Technical Seminar
On
Building Modern Clouds

Guide:
Prof. Sajitha N
Assistant Professor

By:
Nama Venkata Naga Sukesh
1BG16CS063

Contents

- Introduction
- Problem statement
- Literature survey
- System Architecture
- Result Analysis
- Conclusion

Introduction

- The cloud is made up of servers in data centers all over the world and provides convenience to the users accessing it. It reduces the task of maintaining the database and servers by the organisation.
- The cloud enables to access the same files and applications by multiple users from different locations.
- The computing and storing takes place in the data centers instead of locally on the user device.
- Cloud helps to create virtual instances that works similar to the physical computer with its own hardware.

Problem Statement

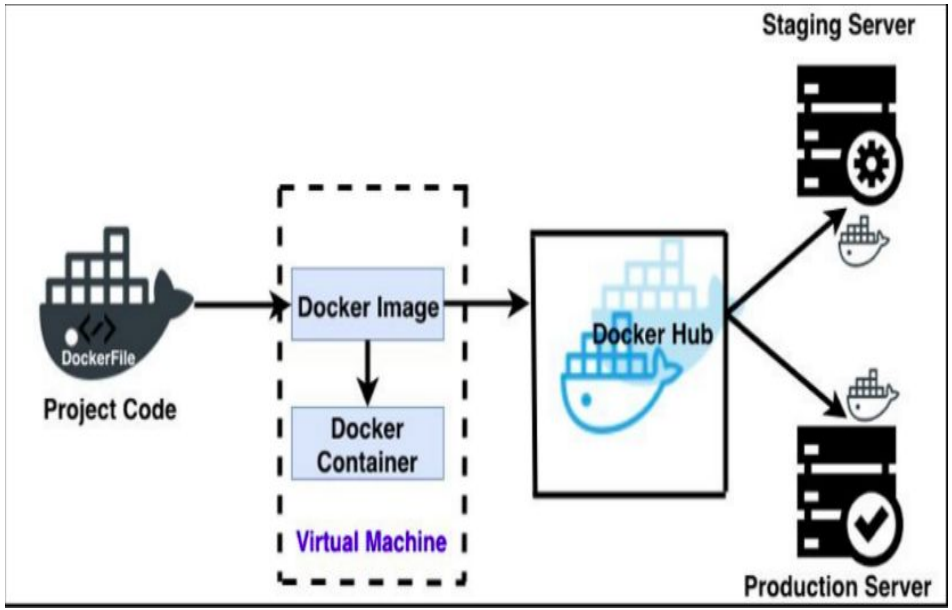
To build modern clouds using Docker, Kubernetes and Google Cloud Platform, handle Non-Relational Databases on BigQuery.

Literature Survey

1.Jay Shah, Dushyant Dubaria, Building Modern Clouds using Docker, Kubernetes & Google Cloud Platform, 2019 IEEE

Using Docker

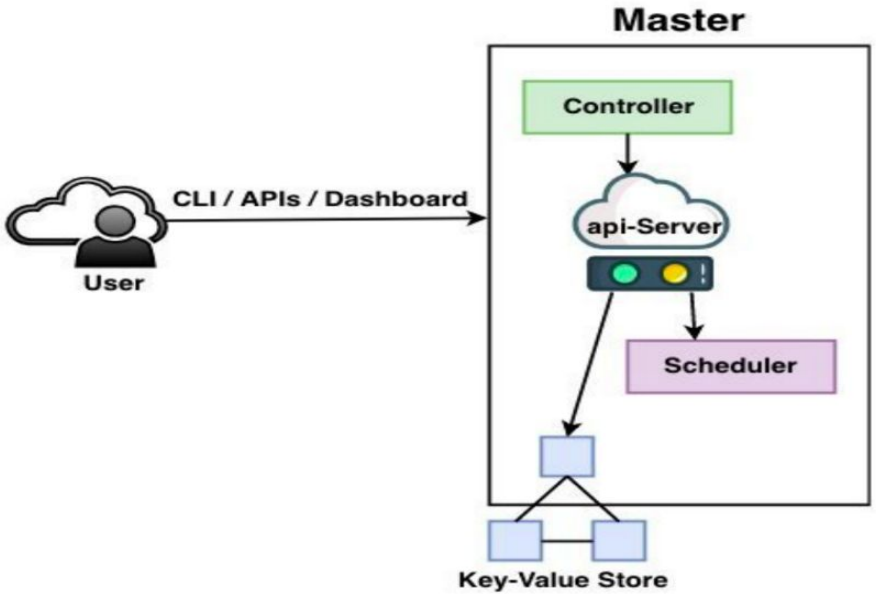
- Docker image
- Docker container
- Docker hub



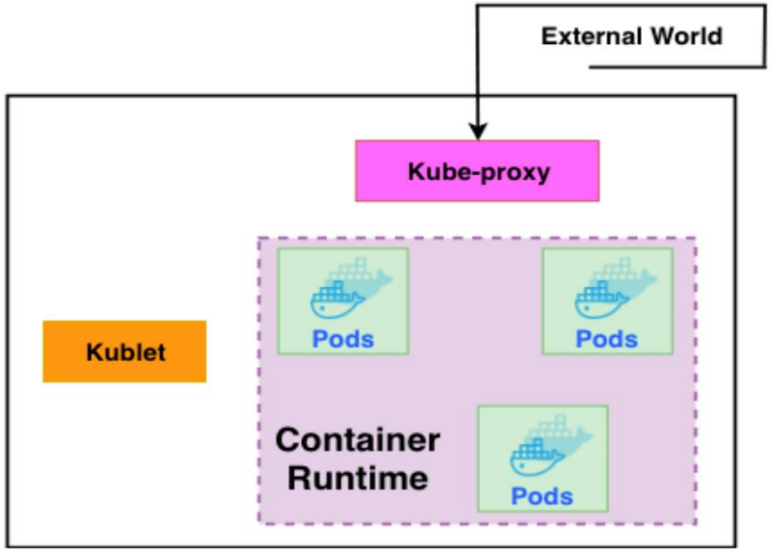
Cont.

Using kubernetes :

Master :

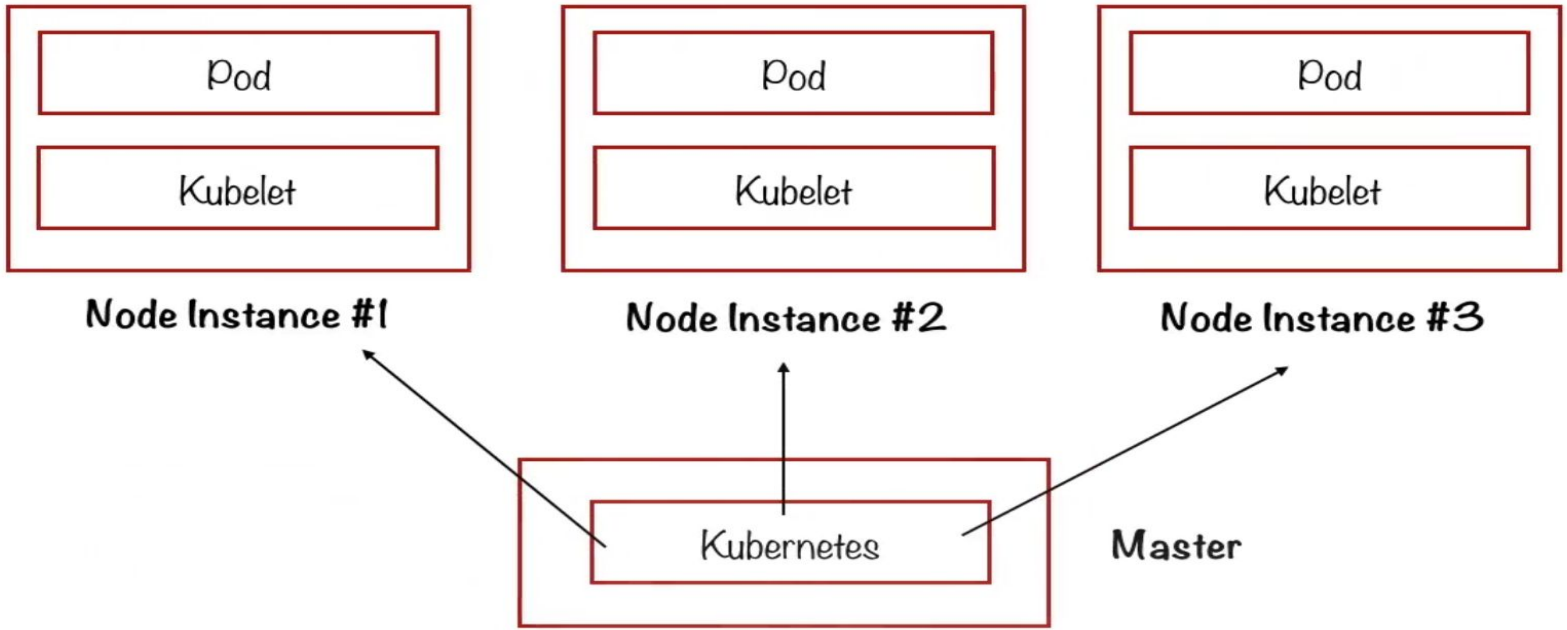


Slave :



Cont.

Kubernetes cluster :



Cont.

Google Cloud Platform:

- Installing and configuring the CLI
- Planning and configuring the cloud solution
- Planning and Configuring Network Resources
- Deployment and implementation App engine and cloud function

Cont.

2. Hera Arif, Hassan Hajjdiabl, Fatima Al Harbi , Mohammed Ghazal, A Comparison between Google Cloud Service and iCloudHera, 2019 IEEE 4th International Conference on Computer and Communication Systems

History of cloud services :

- It started in the 1950
- 1970 - concept of virtualization started
- 1990 - private virtual environment, multiple users can use one physical infrastructure
- 2000 - cloud computing flourished
- 2005 - world entered into cloud computing
- 2009 - started with google apps

Cont.

History of icloud :

- iTools-2000,.Mac-2002,MobileMe-2008
- Presented icloud in 2011 and died before launching and it is updated version of MobileMe(20GB for $99 year)
- Then it provided free 5GB along with App store, iBookstore, iCloud for backup, iTunes for music
- It hired Amazon Web service and Microsoft Azure for icloud services and later shifted some services to Google Cloud

Cont.

History of Google Cloud :

- Started cloud service in 2008 with app engine with limited storage of 500 Mb and lack of prominent languages
- 2010 introduced cloud storage
- 2011 preview of cloud SQL
- 2012 introduced Google Cloud Platform

  And BigQuery
- 2013 introduced Google Compute Engine and added growth to App Engine
- 2016 introduced Dataproc

Cont.

- GCP has many features like Compute engine, App engine, kubernetes, replica Pool, Cloud Storage, Datastore(NoSQL), Cloud sql
- Its has different services like bronze(basic level), silver($150), gold($400), platinum
- It also offers Drive, calendar etc for sync
- iCloud offers Drive, iTunes, iBooks, App Store, location sharing, Safari, Notes, keychain that store sensitive information like passwords

Cont.

Google cloud services :

- computing(Iaas,Paas)
- Storage(SQL and NoSQL)
- Big data(BigQuery,Dataprep,Pub/sub)
- Support(Gmail,Calendar,Suite,Drive,Photos,Photos)

Cont.

3. Bansari Kotecha, Hetal Joshiyara, Handling Non-Relational Databases on Big Query with Scheduling Approach and Performance Analysis ,2018 IEEE

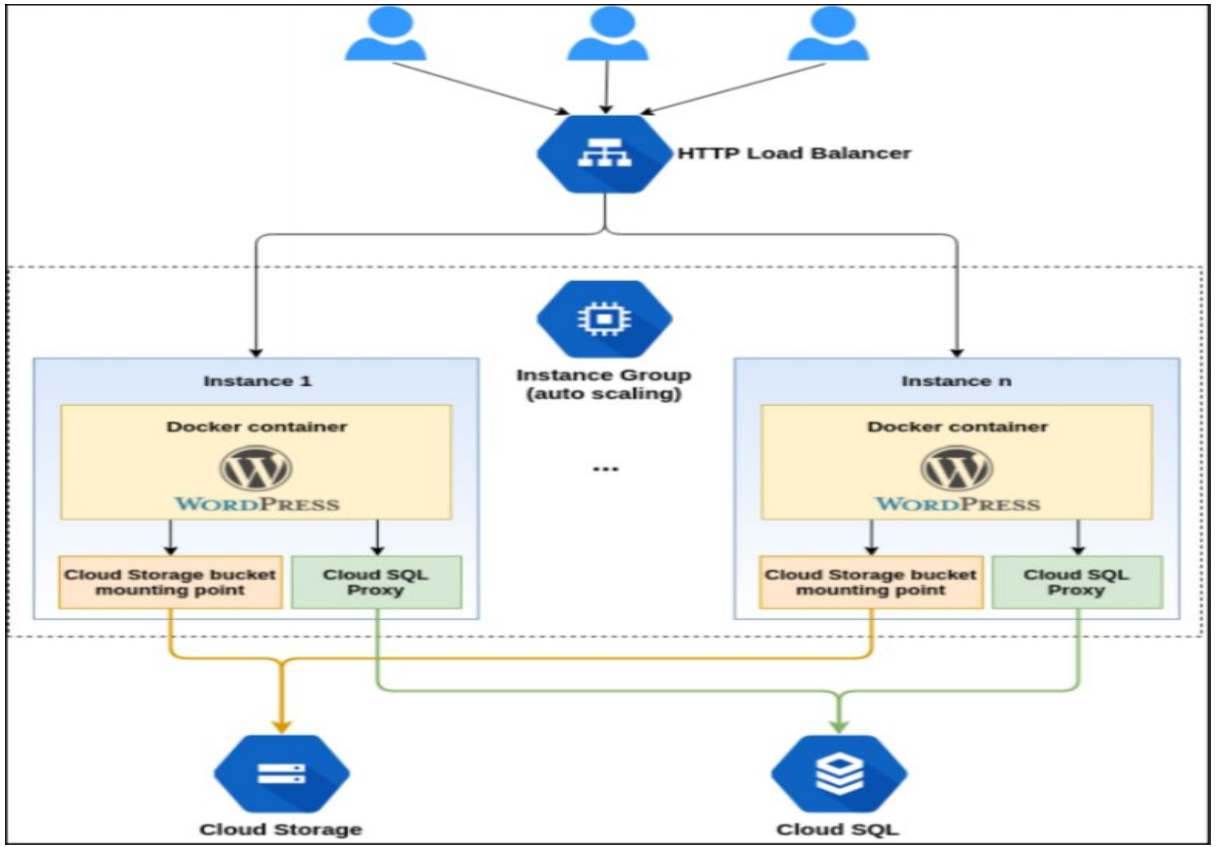
- Non relational database are used to manage large amount the data and overcomes limitations of relational database
- These Non relational databases do not have any schema like relational databases and are more scalable and flexible than relational databases
- Big Query is cloudplaform’s fully managed data warehouse that can query large amount of data.

Cont.

Steps involved in using BigQuery :

- Load your non-relational Database on google cloud
- Query the database
- Note the performance and execution time required for interactive query and batch query
- Generate chart
- Apply proposed scheduling approach

System Architecture

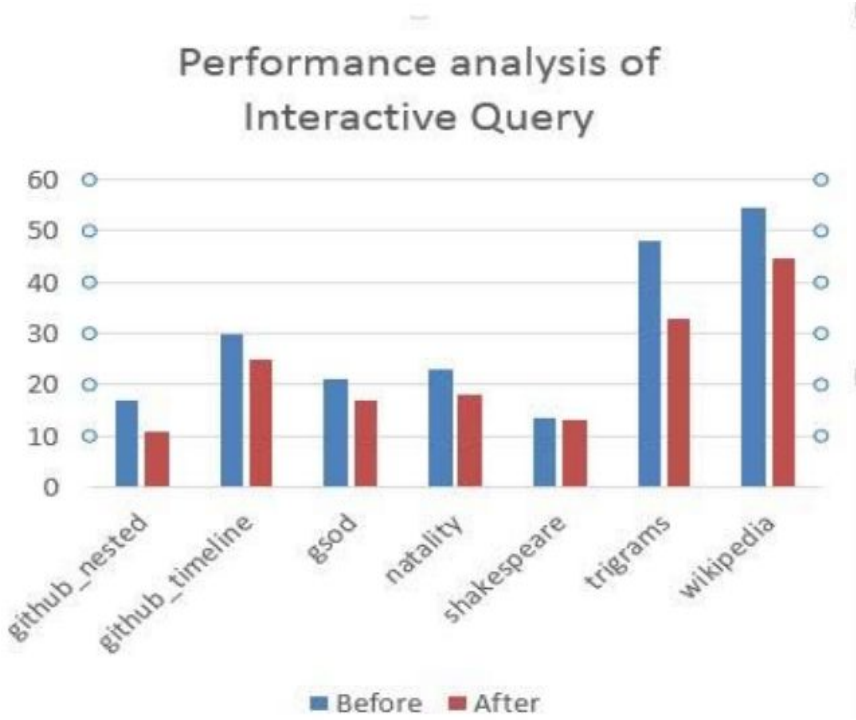


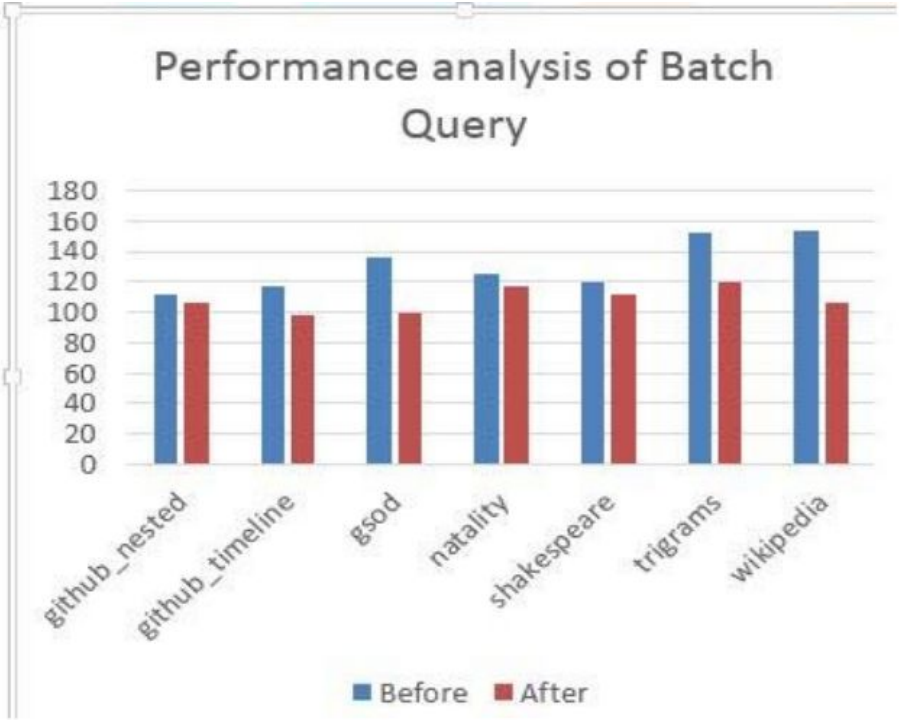
Result Analysis

- This is the dataset used to query with Interactive and batch query
- The dataset is queried before and after using scheduling approach.
- The bar chart is plotted for amount of time taken by each query to process the dataset.

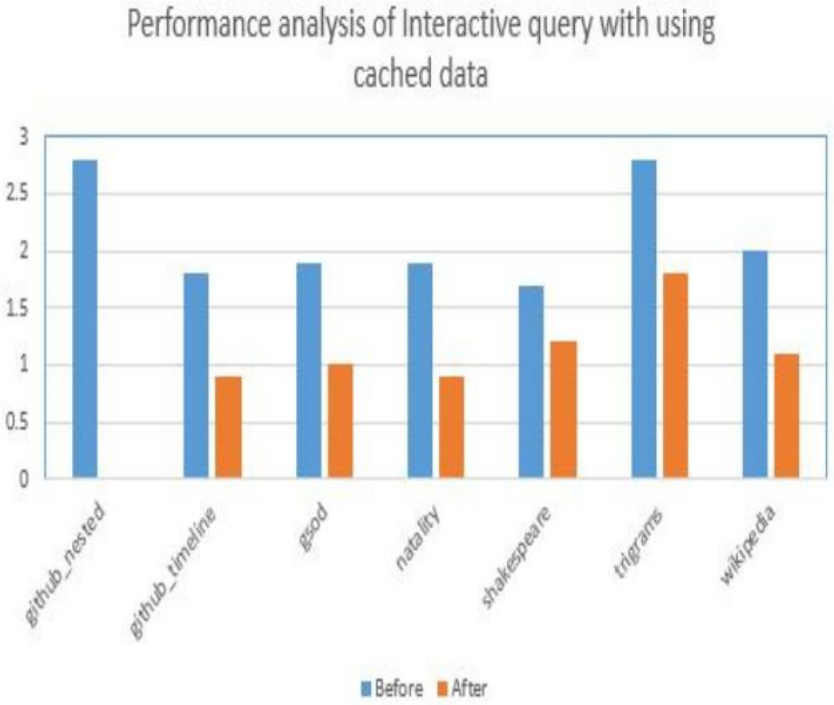
Tabel ID	Table size	Long Term Storage size	Number of Rows
github-nested	1.58 GB	1.58 GB	2,541,639
github-timeline	3.4 GB	3.4 GB	6,219,749
gsod	16.1 GB	16.1 GB	114,420,316
natality	21.9 GB	21.9 GB	137,826,763
shakespeare	6.13 MB	6.13 MB	164,656
trigrams	258 GB	258 GB	68,051,508
wikipedia	35.7 GB	35.7 GB	313,797,035

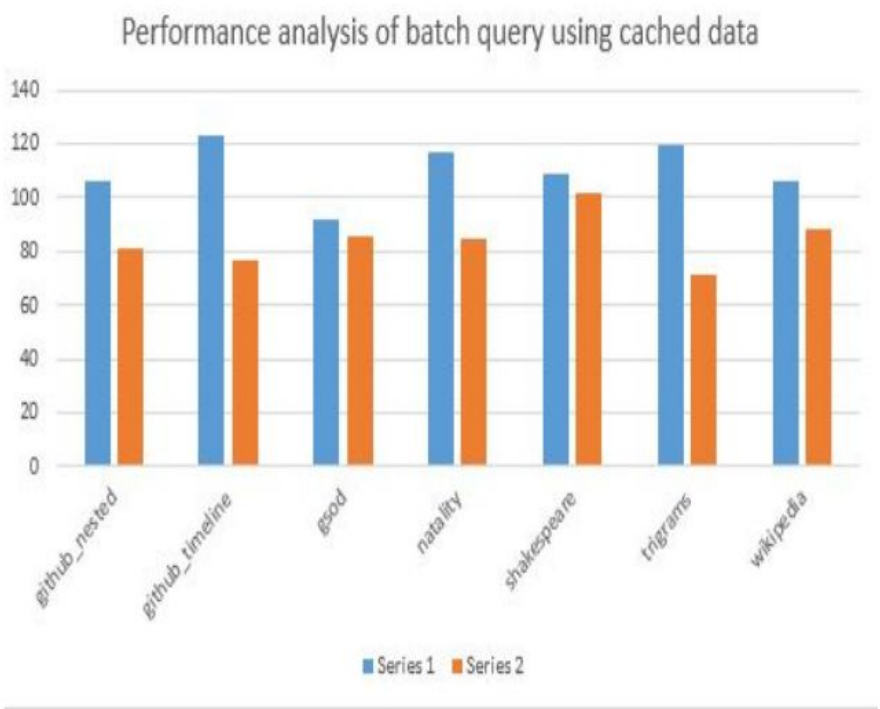
Cont.


Performance analysis of Interactive Query
60
50
40
30
20
10
0
github_nested
github_timeline
gsod
natality
shakespeare
trigrams
wikipedia
Before
After


Performance analysis of Batch Query
180
160
140
120
100
80
60
40
20
0
github_nested
github_timeline
gsod
natality
shakespeare
trigrams
wikipedia
Before
After

Cont.


Performance analysis of Interactive query with using cached data
3
2.5
2
1.5
1
0.5
0
github_nested
github_timeline
gsod
natality
shakespeare
trigrams
wikipedia
Before
After


Performance analysis of batch query using cached data
140
120
100
80
60
40
20
0
github_nested
github_timeline
gsod
natality
shakespeare
trigrams
wikipedia
Series 1
Series 2

Conclusion

- Containerization of dockers makes use of linux features that separate the entire system from other applications that makes docker to run fast and accurate.
- Kubernetes helps to communicate between the applications and allows to scale the applications as per requirement.
- Scheduling approach in BigQuery reduces the query execution time for large databases.

References

[1]. Jay Shah, Dushyant Dubaria, Building Modern Clouds using Docker, Kubernetes & Google Cloud Platform, 2019 IEEE

[2]. Hera Arif, Hassan Hajjdiabl, Fatima Al Harbi , Mohammed Ghazal, A Comparison between Google Cloud Service and iCloudHera, 2019 IEEE 4th International Conference on Computer and Communication Systems.

[3]. Bansari Kotecha, Hetal Joshiyara, Handling Non-Relational Databases on Big Query with Scheduling Approach and Performance Analysis ,2018 IEEE

THANK YOU