iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5395

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: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

Mobile Data Warehouse - Creating Multi-Cluster Web Application which will Track the Entire Process of Data Warehouse

Babita Panja[1], Prachi Bhagwatkar[2], Palak Lohia[3], Ruchita Koli[4] , Prof Kiran Yesugade[5], Prof. Dr. Sonali Kadam[6]
[1,2 ,3 4,5,6]*Computer Department, SPPU*

***Abstract**: The Project's main goal is to create a platform for easy interoperability service for the process in data warehouse system. This project is very user friendly as it provides a secure environment and easy access to the customers' databases. It is auto-scalable and economically feasible. The migration services helps migration of data inter-regionally and from one platform to another to provide them with quick statistical analysis for their convenience. It uses hadoop in its transactions which will make the system faster and speed up the data mining process. It keeps data on cloud rather than shipping the data so that data can accessed whenever and wherever.*
Keywords; Data warehouse, Migration, API

I. INTRODUCTION

Data warehousing emphasizes the collection or the capture of data from varied sources for useful analysis and access. There are two approaches to data warehousing, top down approach and bottom up approach. In top down approach the data warehouse is created first and then the specific data marts are combined (also known as local databases). In bottom up approach firstly the individual data marts are generated and then only they are combined in order to collaborate into a data warehouse.
Docker container is an open source software development platform. The main benefit is to package applications in Containers which allows them to be portable among all systems which can run the Linux operating system (OS).
The Container technology has been in the fuss for a while, but momentum and claims of Dockers approach to containers has pushed this approach to the edge in the last year. It is one form of container technology.
Container technology emerged from the Linux world, based on key features in the Linux kernel, including cgroups and namespaces. These features allow lightweight workloads to be virtualized within the Linux OS. This has become a popular way to build applications because of the speed and agility with which the applications can be built, tested, and deployed.
Linux container features were first exploited most publicly by Google, Inc., which built its data-centre on the technology and also developed the open source project Kubernetes for managing containers and container clusters. It became clear that just having container features in Linux was not enough: management tools would be needed well. Docker containers and Linux containers (LXCs) emerged as ways to manage containerized applications and make them portable across networks.
Docker engineered its containers to be operated by means of a command-line tool called the Docker client. This client may run on the container host, or through a remote interface connected to the container host. The principal job of a Docker client is to pull images of containers from a registry, which is a public or private repository of sources for ready-to-run virtual workloads. Docker Hub is the main public registry operated by Docker Inc., although there are now many others.
Once the Docker daemon pulls a container image, it builds a working model for that container using a short batch of instructions, sometimes numbering as few as one. This build file may also include directives for the daemon to pre-load the container with other components prior to running, or directives to be given to the local command line once the local container image is built (thus necessitating the local, minimalized OS).

II. THEORY

A. Operation

Data warehouse is a complementary database, where data obtained from external operational sources are organized and reshaped into a specific structure and format in order to support decision activities. It contains the conceptual, logical, and physical data models and data model types. The metadata developed into the database component of the reporting system will contain both functional information to highlight the analytical perspective regarding the meaning of data and relationships between them, as well

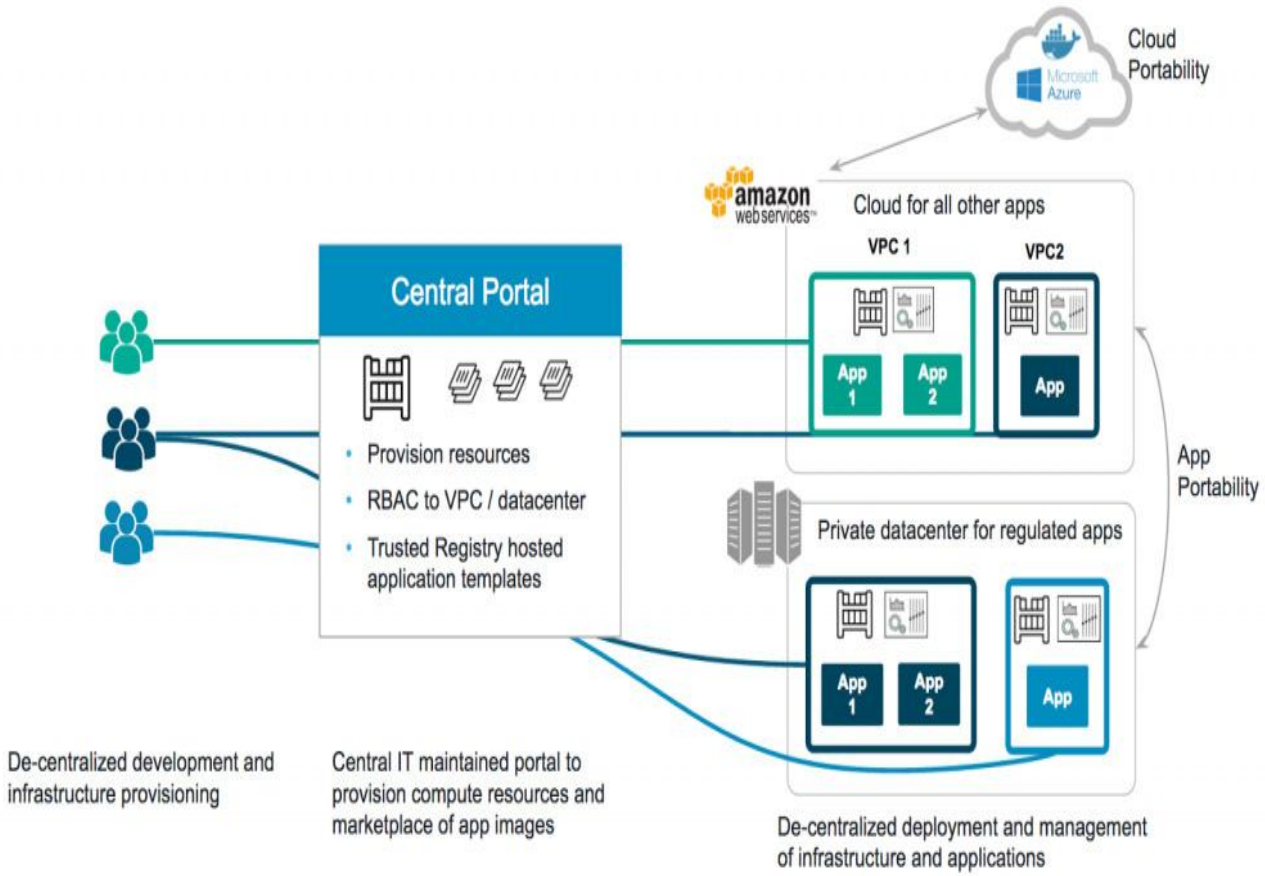
as technical information. The data warehouse database holds all the detailed information in the system. The Data Warehouse components, DDS – Detailed Data Store and SDS

Summarized Data Store (aggregates and result tables) are maintained up to date and the last one is used to provide data for a couple of applications that run over the Data Warehouse. Databases today, irrespective of whether they are data warehouses, operational data stores, or OLTP systems, contain a large amount of information. However, finding and presenting the right information in a timely fashion can be a challenge because of the vast quantity of data involved. So we'll be using Parallel execution. Using parallel execution (also called parallelism), terabytes of data can be processed in minutes, not hours or days, simply by using multiple processes to accomplish a single task. This dramatically reduces response time for data-intensive operations on large databases typically associated with decision support systems (DSS) and data warehouses. Parallelism is the idea of reeking down a task so that, instead of one process doing all of the work in a query, many processes do part of the work at the same time. Apart from all this the user will have the data sharing operation where they (users) will be able to share the data amongst

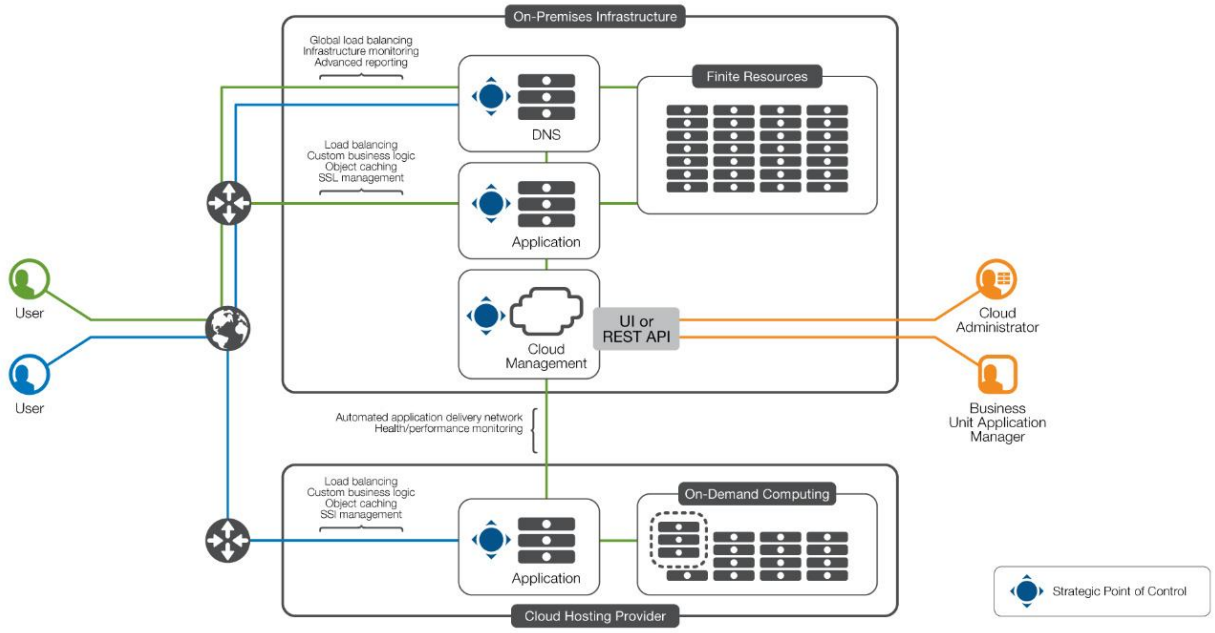
Them-selves with the help of a key which will be provided to all the user when they upload the data or when they want to extract someone's data.

There are users who want to upload a file as well as use data from other user**.** The user uploads the file in the provided field by browsing it though his documents or by using the API calls that will be provided. As the right path to the desired file is selected and upload is hit, the file will get uploaded and now an API of that particular database will be made available for the user for sharing purposes. The file which is uploaded is accessible through our front end to any internet-facing device by the use of good internet connectivity.

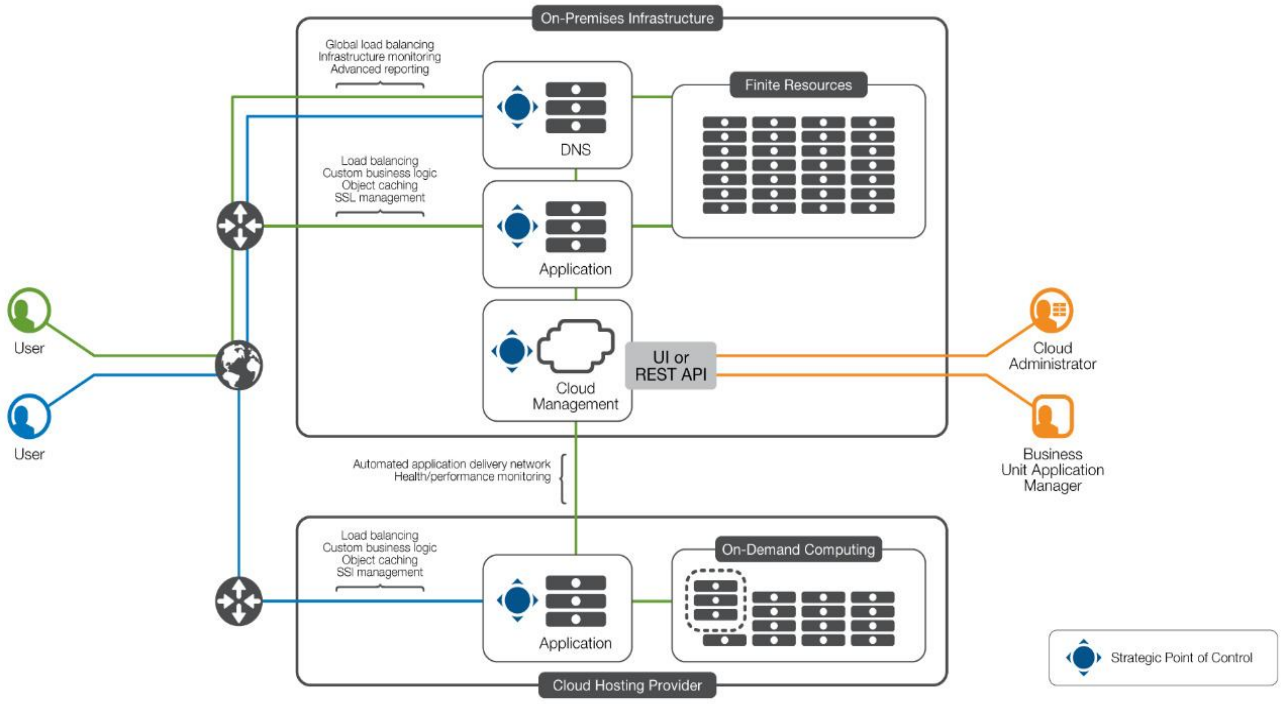
III. SYSTEM ARCHITECTURE



This is the system overview. It has three parts. First being AWS which contains two VPC and second being a private data set for regulated Apps. Third being Cloud Portability. The first VPC contains clusters needed for the application, while the second VPC contains the entire application code. An API provides the app portability between AWS and private data center for regulated apps.



There are three section in this diagram. The first and second are identical, while third section is to represent the third party data warehouse hosting application. When the user logs in he has to go through security to check in, then the DNS service helps to securely connect to the application. After this various options are provided, user to user migration and migration of data to third party application. Operation are executed as per request.



There are 2 sections in this diagram. first being the on-premises infrastructure that is the one attached with the application in the focus and the second being the other cloud hosting provided platforms, i.e. docker and assure. Once the user logs in the application with the request to migrate certain amount of data then it is migrated using on-demand computing. Resources are shifted to the finite resources with respect to the application.

IV. APPLICATION

A. Business

A new efficient decision support system can be designed to adopt this model of data warehouse.

B. Construction

Will help the construction managers to avail the knowledge of internal and external data of easy monitoring.

C. Medical

A controlled medical recent system containing records of each medical taken by each client will help in better treatment.

III. RESULT

A. The main objective is to create a platform for easy interoperability service for the process in data warehouse system.

B. The second objective is to make the entire process of data-warehousing user friendly, So that a middle developer is not needed to do the technical work. It will become an easy process for a common person without technical skills.

C. In today's scenario, data is of such important, rather big data plays a major role in taking business related decisions. In order to store such undeniable and ever growing Big-Data if the platform is auto scalable then it will be more useful for the customer. Hence, Auto scalability is the third objective.

D. Partnerships and collaborations are yet another need to share the customers' big data. But this user to user sharing capability is not yet seen in the market. Hence, the user to user sharing capability is the fourth objective.

E. The Big data is proposed to be stored on cloud so that the customer will be able to access and modify his/her data on–the-go. Such ease-of access will be provided with the mobility capability of the application.

V. DISCUSSION

Data warehouses (DW) are centralized data repositories that integrate data from various transactional, legacy, or external systems, applications, and sources. The data warehouse provides an environment separate from the operational systems and is completely designed for decision-support, analytical-reporting, ad-hoc queries, and data mining. Also is a complementary database, where data obtained from external operational sources and is organized and reshaped into a specific structure and format in order to support decision activities. It gives us the advantages such as scalability, cost reduction, migration etc.

It contains the conceptual, logical, and physical data models and data model types. The metadata developed into the database component of the reporting system will contain both functional information to highlight the analytical perspective regarding the meaning of data and relationships between them, as well as technical information. Databases today, irrespective of whether they are data warehouses, operational data stores, contain a large amount of information. However, finding and presenting the right information in a timely fashion can be a challenge because of the vast quantity of data involved. So we'll be using Parallel execution. Using parallel execution (also called parallelism), terabytes of data can be processed in minutes, not hours or days, simply by using multiple processes to accomplish a single task. Apart from all this the user will have the data sharing operation where they(users) will be able to share the data amongst themselves with the help of a key which will be provided to all the user when they upload the data or when they want to extract someone's data.

IX. CONCLUSION AND FUTURE SCOPE

As the current scenario there are different cloud service providers which give different services for storage and management of the data. Since some of the functionalities are not provided hence the above purposed system will be there to overcome the past cloud service provider service and will give new generation of migration.

The purposed system will provide the facility where different industries will be able to store the large amount of data in the databases. Here, different industries will be able to store the data in such a form that they will also be provided with different functionalities with which the processing will also get easy. Such as the migration function which will be provided where they will be able to migrate the data from user to uses as well as the data migration from one cloud service provider to other cloud service provider will be possible. Such as in the healthcare industry the data from the one of hospital can be stored in the data base and with that with the functionalities provided by the purposed system this huge database can be migrated from one cloud service provider to another and also data from one hospital can transferred to another with just some few steps with the help of a key which will be provided during the login process.

Data warehousing over the cloud computing has potential for elasticity, scalability, deployment time, reliability and reduced costs. The capabilities of the Data warehousing over cloud computing is high, parallel and distributed. It involves easy control of the environment. Data warehousing over cloud computing provides various facilities to customers like self service provisioning, self-service data management, web based upload and download the data and services are delivered over the network. IT also provides abstraction between hardware and computing software.

X. APPENDIX

Write here the appendices, if any. Write here the appendices, if any. Write here the appendices, if any. Write here the appendices, if any. Write here the appendices, if any. Write here the appendices, if any. Write here the appendices, if any. Write here the appendices, if any. Write here the appendices, if any. Write here the appendices, if any. Write here the appendices, if any. Write here the appendices, if any. Write here the appendices, if any.

XI. ACKNOWLEDGEMENT

We wish to express my sincere gratitude to Prof. Dr. S. P. Kadam for her invaluable support , encouragement, supervision and useful suggestions throughout our work. Her moral support and continuous guidance enabled us to complete our work successfully.

REFERENCES

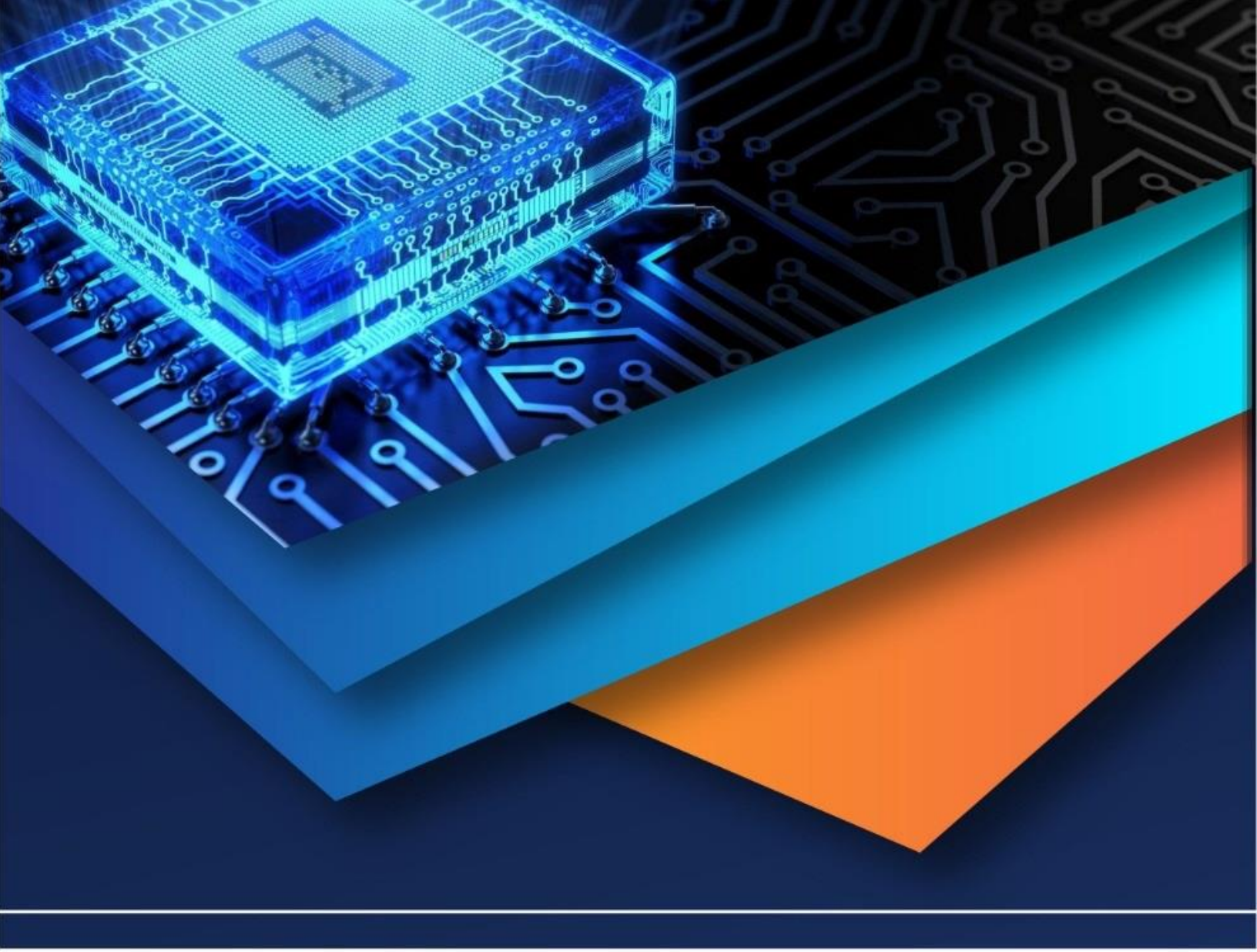
Write references here (in IEEE format). Two such templates, as sample examples, are given hereunder.

A. Journal References

[1] J. Cox, J. Kilian, T. Leighton, and T. Shamoon, “Secure spread-spectrum watermarking for multimedia”, *IEEE Transactions on Image Processing*, Vol. 6, No. 12, pp. 64 – 69, December 1997.

B. Book References

[1] J. G. Proakis and D. G. Manolakis – Digital Signal Processing – Principles, Algorithms and Applications; Third Edition; Prentice Hall of India, 2003.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY