

Study on Usage Pattern of Public Cloud Storage

Khimya S Tinani^{1*}, Bhargav Choithwani², Bhagyashree Patil³, Pathan Faiyazkhan⁴, Tanvi Salat⁵

^{1,2,3,4,5}Dept. of Statistics, Faculty of Science, The Maharaja Sayajirao University of Baroda, Vadodara 390002, India

Corresponding Author: khimya27@yahoo.com & Mobile Number: 09824364929

DOI: <https://doi.org/10.26438/ijcse/v7i6.922927> | Available online at: www.ijcseonline.org

Accepted: 09/Jun/2019, Published: 30/Jun/2019

Abstract- Nowadays a huge amount of data is generated by individual due to sharing of data such as photos, videos, documents, use of online platforms and various gadgets. The major challenge faced by an individual is to manage the data which exceeds the limit of mass storage devices. Cloud storage services allow its users to store their data online at affordable cost and makes it available all the time. Due to accessibility, scalability affordability, etc. the usage of cloud storage is increasing tremendously. In this paper, we aim to explore and identify the factors which affect the usage of cloud data storage and find the association between them. Under this research, a survey was held to gather data from the study area that is Vadodara city. We have used the Chi-square test for independence of attributes to check the association between the factors affecting the usage of data storage and find the degree of association between them using T Schuprow's coefficient of association. Multiple Response Analysis to study the data was done. Statistical Package for Social Sciences (SPSS) software was used for analysis.

Keywords: Cloud Computing, Data Storage, Multiple Response Analysis, Chi-square, T Schuprow's.

I. INTRODUCTION

Cloud computing is defined as a model for enabling on-demand, ubiquitous and convenient network access to a shared pool of configurable computing. Cloud computing is an emerging computing platform and service mode, which tabulate and schedule service based on the internet. Cloud storage is one of the services which provide storage resource and service based on the remote storage servers based on cloud computing. There are many cloud computing and cloud storage providers, such as IBM, Google, Sun Microsystems, Microsoft, Amazon, etc. There are also more and more cloud storage platforms, e.g., HDFS, GFS, Sun Network.com, SkyDrive, Amazon S3, EMC Atoms, Data ONTAP, HP, etc. Cloud storage is a new concept come into being simultaneous with cloud computing, which generally contains two meanings: It's the storage part of the cloud computing, virtualized and highly scalable storage resource pool. Cloud users access to cloud computing services based on the cloud storage resources pool, but not all storage part can be separated in cloud computing. Cloud storage means that storage can be provided as a service over the network to the user. Based on the definition of cloud computing cloud model comes with 5 key characteristics.

On-demand self-services: Where user can w/o any human interaction can choose to manage, schedule and deploy any of cloud service whenever required. Consequently it reduces cost and personnel workload of the service provider [1].

Broad network access: Cloud consumer can access cloud services over the network via standardized interfaces, which can be accessed only by heavy complex devices such as Computer but also by other light weighted devices like smartphones, tablets. In addition, it leads to lowered cost of high bandwidth network that will sustain a high level of utilization [1].

Location independent resource pooling: Consumer generally has no control or knowledge over the location of their resources, so in order to meet consumer needs, cloud uses a technique named "Virtualization" which allows cloud to pool its computing resources [1].

Rapid Elasticity: To meet the characteristics of a self-service characteristic of cloud computing, the cloud has the ability to release and allocate the resources quickly and effectively. This automated process reduces the procurement time for new computing capabilities when required [1].

Measured services: Resources used by cloud computing can automatically measure by cloud computing which is used to provide consumers with the payment model based on "pay-per-use". Therefore resources used can be monitored, reported and controlled providing transparency for both the provider and consumer of the utilized services [1].

Service Models: The service models of cloud computing are particularly important as it defines the scope of cloud

computing. This model depicts various layers of service provided to user by cloud service provider. Clouds could be used to offer as-a-service: a platform to develop on, software to use, or an infrastructure to utilize [10].

Infrastructure as a service: Is a shared internet infrastructure such as servers and storage. Its vendors are Sun Microsystems Cloud Services, Dropbox and OneDrive.

Platform as a service: Is the application platform that provides developer with quick deployment. For example: Google App Engine, Microsoft Azure [10].

Software as a service: Where the cloud service provider delivers a software application over the web. Consumer uses this application over the cloud infrastructure. These are accessible from various client devices through a thin client interface (a web browser based email). Its target audiences are Google Docs, Apples MobileMe [10].

Deployment Model: Cloud computing has four major types of deployment model. Each model has its advantages and disadvantages.

Public Cloud: In this model, general public can ingress the services, storage application offered by the provider. The main disadvantage of this model is lack of high level security. Ex: Google App Engine, Amazon Elastic Cloud Compute, Blue Cloud by IBM.

Private Cloud: Model of this type provides access to the service and system within an organization. This model is majorly opt by finance industries, where security of data is the primary concern. Data stored in private cloud can only be accessed and shared among the users of the organization. The weak point of private cloud is that it is difficult to deploy globally. Some of the examples of this model are Microsoft Private Cloud, Amazon Virtual Private Cloud.

Hybrid Cloud: It is the combination of both public cloud and private cloud. Features of Hybrid cloud are Scalability, cost efficiency, Security, Flexibility [10]. It provides better security when compared to public cloud. This may be managed by either internally or third party [10].

II. RELATED WORK

Like real clouds which are the collection of water molecules, the term 'cloud' in cloud computing is the collection of networks [11]. The user can use the modalities of cloud computing boundlessly whenever demanded. Instead of setting up their own physical infrastructure, the users ordinarily prefer a mediator provider for the service of the internet in cloud computing. The users have to pay only for the services they had used [1]. Cloud Computing provides a surroundings for resource sharing in terms of ascendance frameworks, middleware's and application development platforms, and business applications. The operation models of cloud computing grasp free infrastructure services with

value another platform services, subscription based infrastructure services with supplemental application services, and free services for sellers but sharing of revenues generated from shoppers [2]. Cloud computing system can be end. They each area unit connected with one another through a network, sometimes the net. Front is what the consumer (user) sees whereas the rear end is that the cloud of the system. Face has the client's laptop and therefore the application needed to ingress the cloud and therefore the back has the cloud computing services like numerous computers, servers and information storage [3]. Cloud data storage, also known as a file hosting service, allows users to store their data remotely with the service provider instead of locally on their hard drive. The result of this is that the data can then be accessed over the internet from any location from devices such as desktop computers, laptops, tablets and smartphones which offers increased flexibility and accessibility [4]. The user is required to provide a user name and password to access their data and it is then possible to share files publicly or keep them password-protected [5]. Mobile devices such as smartphones also come with automatic backup-to-cloud options and the result is that when a configured device is connected to the internet, either directly or through a computer, it is updated to the cloud service provider. For example, data such as photographs and contacts can be automatically backed up from an iPhone to Apple's iCloud service [6]. Users of Samsung or HTC Android smartphones can be configured to automatically backup to their Dropbox account. It should be noted though that a user may not be actually aware that they are using a cloud data storage account to handle the backing up of their device. Signing up for this service may be a single step when configuring their smart phone or tablet for the first time. Another important aspect of cloud storage is the ability for an end user to synchronise their data to a local folder on their computer, which can hold a full copy of the data from the remote server [7]. Any changes to documents are automatically synchronised between the local folder and the cloud service provider. The user is not limited to a single device for synchronisation; this can be configured on multiple apparatuses and the result is that an up-to-date version of the data appears to be in the same folder regardless of which device is used to view it [8].

Tschuprow's T: In statistics, Tschuprow's T is a measure which is use to find the association between two nominal variables, whose range is between 0 to 1. If we have a multinomial sample of size n, then T can be estimated by using the formula

$$\hat{T} = \sqrt{\frac{\sum_{i=1}^r \sum_{j=1}^c \frac{(p_{ij} - p_{i+} p_{+j})^2}{p_{i+} p_{+j}}}{\sqrt{(r-1)(c-1)}}}$$

Where $p_{ij} = n_{ij}/n$ is the above estimated value is the empirical value of T. With the Pearson chi-square statistics, the above formula can also be written as

$$\hat{T} = \sqrt{\frac{\chi^2/n}{\sqrt{(r-1)(c-1)}}}$$

III. PROBLEM STATEMENT

This research is aimed to identify the factors which affects the usage of public cloud storage. The research is further aimed to find the association among those factors.

IV. OBJECTIVES

The objective of this paper is to gain insight about the usage patterns of public cloud storage by the people of Vadodara

V. RESEARCH METHODOLOGY

The research is done by the means of survey, based on the questionnaire which provides the qualitative data such as age, gender, education & profession of users, the cloud storage facilities they use, the cloud service provider they prefer and other related questions. The survey held with specific sampling technique and sample size determination method. Sample Size determination method is used to choose the number of observations to include in a statistical sample. It is important for any statistical study to make inference about the population based on a sample. In practice the sample size is usually determined based on the cost, time or method of collecting the data and the need for it to offer sufficient statistical power. For our study the sample size is determined based on the pilot survey results. In the pilot survey the sample of size 30 was selected, from the pilot survey the proportion of cloud storage users were 0.7 and of non-users was 0.3. The sample size for the main

city. To know the facilities which are mostly used by people – gender wise, age wise, profession wise.

(i) Which Cloud Company mostly people prefer?
 (ii) Which Mass Storage device is used by the people in Vadodara city?

(iii) To know which factor influence the usage of cloud storage.

(iv) To check dependency of specific cloud storage facilities usage on cloud company.

(v) To check whether the use of different cloud storage facilities depends on profession or not.

survey was determined by using proportion method, $p =$ proportion of users = 0.7, $q =$ proportion of non-users, taking margin of error $e = 0.05$, the sample size of the main survey was 364. For this study the primary data was collected using the questionnaire. The survey was conducted in Vadodara city. The two stage sampling technique was used for the collection of data. In first stage cluster sampling was used. The population of Vadodara city was divided into 13 clusters based on wards. Then 6 clusters were selected randomly. As the ward size varies with respect to the population, in second stage Probability proportion to size (PPS) sampling method was used. In PPS cumulative method was used to determine the no. of samples to be select from each ward. After determining the no. of samples to be obtain from each ward, individuals were selected by using simple random sampling (SRS). The below table shows the number of samples to be drawn from each ward.

Ward no.	Ward name	Size of ward	Cumulative Total	Range	No. of sample
1	Raopura	43555	43555	1 – 43555	29
8	Karelibaugh	98723	142278	43556-142278	62
9	Waghodiya	236097	378375	142279-378375	72
6	Akota	130715	509090	378376-509090	41
7	Harni	122741	631831	509091-631831	89
10	Alkapuri	177287	809118	631832-809118	71

VI. RESULTS AND DISCUSSION

VI.I. Frequency distribution:

(i) **Objective:** Which Cloud Company mostly people prefer?

Which public cloud company you use?

(a) Google Drive (b) Dropbox (c) One Drive (d) icloud (e) Others [specify].....

Cloud company Frequencies			
Cloud company	Responses		Percent of Cases
	N	Percent	
Google Drive	151	51.9%	84.8%
Dropbox	51	17.5%	28.7%

One Drive	34	11.7%	19.1%
icloud	47	16.2%	26.4%
Others cloud	8	2.7%	4.5%
Total	291	100.0%	163.5%

We observe from our survey that maximum (51.9%) of people are using Google Drive followed by Dropbox (17.5%).

VI.II. Frequency of Multiple Response analysis:

(ii) **Objective:** Which Mass Storage device is used by the people in Vadodara city.

Based on this question we conducted survey? (Allow to tick multiple answers)

Which mass storage device you use to store data?

(a) SD card (b) CD (c) Pen Drive (d) Hard Disk

Mass Storage Frequencies

Mass Storage	Responses		Percent of Cases
	N	Percent	
SD card	64	21.5%	36.0%
CD	31	10.4%	17.4%
Pen Drive	113	38.0%	63.5%
Hard Disk	68	22.9%	38.2%
None	21	7.1%	11.8%
Total	297	100.0%	166.9%

We observe from our survey that maximum (38%) of people are using Pen Drive to storage their data.

VI.III. Chi-square Test for independence of attributes :

(iii) **Objective:** To check whether the usage of cloud storage is influenced by different factors such as gender, education, profession, operating system and internal space of mobile.

H_0 : There is no association between cloud storage usage and gender, education, profession, operating system and internal space of mobile.

H_1 : There is association between cloud storage usage and gender, education, profession, operating system and internal space of mobile.

The following table gives the list of attributes and corresponding chi square value, p-value, conclusion and degree of association between them.

Attribute	Category	VS	Attribute	Category	Chi-Sq.	p-value	Association
Cloud Storage Usage	Yes	VS	Gender	Male	5.608	0.018	12.4 %
	No			Female			
Cloud Storage Usage	Yes	VS	Education	SSC	38.044	0.000	21.6 %
	No			HSC			
				Diploma			
				Graduate			
				Post Graduate			
			PHD				
Cloud Storage Usage	Yes	VS	Profession	Student	8.245	0.041	11.4 %
	No			Job			
				Business			
				Unemployed			
Cloud Storage Usage	Yes	VS	Operating system	Android	19.838	0.000	19.6 %
	No			IOS			

				Windows			
Cloud Storage Usage	Yes	VS	Internal Space	4 GB	25.236	0.000	16.8 %
	No			8 GB			
				16GB			
				32 GB			
				64 GB			
				128 GB			
				256 GB			

REFERENCES

- [1] Garrison, G., Kim, S., Wakefield, R.L. "Success Factors for Deploying Cloud Computing". Communications of the ACM. 55, 62–68, 2012.
- [2] Liang-Jie Zhang, Qun Zhou, "CCOA: Cloud Computing Open Architecture ". In proceeding of IEEE International Conference on Web Services (ICWS), pp. 607-616, 6-10, July 2009.
- [3] Yashpalsinh Jadeja, Kirit Modi, "Cloud Computing Concepts, Architecture and Challenges". In Proceeding of International Conference on Computing, Electronics and Electrical Technologies [ICCEET], 2012.
- [4] Quick, D. and Choo, K. K. R. "Digital droplets: Microsoft SkyDrive forensic data remnants". Future Generation Computer Systems-the International Journal of Grid Computing and Escience, 29, 1378-1394, 2013a.
- [5] Google Drive Support. "How to share - Drive Help.". [Online] Available at: <https://support.google.com/drive/answer/2494822?hl=en>, 2015.
- [6] Apple. "Backup and restore your iPhone, iPad, or iPod touch using iCloud or iTunes"- Apple Support. [Online] Available at: <https://support.apple.com/en-ie/HT203977>, 2015.
- [7] Drago, I., Mellia, M., M Munafo, M., Sperotto, A., Sadre, R., and Pras, A. "Inside Dropbox". The ACM Conference, 481, 2012.
- [8] Geel, M. "Cloud Storage: File Hosting and Synchronisation 2.0." [Online] Available at: https://www.vis.ethz.ch/de/visionen/pdfs/2012/visionen_2012_3.pdf?end=15&start=11 Google Drive Support. 2013.
- [9] M. M. Dahshan and Dr. Sherif El- Kassas, Data Security in Cloud Storage Services, 2013
- [10] Dean O’Gorman, Cloud Computing Data Storage: Usage, security, and privacy issues for individual users
- [11] Priyanshu Srivastava and Rizwan Khan, A Review Paper on Cloud Computing, International Journals of Advanced Research in Computer Science and Software Engineering, June 2018

Authors Profile

Dr. (Mrs.) Khimya S Tinani pursued Ph.D. from The Maharaja Sayajirao University of Baroda, Baroda in 2014. She is associated as a faculty member from last 20 years with Department of Statistics, Faculty of Science, The Maharaja Sayajirao university of Baroda. She has published 11 research papers in reputed UGC approved national and international journals and they are also available online. Her main research work focuses on Operations Research. She has 20 years of teaching experience and 10 years of Research. She has kept herself updated by conducting and attending various workshops and presented research papers in international and national conferences.



Mr. Bhargav N. Choithwani pursued Bachelor of Science in Statistics from Maharaja Sayajirao University of Baroda, India in 2018. He is currently pursuing his Master of Science in Statistics from Maharaja Sayajirao University of Baroda, India. He has done two internships first as a Research Associate at MAXIMAIZE MARKET RESEARCH PVT. LTD. Pune India in 2018. Second as a Project intern at LARSEN & TOUBRO Electrical & Automation LTD. [VSW] in 2018. He has done two projects first "SWITCH GEAR TEST DATA ANALYSIS" at L&T Electrical & Automation VSW and second "CLOUD DATA STORAGE: USAGE PATTERN OF VADODARA CITY AND THEIR CONCERNS RELATED TO CLOUD STORAGE".



Ms. Bhagyashree J. Patil pursued Bachelor of Science in Statistics from Maharaja Sayajirao University of Baroda, India in 2018. She is currently pursuing Masters of Science in Statistics from Maharaja Sayajirao University of Baroda, India. She has over 3 internship experience. Firstly she worked as Data Analyst under the title of "A study to assess the utilization of mobile health services and its cost effectiveness to the beneficiaries in the rural and tribal areas of Gujarat" at Deepak foundation, Baroda in 2017. Secondly she work as Research Analyst under the project named "MY SKYLL" at Alpha Prediction LLP, Pune, India in 2018. Thirdly she work as Research Associate at Maximize Market Research PVT. LTD, Pune, India in 2018. She has done project on "CLOUD DATA STORAGE: USAGE PATTERN OF VADODARA CITY AND THEIR CONCERNS RELATED TO CLOUD STORAGE".



Mr. Pathan Faiyazkhan pursued Bachelor of Science in Statistics from Maharaja Sayajirao University of Baroda, India in 2015. He is currently pursuing Master of Science in Statistics from Maharaja Sayajirao University of Baroda, India. He has worked as a Data analyst in the Bio statistical study “oral health of pregnant women”. He is working as a Data analyst in the Bio statistical study “Fluoride level in coffee”. He has done project on “CLOUD DATA STORAGE: USAGE PATTERN OF VADODARA CITY AND THEIR CONCERNS RELATED TO CLOUD STORAGE”.



Ms. Tanvi S. Salat pursued Bachelor of Science in Statistics from Maharaja Sayajirao University of Baroda, India in 2018. She is currently pursuing Masters of Science in Statistics from Maharaja Sayajirao University of Baroda, India. She worked as Data Analyst intern under the title of “A study to assess the utilization of mobile health services and its cost effectiveness to the beneficiaries in the rural and tribal areas of Gujarat” at Deepak foundation, Baroda in 2017. Further she was appointed as Field Assistant in Research Project [MoEF] under the title of “ Biodiversity Assessment of Central Gujarat using Geospatial and Conventional Approach with Emphasis on Conservation of Important Threatened Taxa” in Maharaja Sayajirao University, 2018. She has done project on “CLOUD DATA STORAGE: USAGE PATTERN OF VADODARACITY AND THEIR CONCERNS RELATED TO CLOUD STORAGE”.

