BIG DATA AND CLOUD READINESS

[1]N.Kamala Vikasini, Assistant Professor, CSE(DS), vikasini574@gmail.comSwarna Bharathi institute of science and technology,
Khammam

[2]V.Krishnaveni, Assistant Professor , CSE,veni.ch1801@gmail.com
Swarna Bharathi institute of science and technology
,Khammam

[3]N.Saritha, Associate Professor CSE(DS),
sarithanune@gmail.com
Swarna Bharathi institute of science and technology,
Khammam

ABSTRACT

How companies reach their goals may be drastically altered by combining big data with an adaptable cloud platform. A lot of businesses are making the switch to the cloud, but it seems that the trust problem is slowing them down. The elements that influence Trust as a result of Service Satisfaction were examined in this article. Researchers employed the PLS-SEM method to analyze the connection of the variables since the sample was not normally distributed.Cloud advantages, data privacy, service level agreement (SLA), risk management, trust, service satisfaction, and data security are the factors. Qualitative study grounded in theory established relationships between the variables, and further validation of these associations was achieved by quantitative data analysis. According to the quantitative data analysis, trust is influenced by service satisfaction, which in turn is influenced by data security, cloud benefits, reputation, and SLA. Big Data, Cloud, and Trust: Key Concepts.

1. INTRODUCTION

Companies are exploring the possibility of storing their data on the cloud due to the many advantages it offers [1]. The proliferation of businesses offering big data and cloud solutions has accelerated the shift towards cloud computing. Specialists are available to sift through mountains of data in order to provide useful information to the general population [2].With the cloud's massive storage capacity and powerful data analysis capabilities, organizations may get a better understanding of their data, which in turn allows them to deploy better solutions and make better decisions [3]. Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Business Intelligence (BI), and Document Management Systems (DMS) are just a few of the programs that help complicated enterprises run smoothly on the cloud [3]. As the demand for internet services and communications continues to rise, cloud computing is assuming more and more significance [2]. The sheer volume of services, data, and devices kept in the cloud, however, raises security issues [2]. Despite the cloud's rising popularity and many benefits, businesses are nonetheless hesitant to transfer critical data to it [4]. Due to a lack of faith in the cloud service, businesses are very wary about providing them with their most critical data [5]. The broad adoption of big data in cloud computing is hindered by a crucial challenge: creating trust among cloud members, which is necessary for data storage in the cloud [6][7]. Therefore, this research uses the organizational system-based theory to suggest that businesses place their faith in cloud computing and big data [8].There are several parts to this work. In Section 2, we cover the reasons for doing the study.

The study's goal is outlined below. The approach is detailed in Section 3.The reflective measuring scale is presented in Section 6, descriptive statistics are explained in Section 5, and the data analysis procedure is shown in Section 4. In Section 6, we cover the reflective measurement scale; in Section 7, we go over the validation method; and in Section 8, we examine the data through the lens of the structural model. Finally, Section 11 provides a brief summary of the paper's main points, Section 10 details next projects, and Section 9 showcases the updated model with data analysis findings and debates.

2. MOTIVATION

To make sure big data in the cloud is utilized effectively, you may follow the inquiry that leads to Service Satisfaction and Trust [9][10]. Consequently, it is necessary to verify the components of Qualitative Data Analysis that are grounded in theories.The study's overarching goal is to provide light on what factors lead to trust and satisfaction with service based on quantitative data analysis. Organization theory provides support for quantitative data analysis, which in turn validates the framework established from the qualitative research study [11]. References [12] and [13].

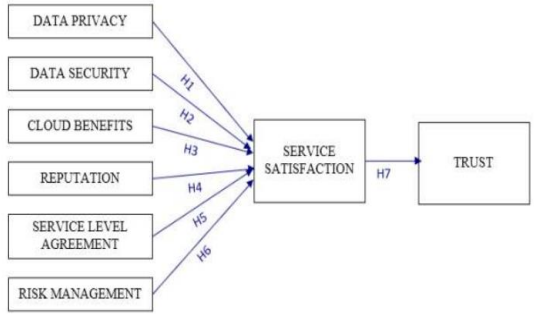

Figure 1. Organization System Theory: Trust for Big Data in Cloud

3. METHODOLOGY

The research methodology details the steps to take in order to address the research questions. What data was acquired and how was it collected? Who or what provided the data? How was it analyzed? All of these questions are answered in the methodology section of the study. The researcher is instructed on how to explore and resolve an issue in the methods chapter. Research using a combination of qualitative and quantitative techniques is known as mixed-methods research. Researchers might get a quantitative research approach by first doing qualitative research. In order to gain better informed opinions from qualified IT candidates, researchers should do qualitative research first. Candidates are given the opportunity to provide detailed responses by way of the interview questionnaire's use of open-ended questions. A research framework is developed by the researcher via the study of gathered data that is backed by theories. This study's sample was selected using a method known as intentional sampling. In order to find the IT staff responsible for cloud-based big data use and to get correct answers to questions on cloud computing and big data, the researchers used deliberate sampling.

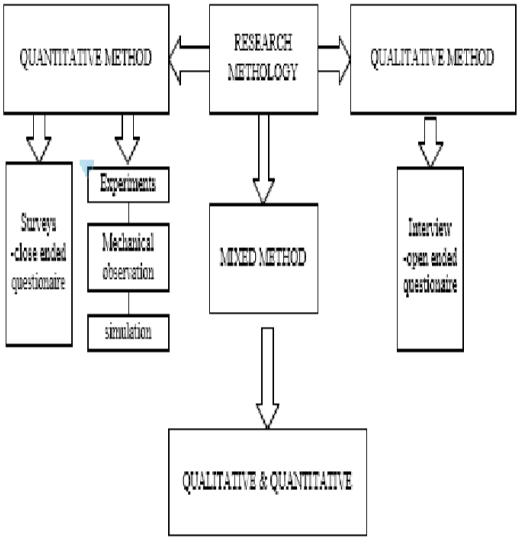

Figure 2. Research Methodology

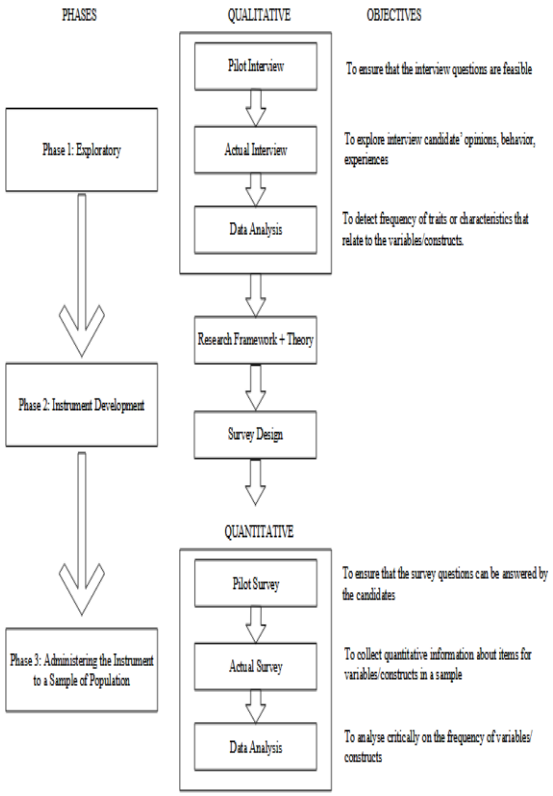

Figure 3. Exploratory Sequential Mixed Methodology

Most individuals worked from home during the epidemic, therefore that's when the data was collected. Consequently, the researchers contacted them using the web-based instrument. Prospective participants were provided with a link to the survey, which was uploaded to Google Form in the cloud. We reached out to the participants using LinkedIn and WhatsApp, among other social media. Connections made via LinkedIn attracted most of the participants.

Table 1. Quantitative Data Collection

Task	Start Date	End Data	No of Days
Pilot Data Collection	30 March 2021	6 April 2021	7 days
Actual Data Collection	7 April 2021	22 July 2021	107 days

Information gathering for the pilot project started on March 30, 2021, and ceased once the participant count reached 30. It took seven days to finish collecting data for the pilot. Researchers used the results of the pilot data analysis to confirm that the questionnaire questions were accurate and relevant to the applicants. This ensured that the remaining sample would be able to correctly answer the questions. Data collecting started once the pilot's analysis was finished. A total of 107 days, from April 7th to July 22nd, 2021, were devoted to data collection. With Linked-in, it's much easier to identify competent individuals since they may be picked using specific criteria. Consequently, there were enough participants for the researchers to conduct deliberate sampling.The researchers conducted three layers of screening to guarantee that the applicants were appropriate. Two checks are included in the survey: the first is in the authorization letter to administer the survey, and the second is a short statement of the survey's purpose on the questionnaire's cover page. As a last step, the questionnaire's third screening takes place at Question 6, which inquires about the organization's use of cloud computing services. The survey will close if they choose "No". If the candidates are selected incorrectly, the researchers feel that the results of the investigation will be meaningless. The eight (8) constructs and several indicators for each construct are handled using PLS-SEM in this study. Each construct has anything from three to six indications.

4. DATA ANALYSIS PROCESS

Researchers used descriptive statistics, a measurement model, and a structural model to examine the quantitative data that was gathered, as shown in Figure 3. The researchers began their investigation into the organization's big data consumption in the cloud with descriptive statistics, while data analysis is done using PLS-SEM.

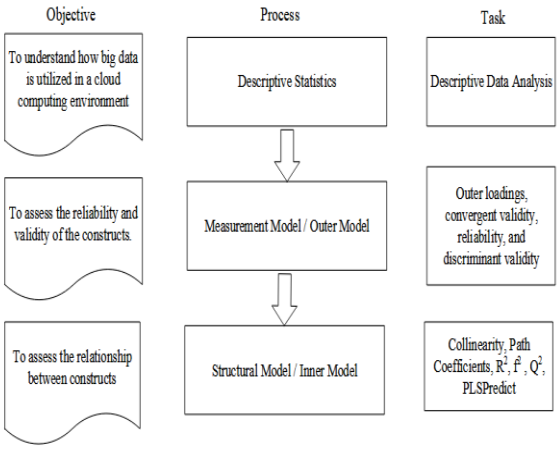

Figure 4. PLS SEM Data Analysis Process

Following this, the researchers proceeded to evaluate the construct's validity and reliability using a measurement model, also known as an outer model. As a last step, researchers might use the structural model, also called the inner model, to evaluate the interrelationships of the constructs. By following each step in Figure 3, the researchers were able to identify the important and insignificant relationships between the components.

5. PLS-SEM DESCRIPTIVE

The researchers began the real data analysis by using SEM descriptive data analysis, as shown in Table 2. In order to explore how big data is used in a cloud computing setting, the researchers descriptively evaluated the actual data that was gathered.

Table 2. PLS-SEM Descriptive

Variable	Category	Percen
Role	Director/CEO/Owner	1
	Manager	3
	Executive	3
	Others	
Gender	Male	
	Female	
Age	Less than 21	
	21 – 30	
	31 – 40	
	41 -50	
	Above 50	
Highest Education	Secondary	
	College/Matriculation/Polytech	
	University	
Years with company	1-2 years	
	3-5 years	
	6-10 years	
	More than 10 years	
IT Knowledge	None at all	
	Minimal	
	Knowledgeable	
	Very knowledgeable	
Cloud Experience	1-2 years	
	3-4 years	
	More than 4 years	
Organization Involvement	Services (IT-based)	54.5
	Services (Non-IT based)	21.3
	Others	24.2
Type of Organization	Sole proprietorship	17.1
	Sdn Bhd	24.4
	GLC Company	20.9
	Government	9.8
	Others	27.8
Years Established	1-5 years	11.9
	6-10 years	8.9
	11-15 years	13.7
	More than 15 years	65.5
Number of Employees	Less than 200	33.2
	More than 200	66.8
Number of IT Staffs	1-5	22.5
	6-10	8.1
	More than 10	69.4

Since the ideal participants have sufficient background in cloud and big data management, their profile is well-suited to the study. In most cases, years of service, job title, age, and level of education correlate with experience. The fact that they are mostly men suggests that they use cloud computing services in a very collaborative and networked way. People with over a decade of expertise in the field, familiarity with IT in general and cloud computing in particular, make up a significant portion of the sample. So, it's safe to say that the sample is well-versed in cloud computing and big data, which gave them little trouble answering the questions.According to the statistics, Sdn Bhd firms have the greatest proportion of organizations with over 200 workers. Additionally, more than half of these companies have been in operation for more than fifteen years. These companies have been around for a while and have plenty of expertise with big data on the cloud, according to the data we have. No matter whether the sample's enterprises are organized, semi-structured, or unstructured, they are all making use of big data services. Microsoft Azure is the cloud platform of choice for more organizations than Amazon Web Services, Google Cloud Platform, or Cloudera, with Cloudera having the fewest users. With the As a Service model, businesses have the option to choose many service models. CRM, MIS, ERP, HRM, LMS, CM, and GIS are all part of SaaS. Other applications in this category include DBMS and CAD.Application design, development, testing, and deployment are all part of PaaS. Infrastructure as a service (IaaS) handles software packages, public IP addresses, virtual local area networks (VLANs), storage, firewalls, and load balancers. Data as a service (DaaS) lets you clean and enhance data before making it available to other systems, apps, or people. In conclusion, XaaS integrates SaaS, PaaS, and IaaS providers. Companies employ a combination of cloud components as a service, thus. A hybrid cloud is the preferred delivery strategy for many organizations. They may be reluctant to use the

public cloud because of security and privacy concerns, but they need cloud services and have decided to go with a hybrid solution instead. This is why they lean toward hybrid cloud architectures, in which certain services reside in public clouds while others continue to operate in private ones. If companies want to get the benefits of cloud-based big data without sacrificing service quality, a hybrid cloud is the way to go. Nearly 50% of those surveyed are employed by IT firms, and even more are large organizations (over 200 people) than that. Eighty-three percent of respondents said they utilized structured data, sixty-three percent used semi-structured data, and the other participants said they used unstructured data formats including Word, PDF, and social media. They are all using cloud systems like Cloudera, Amazon Web Services, Google Cloud Platform, and Microsoft Azure.In order to study what leads to Service Satisfaction and, ultimately, Trust, it is crucial to choose a suitable sample. People in charge of the organization's big data use are the intended recipients of the Trust in the use of Big Data in the Cloud. Their likelihood of using the cloud will decrease if they lack confidence. In conclusion, the researchers were able to learn more about the scope of cloud-based big data use by analyzing the descriptive statistics provided above. Concerns about service satisfaction and trust seemed to arise throughout cloud service utilization.The indicators are derived from the quantitative survey questionnaire's itemized questions. Data Privacy (DP), Data Security (DS), Cloud Benefits (CB), and so on are the factors that determine the labels given to the indicators. The PLS-SEM Data Analysis is used for quantitative data. Finally, the researchers were able to achieve Figure 7. After adjusting for loading, reliability, and validity, Figure 7 displays the revised study framework.

6. REFLECTIVE MEASUREMENT MODEL

One aspect of the model is the reflective measurement scale, which includes latent constructs (variables) that have an effect on, influence on, or cause the indicators that make up the model [14]. When the latent construct changes, all indicators will alter as well. Indicators stand in for latent variables, acting as empirical stand-ins or proxy variables [15]. There has to be extensive testing of the indicators' validity and reliability because of how interdependent and interchangeable they are. A total of four observable indicators—DP1, DP2, DP3, and DP4—make up data privacy. The three (3) visible signs that make up data security are DS1, DS2, and DS4. Because of the low factor loading, DS3 has been eliminated. The five indicators that make up Cloud Benefits are CB1, CB2, CB3, CB4, and CB5.

The three markers of reputation are REP1, REP2, and REP3. The SLA is characterized by four (4). Three indicators are used in risk management. There are a total of six (6) indicators for both Trust and Service Satisfaction. Evaluation and reporting of outer loadings, composite reliability, AVE, and square root is required. As seen in Figure 6, the causation direction in a reflective measuring scale travels from the blue-colored latent variable to the yellow-colored indicators.

7.VALIDATION PROCESS

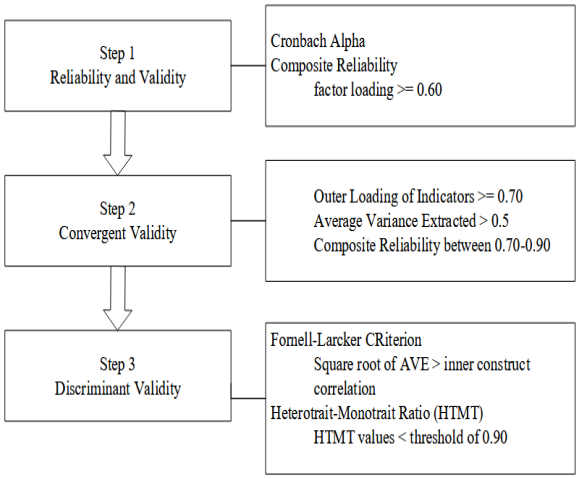

Figure 5. Validation Process

Evaluating the constructs' validity and reliability is the job of the measurement model, often known as the outer model. The measuring model, often called the outside model, shows the connection between the components and the indicators. The reliability, discriminant validity, convergent validity, and outer loadings of an indicator are used to evaluate its relationship to its concept [14]. It is possible to construct a measurement model using the PLS Algorithm. The inner workings of the fundamental PLS algorithm, as used in SmartPLS 3.0 [14], are described in depth in [16] and [17].

Table 3. Types of Variables and Indicators

Independent	Indicator	Mediator	Indicator	Dependent
Data Privacy	DP1	Service Satisfaction	SS1	Trust
	DP2		SS2	
	DP3		SS3	
	DP4		SS4	
Data Security	DS1		SS5	
	DS2		SS6	
	DS3			
	DS4			
Cloud Benefits	CB1			
	CB2			
	CB3			
	CB4			
	CB5			
Reputation	REP1			
	REP2			
	REP3			
Risk Management	RM1			
	RM2			
	RM3			
	RM4			

7.Step 1: Reliability and Validity

The variables' validity and reliability were examined by the researchers via the use of Composite Reliability (CR) and Cronbach Alpha.
Table 3's indicators are labeled according to the variables in Figure 1, where DP denotes Data Privacy, DS denotes Data Security, CB denotes Cloud Benefits, etc. We eliminated indicators that had factor loadings below 0.6. The study excluded three items (DS3, RM1 and RM5) due to low factor loadings (<0.600). You can see the item factor loadings and reliability and validity findings here. The constructions (variables) are legitimate and dependable. When it comes to measuring data security, indicators DS1, DS2, and DS4 all converge. When it comes to measuring cloud benefits, CB1, CB2, CB3, CB4, and CB5 all come together. All of the other indicators in Table 3 follow the same pattern.

8.STRUCTURAL MODEL/ INNER MODEL

The structural model determines how the variables are related to one another. After a successful evaluation of the measurement model, the structural model of the PLS-SEM data is examined. A reflective construct is shown in Figure 5. It comprises of a single mediator variable, Service Satisfaction, six independent variables, a dependent variable, Data Privacy, Data Security, Cloud Benefits, Reputation, SLA, and Risk Management. As a result of factor loading being less than 0.60, one indication (DS3) and two indicators (RM1 and RM5) were eliminated.Section 8.1. Initial Step: Assess Collinearity in the Structural Model
It is possible to check for multicollinearity by looking at VIF values; if they are less than 3.0, then no problems exist [17]. Benefits of the Cloud, Data Privacy and Security, Reputation, Risk Management, and Service Level Agreement (SLA) with Regard to Customer Satisfaction (Table 6)

9. RESULT AND DISCUSSION

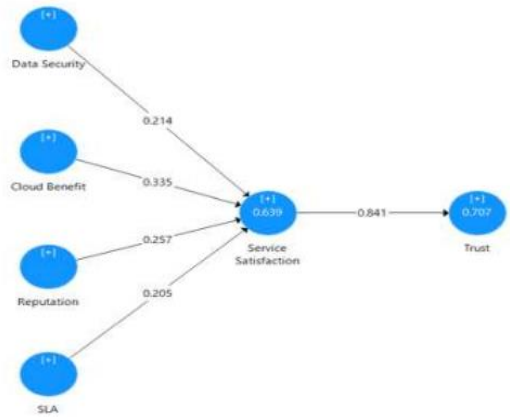

Figure 7. Revised Model

10. FUTURE WORK

Experts agree that sampling should be meticulous and exhaustive. As a result of differences in responsibilities, several IT managers approached the topic of cloud computing from different angles. While some saw the cloud as a place to save massive data sets, others saw it more as a tool for analytics. IaaS, PaaS, and SaaS are the three distinct varieties of SaaS. Purposive sampling that is both sufficient and accurate is therefore necessary. The sample has to be carefully selected before being included in the interview questionnaire. In terms of cloud administration, they ought to be the ideal candidate for that particular IT job. Making sure they can accurately reflect the population and provide accurate answers is the goal of the screening process. One possible explanation is population. Possessing a SLA with them may do wonders for your Trustworthiness, in addition to their Reputation. Users are able to transfer the responsibility for data security to cloud providers via SLAs. it may be challenging to find the right candidate, but it can help in choosing the most appropriate constructions to describe the

11. CONCLUSIONS

In order to reap the benefits of the cloud without unnecessary worries or reservations, this research is crucial.Big data processing on the cloud comes with its fair share of unknowns, therefore it's important for customers to have clear standards before committing to the service. The research concludes by shedding light on how businesses may secure the cloud storage of their big data, thereby contributing to the body of knowledge. Cloud computing has much more benefits than drawbacks. Due to the many benefits, cloud customers rely on service level agreements (SLAs) and provider reputation to alleviate their worries about data security. Customers will only trust cloud suppliers with stellar reputations. Having a service level agreement (SLA) with them may assist build trust, in addition to reputation. Through SLAs, cloud customers are able to transfer the responsibility of data security to cloud providers.

REFERENCES

[1] B. M. Balachandran and S. Prasad, "Challenges and Benefits of Deploying Big Data Analytics in the Cloud for Business Intelligence," Procedia Comput. Sci., vol. 112, pp. 1112–1122, 2017.

[2] A. Sether, "Cloud Computing Benefits," SSRN Electron. J., no. January, 2016.

[3] K. Tatic, "The benefits of using cloud technology in Bosnia and Herzegovina," no. April, 2020.

[4] S. M. Din, R. Ramli, and A. A. Bakar, "A Review on Trust Factors affecting Purchase Intention on Instagram," 2018 IEEE Conf. Appl. Inf. Netw. Secur., no. November, pp. 49–53, 2018.

[5] F. Selnes, "Consequences of Trust and Relationships," Eur. J. Mark., vol. 32, no. 3/4, pp. 305–322, 1998.

[6] J. K. Adjei, "Explaining the role of trust in cloud computing services," Info, vol. 17, no. 1, pp. 54–67,2015.

[7] Z. Tbatou, A. Asimi, and C. El Balmany, Trust in Cloud Computing Challenges: A Recent Survey, vol. 81, no. January. Springer International Publishing, 2020.

[8] R. R. Greene, General systems theory, no. January 2013. 2017.

[9] F. Zohra Filali and B. Yagoubi, "Global Trust: A Trust Model for Cloud Service Selection," Int. J.Comput. Netw. Inf. Secur., vol. 7, no. 5, pp. 41–50, 2015.

[10] M. Felici and S. Pearson, "Accountability, Risk, and Trust in Cloud Services: Towards an Accountability-Based Approach to Risk and Trust Governance," 2014 IEEE World Congr. Serv., pp. 105–112, 2014.

[11] I. Encalada, "General systems theory . Practical approach in education," no. October, 2019.

[12] A. Šijan, D. Karabašević, and D. Rajčević, "The importance of the general system theory for the modern world," Trendovi u Posl., vol. 7, no. 2, pp. 87–94, 2019.

[13] A. Steblyanskaya, Y. Jun, and M. Vasiev, "General systems theory for sustainable green development," no. September, pp. 138–141, 2021.

[14] J. F. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM : Indeed a Silver Bullet," vol. 19, no. 2, pp.139–151, 2011.

[15] J. F. Hair, J. J. Risher, M. Sarstedt, and C. M. Ringle, "When to use and how to report the results of
PLS-SEM," Eur. Bus. Rev., vol. 31, no. 1, pp. 2–24, 2019