https://doi.org/10.1155/2013/350934

[25] Zhang, Y., & Xu, K. (2020). A survey of resource management in Cloud and edge computing. Network Management in Cloud and Edge Computing, 15–32. https://doi.org/10.1007/978-981-15-0138-8_2

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)