TRUST FOR BIG DATA USAGE IN CLOUD

Hazirah Bee Yusof Ali and Lili Marziana Abdullah

Kulliyyah of Information and Communication Technology (KICT),
International Islamic University Malaysia, Kuala Lumpur,
Malaysia

ABSTRACT

Integrating big data with an agile cloud platform can significantly affect how businesses achieve their objectives. Many companies are moving to the cloud, but the trust issue seemed to make a move to the cloud slower. This paper investigated the factors that affect Service Satisfaction that led to Trust. Since the sample was not normally distributed, the researchers used the PLS-SEM tool to analyse the relationship of the variables. The variables are Data Security, Data Privacy, Cloud Benefits, Reputation, Service Level Agreement (SLA), Risk Management, Service Satisfaction and Trust. The variables were linked together based on the analysis from qualitative research supported by theories, and the linkages were being validated through quantitative data analysis. The quantitative data analysis found that Data Security, Cloud Benefits, Reputation and SLA influence Service Satisfaction and Service Satisfaction influences trust.

KEYWORDS

Trust, Big Data, Cloud.

1. INTRODUCTION

Big data and cloud computing have altered the way companies operate. By the end of the day, most business and leisure activities are conducted in the cloud. Companies cannot afford not to maximise big data in the cloud because the benefits of the cloud are too appealing [1]. Similarly, several companies are there to provide solutions for analysing large datasets, assisting in disseminating helpful information to the general public [2]. Powerful data analysis tools, combined with the cloud's large storage capacity, enable businesses to understand their data better and, as a result, implement excellent solutions and improve their decision-making skills [3]. Cloud is a platform that provides solutions for increasingly complex businesses, including Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Business Intelligence (BI), and Document Management Systems (DMS), among others [3]. Cloud computing is becoming increasingly important because of the ever-increasing need for internet services and communications [2].

However, there are security concerns due to the high number of services, data, and devices housed in cloud environments [2]. Businesses are still reluctant to move sensitive data to the cloud, despite its many advantages and growing popularity [4]. Businesses are highly cautious about entrusting their most sensitive data to a cloud service provider because they lack confidence in the cloud service [5]. Storage data in the cloud means building Trust among cloud members, and having enough Trust is a critical challenge for the widespread adoption of big data in cloud computing [6][7]. As a result, the paper proposes organisations' Trust in big data and cloud computing using the organisation system-based theory [8].

David C. Wyld et al. (Eds): EMSA, SEA, AIFU, NLCAI, BDML, BIoT, NCOM, CLOUD, CCSEA, SIPRO - 2022
pp. 287-304, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120624

This paper is divided into a few sections. Section 2 is the motivation of why the research was conducted. Here, the objective of the study is defined. Section 3 explains the methodology. Section 4 displays the data analysis process, and Section 5 explains descriptive statistics, while Section 6 presents the reflective measurement scale. After explaining the reflective measurement scale, Section 7 talks about the validation process, and Section 8 analyses data from the angle of the structural model. Section 9 illustrates the revised model along with results and discussions based on the data analysis, Section 10 states the future works, and Section 11 concludes the paper.

2. MOTIVATION

The investigation that contributes to Service Satisfaction leading to trust can be used as a guideline to ensure the successful usage of big data in the cloud [9][10]. Thus, the factors acquired from Qualitative Data Analysis supported by theories need to be validated.

The paper's objective is to present the findings of the quantitative data analysis on the factors that contribute to Service Satisfaction, which leads to Trust. The investigation through quantitative data analysis validates the framework captured from the analysis of qualitative research, which is supported by the Organization theory [11][12][13].

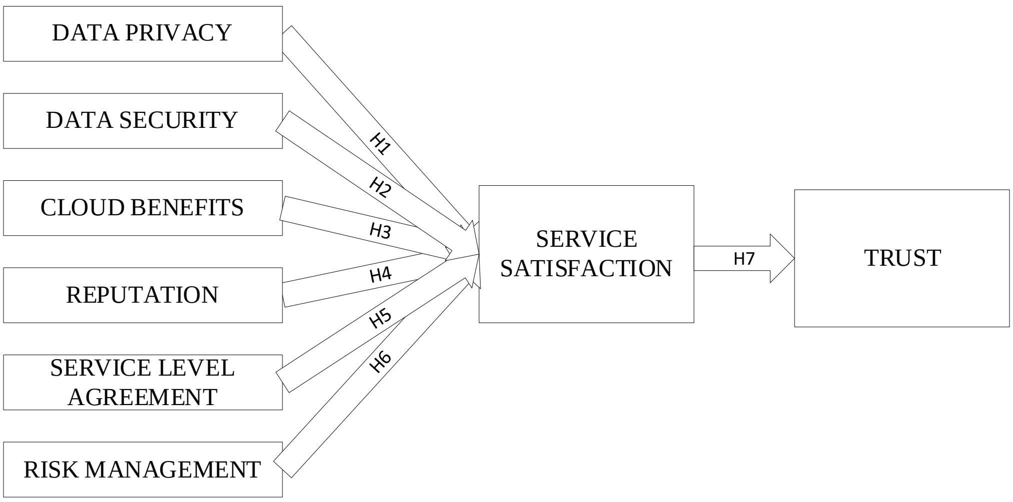

Figure 1. Organization System Theory: Trust for Big Data in Cloud

3. METHODOLOGY

Mixed method research is chosen where the qualitative method is done first, followed by a quantitative research method. By doing qualitative research first, the researchers are able to get more input from knowledgeable IT candidates. The interview questionnaire consists of open-ended questions, which enables the interview candidates to elaborate when answering the questions. The analysis of collected data supported by theories helps the researcher to produce a research framework.

The sample for this study was chosen through purposeful sampling. The researchers utilised purposeful sampling to discover IT personnel in charge of big data utilisation in the cloud and quickly and accurately answer big data and cloud questions.

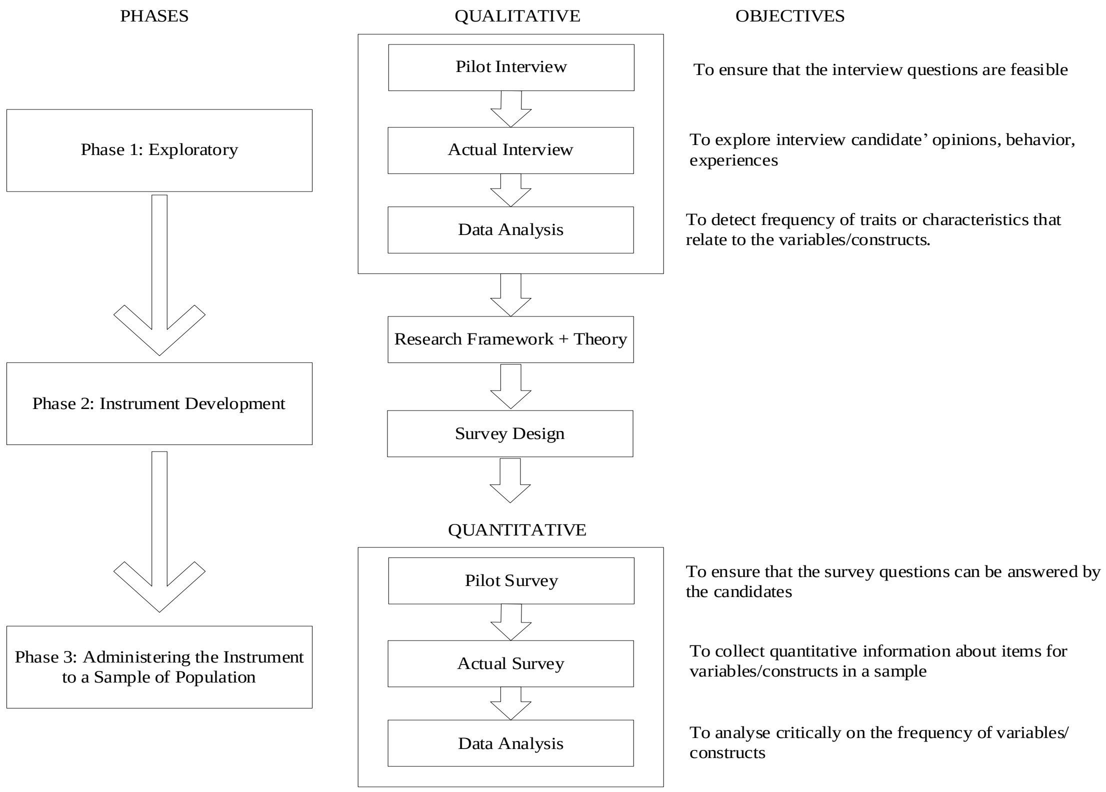

Figure 2: Methodology

Data collection was done during the pandemic when most people worked from home. So, the researchers used the online tool to communicate with them. The survey questionnaire was uploaded to Google Form in the cloud, and the link was given to prospective participants to complete the questionnaires. The participants were contacted via social media platforms such as WhatsApp and LinkedIn. The majority of the participants were acquired using LinkedIn.

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

Pilot data collection started on March 30th, 2021, and ended once data reached 30. The pilot data collection took seven days to complete. The researchers used the findings from pilot data analysis to confirm that the questions in the questionnaire are correct and related to the candidates so that they can be answered by the rest of the sample accurately.

Once data analysis for the pilot was completed, the actual data collection began. The actual data collection took 107 days, from 7 April 2021 to 22 July 2021. Linked-in enables the selection of candidates based on specific criteria so, the task of looking for suitable candidates became much more manageable. As a result, the researchers managed to get a sufficient number of candidates for purposeful sampling research.

To ensure suitable candidates, the researchers used three screening levels. The first screening is by using the permission letter to conduct the survey, and the second screening is on the front page of the questionnaire, where it states the objective of the survey briefly. Finally, the third screening is at Question 6 of the questionnaire, which says: "Does your organisation use cloud computing services?" If they select "No", then the questionnaire will exit. The researchers believe that the research output will be irrelevant if the candidates are chosen wrongly.

For this research, PLS-SEM is used to handle eight (8) constructs with many indicators for each construct. The indicators range from three (3) to six (6) indicators per construct.

4. DATA ANALYSIS PROCESS

The collected quantitative data was analysed based on the guidelines of Figure 3 where the researchers followed all the processes of descriptive statistics, measurement model, and structural model. As data analysis is done using PLS-SEM, the researchers started with descriptive statistics to investigate the extent of big data usage in the cloud for the organisation.

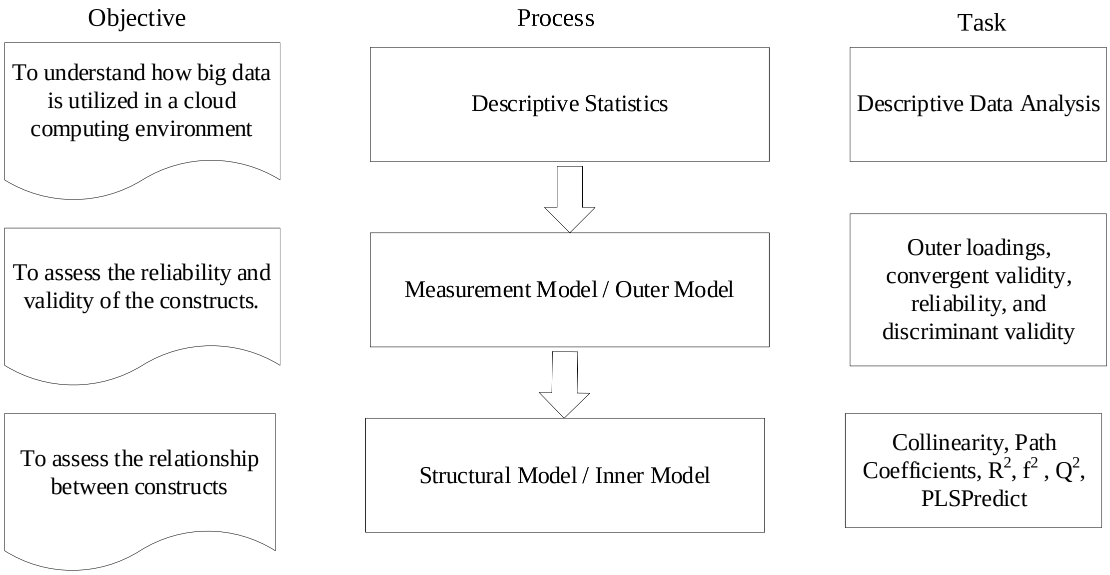

Figure 3: PLS SEM Data Analysis Process

Then the researchers continued with a measurement model or outer model to assess the reliability and validity of the construct. Finally, the structural model or inner model helps the researchers assess the relationship between constructs. Going through all the processes in Figure 3, the researchers determined which relationship of constructs are significant and insignificant.

5. PLS-SEM DESCRIPTIVE

SEM descriptive data analysis is the first process that the researchers implemented for the actual data analysis, as displayed in Table 2. The actual data collected was analysed descriptively so that the researchers could investigate the extent to which big data is utilised in a cloud computing environment.

Table 2: PLS-SEM Descriptive

Variable	**Category**	**Quantity**	**Percentage(%)**
Role	Director/CEO/Owner		14.6
	Manager		35.2
	Executive		32.4
	Others		17.8
Gender	Male		70.8
	Female		29.2
Age	Less than 21		1.4
	21 – 30		8.7
	31 – 40		28.3
	41 -50		37.4
	Above 50		24.2
Highest Education	Secondary		2.6
	College/Matriculation/Polytech		9.1
	University		87.7
Years with company	1-2 years		23.7
	3-5 years		21.5
	6-10 years		13.7
	More than 10 years		41.1
IT Knowledge	None at all		0
	Minimal		32.4
	Knowledgeable		56.6
	Very knowledgeable		32.4
Cloud Experience	1-2 years		25.7
	3-4 years		25.2
	More than 4 years		49.1
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

The profile is suitable for the research since the target candidates are experienced enough in the management of big data and the cloud. Most of the time, experience tally with the length of service, position, age, and education. Being a male-dominated group indicates strong networking and collaboration among them in using cloud computing services.

A significant percentage of the sample consists of those with more than ten years of experience, know about IT technology and specifically cloud computing itself. Therefore, the sample is experienced and knowledgeable in big data and cloud, and that specific characteristic enabled them to answer the questionnaire comfortably.

The data shows that Sdn Bhd companies have the highest percentage with more than 50% of the organisation are more than 15% years of establishment, and most of them have more than 200 employees. This information tells us that the organisations are well established with many years of big data in cloud experiences.

The sample is using big data services regardless of structured, semi-structured and unstructured businesses.

More organisations prefer Microsoft Azure as their cloud platform than Google, Amazon and Cloudera, and Cloudera has the least number of organisations using it as the cloud platform.

In As a Service business model, they may choose more than one service business model. SaaS includes DBMS, CAD, MYOB, CRM, MIS, ERP, HRM, LMS, CM, and GIS. PaaS includes facilities for application design, application development, testing, and deployment. IaaS deals with virtual machines, storage, firewalls, load balancers, IP addresses, VLANs, and software bundles. DaaS involves cleaning, enriching data, and offering it to different systems, applications, or users. Finally, XaaS combines SaaS, PaaS, and IaaS offerings. So, we can conclude that organisations use more than one cloud component as a service.

Many organisations prefer a hybrid cloud as a delivery model. The choice of hybrid may be due to their need to use the cloud as a service but are hesitant to use the public cloud due to security and privacy issues. Because of that, they prefer hybrid where some parts of the cloud services are from the public cloud, and others remain in a private cloud. The hybrid cloud choice is to ensure that they are satisfied with the service and Trust to benefit from using big data in the cloud.

Half of the sample work with IT companies, and more than half are big organisations with more than 200 employees. 83.3% of the sample claimed to use structured data, 63.5% used semi-structured data, while others used unstructured data such as word, pdf, and social media. All of them are using cloud platforms such as Microsoft Azure, Google, Amazon, and Cloudera.

The selection of a proper sample is essential to investigate the factors that contribute to Service Satisfaction and finally towards Trust. The Trust in the usage of big data in the cloud is specifically targeted to those in charge of the big data usage for the organisation. If they do not trust, then their chances to use the cloud will be reduced. To summarise, the analysis of the descriptive statistics above does help the researchers investigate the extent of big data usage in the cloud. There seemed to be some issues in the usage of cloud services that relate to service satisfaction and Trust.

The indicators are taken from the itemised questions of the quantitative survey questionnaire. The indicators are labelled based on the variables where DP stands for Data Privacy, and DS stands for Data Security, CB for Cloud Benefits and so on. Quantitative data is analysed using PLS-

SEM Data Analysis. The researchers managed to get Figure 6 as the final outcome. Figure 6 is the adjusted research framework considering the loading, reliability, and validity of Figure 4.

6. REFLECTIVE MEASUREMENT MODEL

The model has a reflective measurement scale in which latent constructs (variables) whose indicators are influenced, affected, or caused by the latent variable [14]. All indicators will change when the latent construct changes. The indicators serve as empirical substitutes (proxy variables) for the latent variables [15]. Because the indicators are strongly connected and interchangeable, they must be rigorously tested for reliability and validity.

Data Privacy is made up of four (4) observed indicators: DP1, DP2, DP3, and DP4. Data Security is made up of three (3) observed indicators: DS1, DS2, and DS4. DS3 has been removed due to low factor loading. Cloud Benefits has five (5) indicators: CB1, CB2, CB3, CB4 and CB5. Reputation has three (3) indicators: REP1, REP2, and REP3. SLA has four (4) indicators. Risk Management has three (3) indicators. Both Service Satisfaction and Trust have six (6) indicators each.

Outer loadings, composite reliability, AVE, and square root should be examined and reported. In a reflective measurement scale, the causality direction goes from the blue-colour latent variable to the yellow-colour indicators of Figure 6.

7. VALIDATION PROCESS

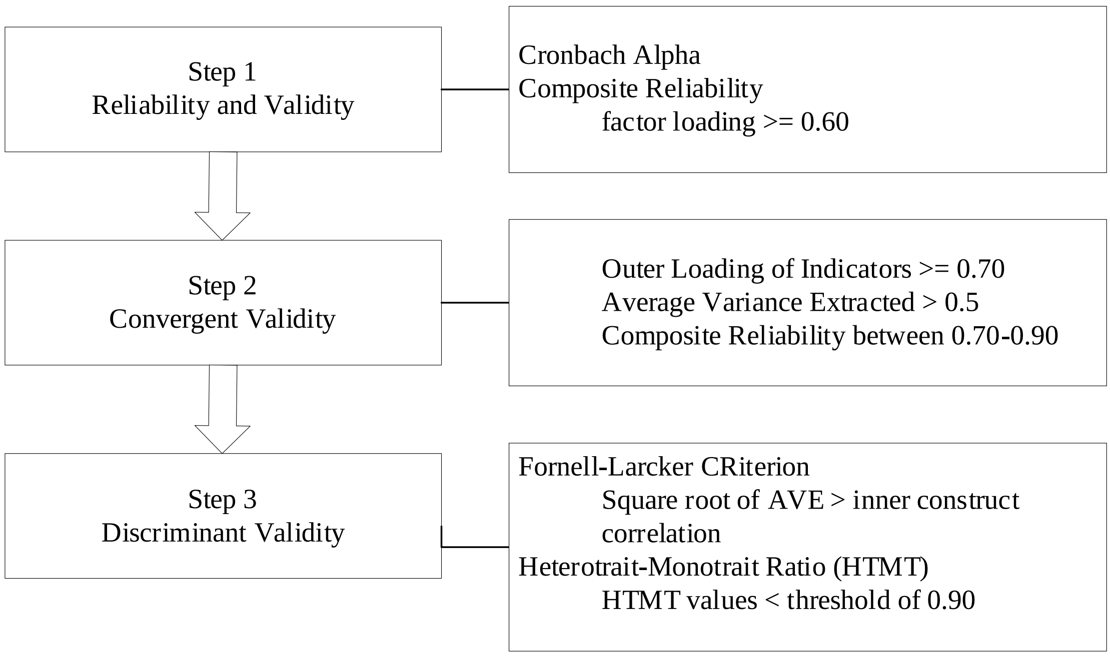

Figure 4: Validation Process

.

The measurement model /outer model is to assess the reliability and validity of the constructs. The relationship between the constructs and the indicators is shown in the measurement model, also known as the outer model. Outer loadings, convergent validity, reliability, and discriminant validity are used to assess the link between indicators and their constructs [14]. The PLS

Algorithm can be used to create a measurement model. [16] and [17] provide detailed explanations of how the basic PLS algorithm operates as implemented in SmartPLS 3.0 [14].

Table 3: Types of Variables and Indicators

Independent	Indicator	Mediator	Indicator	Dependent	Indicator
Data Privacy	DP1 DP2 DP3 DP4	Service Satisfaction	SS1 SS2 SS3 SS4 SS5 SS6	Trust	T1 T2 T3 T4 T5 T6
Data Security	DS1 DS2 DS3 DS4				
Cloud Benefits	CB1 CB2 CB3 CB4 CB5				
Reputation	REP1 REP2 REP3				
Risk Management	RM1 RM2 RM3 RM4				

7.1. Step 1: Reliability and Validity

Table 4: Loading, Reliability and Validity

	Loading	**Cronbach Alpha**	**Composite Reliability**	**Ave**
CB1	0.744	0.870	0.906	0.658
CB2	0.807			
CB3	0.835			
CB4	0.843			
CB5	0.821			
DP1	0.703	0.792	0.865	0.617
DP2	0.880			
DP3	0.809			
DP4	0.738			
DS1	0.838	0.794	0.880	0.710
DS2	0.907			
DS4	0.777			
REP1	0.869	0.897	0.936	0.829
REP2	0.935			
REP3	0.925			
RM2	0.858	0.906	0.941	0.843

RM3	0.947			
RM4	0.947			
SLA1	0.824	0.859	0.904	0.703
SLA2	0.838			
SLA3	0.819			
SLA4	0.871			
SS1	0.864	0.937	0.950	0.762
SS2	0.886			
SS3	0.917			
SS4	0.909			
SS5	0.884			
SS6	0.770			
T1	0.807	0.909	0.930	0.688
T2	0.849			
T3	0.822			
T4	0.841			
T5	0.812			
T6	0.843			

The researchers tested the reliability and validity of the variables using Cronbach Alpha and Composite Reliability (CR). The indicators in Table 3 are labelled based on the variables in Figure 1 where DP stands for Data Privacy, DS stands for Data Security, CB for Cloud Benefits and so on.

Indicators with factor loading less than 0.6 were removed. Three (3) items (DS3, RM1 and RM5) were removed from the analysis because of low factor loadings (<0.600). The results for reliability and validity, along with the factor loadings for the items, are presented. The variables (constructs) are reliable and valid. Indicators DS1, DS2, and DS4 converge and measure Data Security. CB1, CB2, CB3, CB4, and CB5 converge and measure Cloud Benefits. The same applies to all other indicators in Table 3.

7.2. Step 2: Convergent Validity

Convergent validity refers to how indicators that measure the same construct agree with each other since they measure the same construct [16]. For convergent validity, the outer loadings of the indicators have to be at 0.70 or higher [18]average Variance extracted (AVE), AVE > 0.50 [16], and composite reliability (CR) should be between 0.70-0.90 [15]

All the Alpha values and CRs were higher than the recommended value of 0.700. The average Variance Extracted (AVE) and CRs were all higher or close to 0.5 and 0.7, which corroborates convergent validity.

Table 5: Heterotrait-Monotrait Ratio (HTMT)

	Cloud Benefits	Data Privacy	Data Security	Reputation	Risk Management	SLA	Service Satisfaction	Trust
Cloud Benefits	***0.811***							
Data Privacy	0.486	***0.786***						
Data Security	0.483	0.328	***0.842***					
Reputation	0.594	0.262	0.480	***0.910***				
Risk Management	0.489	0.412	0.449	0.447	***0.918***			
SLA	0.509	0.341	0.401	0.551	0.546	***0.838***		
Service Satisfaction	0.688	0.341	0.577	0.682	0.530	0.605	***0.873***	
Trust	0.663	0.365	0.546	0.604	0.551	0.677	0.830	***0.829***

Note: Values in italic represent the square-root of AVE.

Step 7.3: Discriminant Validity

Fornell-Larcker Criterion

Discriminant validity was assessed by fornell-larcker criterion. The square root of AVE for the construct was greater than inter construct correlation.

Heterotrait-Monotrait Ratio (HTMT)

Discriminant validity was also assessed by Heterotrait-monotrait ratio of correlation [16] with values below the threshold of 0.90, Since all the values are below 0.90. then, discriminant validity is established.

8. STRUCTURAL MODEL/ INNER MODEL

Structural model assesses the relationship between variables. When the measurement model assessment is satisfactory, the next step is evaluating the structural model of PLS-SEM results. Figure 5 has reflective constructs where it consists of one mediator variable (Service Satisfaction), six independent variables (Data Privacy, Data Security, Cloud Benefits, Reputation, SLA, and Risk Management), and one (1) dependent variable (Trust). One indicator (DS3) and two indicators (RM1 and RM5) were removed due to low factor loading (< 0.60).

8.1. Step 1: Evaluate structural model collinearity

VIF values can be examined, and if they are below 3.0, then there is no issue with multicollinearity [17].Table 6 shows that Inner VIF for Cloud Benefits, Data Privacy, Data Security, Reputation, Risk Management, and SLA toward Service Satisfaction is below 3. So, there is no multicollinearity issue. The same applies to Inner VIF of Service Satisfaction towards Trust.

Table 6: VIF Values

Construct	Service Satisfaction	Trust
Cloud Benefits	2.668	
Data Privacy	1.680	
Data Security	1.759	
Reputation	2.338	
Risk Management	1.950	
SLA	2.132	
Service Satisfaction		1.000
Trust		

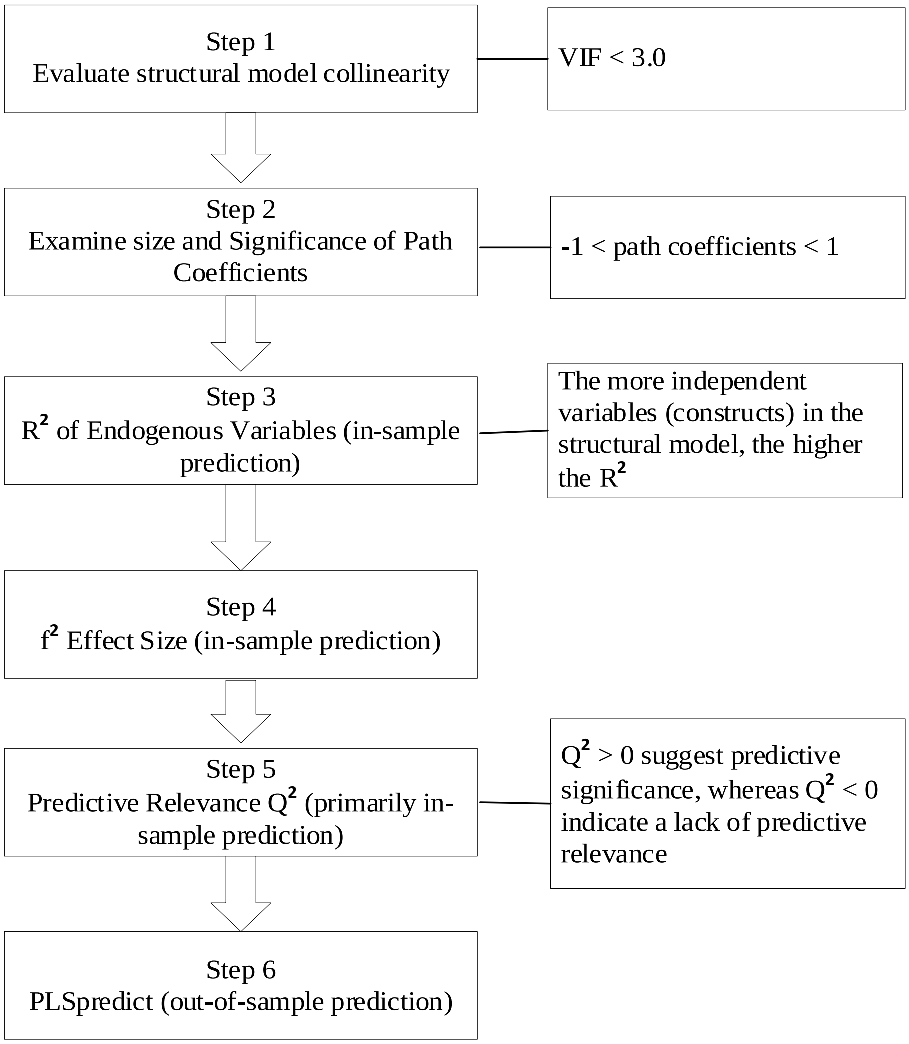

Figure 5. Structural Model

8.2. Step 2: Examine size and Significance of Path Coefficients

If multicollinearity is not an issue, the next step is to look at the path coefficients' size and significance. The path coefficients are standardised values ranging from +1 to 1 but rarely approaching +1 or 1, especially for complicated models when the structural model has numerous independent constructs[15]. The researchers can use this method to test the hypothesised relationships among the constructs.

The path coefficient values are weaker in predicting dependent (endogenous) constructs the closer they are to 0. The stronger the dependent constructs, the closer they are to the absolute value of 1[15]. The researchers should test the structural model's prediction ability as the final step. The four metrics to analyse structural model prediction are outlined in steps 3 through 6 of Figure 5.

In Table 8, the path coefficient for Cloud Benefits is the highest at 0.383, followed by Reputation at 0.236, Data Security at 0.226, SLA at 0.201, Risk Management at 0.055, and finally, Data Privacy at -0.103. Path coefficient shows that Cloud Benefits are the strongest in predicting Service Satisfaction, followed by Reputation, Data Security, SLA, Risk Management, and Data Privacy. Data Privacy is the weakest in predicting Service Satisfaction. Service Satisfaction is strong in predicting Trust as the path coefficient of 0.896 is near +1.0.

Table 8: Significance of Path Coefficients

Construct	**Service Satisfaction**	**Trust**
Cloud Benefits	0.383	
Data Privacy	-0.103	
Data Security	0.226	
Reputation	0.236	
Risk Management	0.055	
SLA	0.201	
Service Satisfaction		0.896
Trust		

- The hypothesised path relationship between Cloud Benefits and Service Satisfaction is statistically significant.
- The hypothesised path relationship between Data Security and Service Satisfaction is statistically significant.
- The hypothesised path relationship between Reputation and Service Satisfaction is statistically significant.
- The hypothesised path relationship between SLA and Service Satisfaction is statistically significant.
- The hypothesised path relationship between Service Satisfaction and Trust is statistically significant.

However, the hypothesised path relationship between Risk Management and Service Satisfaction is statistically insignificant. The same applies to Data Privacy, where the hypothesised path relationship between Data Privacy and Service Satisfaction is statistically insignificant.

This is because the standardised path coefficients for Risk Management is 0.055 and for Data Privacy is –0.103, and both are significantly lower than 0.1. So, Risk Management and Data Privacy are poor predictors of Service Satisfaction and must be deleted. On the other hand, Cloud Benefits, Reputation, Data Security, and SLA are moderately strong predictors of Service Satisfaction, while Service Satisfaction is a significant predictor of Trust.

For path coefficient, in order to claim that the path is significant, weight of impact > 0.20, t-value > 1.96 and P value < 0.05. Path is insignificant if t value < 1.96 [15].

8.3. Step 3: R^2 of Endogenous Variables (in-sample prediction)

Data Privacy, Data Security, Cloud Benefits, Reputation, SLA, and Risk Management account for 73.8% of Service Satisfaction. Data Security, Cloud Benefits, Reputation, and Service Level Agreements (SLA) are significant, but Data Privacy and Risk Management are insignificant. Service Satisfaction is responsible for 80.5% of Trust.

Table 9: The coefficient of determination, R^2

	R^2
Service Satisfaction	0.739
Trust	0.803

The coefficient of determination, R^2, is 0.803 for the Trust endogenous latent variable. The latent variables (Service Satisfaction) moderately explain 80.3% of the Variance in Trust.

R^2 is the most commonly used metric to assess structural model prediction in multiple regression models. R^2 is the coefficient of determination used to evaluate all endogenous constructs' in-sample prediction (Trust). The prediction is merely a measure of the predictive ability for the sample of data used in the calculations, and R^2 should not be inferred to the entire population [14]. R^2 is set to a minimum of 0.

The more independent variables (constructs) in the structural model, the higher the R^2, the independent variables are related to the dependent variable constructs (Trust). Table 9 displays R^2 for Self-Satisfaction equal 0.739 and R^2 for Trust equal 0.803. For the Trust endogenous latent variable, the coefficient of determination, R^2, is 0.803. This suggests that the latent variables (Service Satisfaction) account for 80.3% of the Variance in Trust. Data Privacy explains 73.9%of the Variance in Service Satisfaction, Data Security, Cloud Benefits, Reputation, SLA, and Risk Management.

8.4. Step 4: f^2 Effect Size (in-sample prediction)

R^2, f^2, and Q^2 predictive validity are useful in evaluating the predictive strength of a model based on in-sample data [15] In-sample prediction estimates the model. It predicts responses using the same sample, likely to exaggerate the model's predictive power. This is known as an overfitting problem (a greater forecast than is reasonable). It implies that the model may have limited value in predicting observations outside the original sample. When utilising PLS-SEM, [19] provided a method for assessing out-of-sample prediction. The technique entails first estimating the model on a training (analytical) sample and then using the model's outputs to predict other data in a separate holdout sample.

Table 10: f^2 effect size

Construct	**Service Satisfaction**	**Trust**
Cloud Benefits	**0.210**	
Data Privacy	0.024	
Data Security	0.111	
Reputation	0.091	
Risk Management	**0.006**	
SLA	0.073	
Service Satisfaction		**4.086**
Trust		

8.5. Step 5: Predictive Relevance Q^2 (primarily in-sample prediction)

Prediction is also assessed via the Q^2 value (blindfolding) [20]. When interpreting Q^2, numbers greater than zero suggest predictive significance, whereas values less than zero indicate a lack of predictive relevance. Furthermore, Q^2 values greater than 0.25 and 0.50 show the PLS-SEM model's medium and sizeable predictive relevance, respectively.

Table 11 shows Q^2 for Service Satisfaction as 0.462 and Q^2 for Trust as 0.438. So we can conclude that both Service Satisfaction and Trust represent a medium predictive because both of them is higher than 0.2 but lesser than 0.5.

Table 11: Q^2

	Q^2
Service Satisfaction	0.462
Trust	0.438

8.6. Step 6: PLSpredict (out-of-sample prediction)

The RMSE should be used as the prediction statistic in most cases. However, the MAE should be used if the prediction error distribution is extremely non-symmetrical [19]. The RMSE values are compared to a naive value derived by a linear regression model to measure the prediction error of a PLS-SEM analysis (LM). This is done to make predictions for the measured variables (indicators). In the PLS path model, the LM process uses a linear regression model to predict each endogenous construct's indicators from all exogenous latent variable indicators. However, the stated model structure represented by the measurement and structural theory is not included in the LM process[19]. Depending on the symmetry of the prediction error distribution, the RMSE and MAE values are both acceptable prediction benchmarks.

- The model lacks predictive power when the RMSE or MAE has higher prediction errors than the naive LM benchmark for all dependent variable indicators.
- The model has low predictive power when the dependent construct indicators have higher prediction errors than the naive LM benchmark.
- The model has medium predictive power when an equal or minor number of dependent construct indicators have higher prediction errors than the naive LM benchmark.
- The model has high predictive power when none of the dependent construct indicators has higher RMSE or MAE prediction errors than the naive LM benchmark.

Table 12: PLSPredict

	RMSE (PLS)	RMSE (LM)	MAE (PLS)	MAE (LM)
SS2	0.593	0.626	0.441	0.443
SS3	0.595	0.642	0.404	0.442
SS5	0.665	0.718	0.476	0.513
SS6	0.876	0.991	0.646	0.665
SS1	0.651	0.681	0.453	0.444
SS4	0.710	0.745	0.477	0.493
T6	0.653	0.660	0.431	0.416
T1	0.669	0.651	0.461	0.451

T4	0.811	0.843	0.629	0.637
T3	0.702	0.738	0.538	0.554
T2	0.753	0.811	0.557	0.575
T5	0.661	0.733	0.445	0.457

Table 12 indicates that the RMSE or MAE for PLS has lower prediction errors than the naive LM benchmark for all dependent variable indicators except T1 and T6, meaning that the model has predictive power. Compared to the naive LM benchmark, T1 and T6 have MAE for PLS, and T1 has RMSE with more significant prediction errors.

The model has medium to high predictive power because most dependent construct indicators have lower prediction errors than the naive LM benchmark. From the quantitative data analysis above, we can conclude that Quantitative Data Analysis help the researchers to validate the research framework.

9. RESULT AND DISCUSSION

First, the descriptive statistics analysis helps the researchers investigate the extent of big data usage in a cloud computing environment. Most of them use big data extensively, yet they do not fully utilise cloud computing services since many prefer hybrid cloud compared to the public cloud. The findings from descriptive statistics analysis can help the researchers answer some of the research questions.

Second, Figure 6 is produced after considering the reliability and validity of the constructs. Some indicators are removed (DS3, RM1 and RM5) as the factor loading is less than 0.60. Thus, the measurement model helps assess the constructs' reliability and validity.

Finally, the structural model helps to assess the relationship between constructs. Following the guidelines from Step1 to Step 6 of the structural model in Figure 5, the researchers can conclude which relationship of constructs is significant and insignificant. So, it is found that Data Security, Cloud Benefits, Reputation, and SLA are significant while Data Privacy and Risk Management are insignificant. Therefore, the findings validate the research framework.

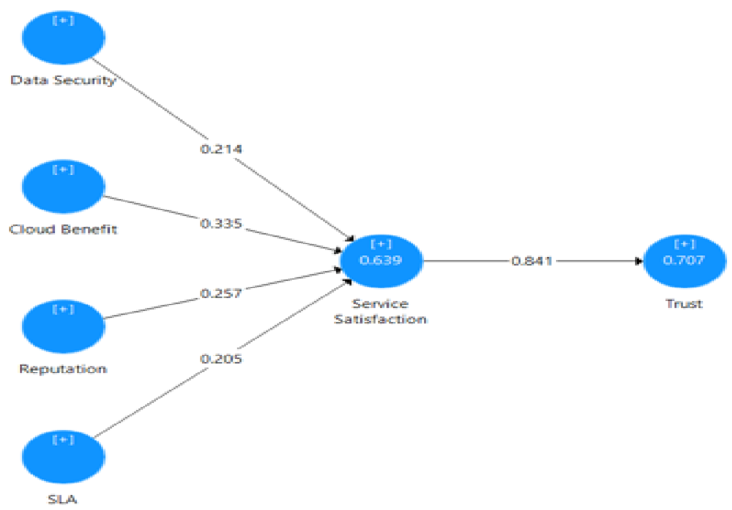

Figure 6: Revised Model

Finally, structural model helps to assess the relationship between constructs. Following the guidelines from Step1 to Step 6 of the structural model in Figure 5, the researchers can conclude which relationship of constructs is significant and insignificant. So, it is found that Data Security, Cloud Benefits, Reputation, and SLA are significant while Data Privacy and Risk Management are insignificant. Therefore, the findings validate the research framework.

Table 13: Hypothesis testing

H	Hypothesis	Supported/Not supported
H1	There is a significant positive relationship between data privacy and service satisfaction.	Not supported
H2	There is a significant positive relationship between data security and service satisfaction.	Supported
H3	There is a significant positive relationship between cloud benefits and service satisfaction.	Supported
H4	There is a significant positive relationship between reputation and service satisfaction.	Supported
H5	There is a significant positive relationship between service level agreement and service satisfaction.	Supported
H6	There is a significant positive relationship between risk management and service satisfaction.	Not supported
H7	There is a significant positive relationship between service satisfaction and Trust.	Supported

10. FUTURE WORK

The researchers recommend for the sampling be done thoroughly and accurately. Due to the differences in job scope in managing IT in the organisation, many IT managers perceived cloud usage from various angles. Some perceived the cloud as a place for big data storage, some perceived the cloud for data analytic purposes, while others may see the cloud as a platform to utilise. SaaS, PaaS and IaaS. So, purposive sampling should be done accurately and adequately. The sample needs to be screened thoroughly before they are being selected for the interview questionnaire. They must be the right candidate with the right position of IT in managing cloud usage. The screening is to ensure; they answer the questionnaire accurately and are able to represent the population. Selecting the right candidate can be difficult, but it does help in getting the most relevant constructs to represent the population.

11. CONCLUSIONS

This research is highly needed to benefit the cloud without any hesitation and unnecessary worries. Cloud users need guidelines on the criteria required before they decide to use the cloud for their big data processing, and to follow the guideline make them ready for all the uncertainties the cloud might have.

In conclusion, the research contributes to knowledge as it can give organisations guidelines on how to put their big data in the cloud with less worry. The benefits of the cloud far exceed the fear that they have. Because of the overwhelming benefits, cloud users transfer their anxiety on Data Security to the cloud providers using Reputation and SLA. The cloud providers must have an excellent image so that cloud users can trust them. Besides Reputation, the Trust can be strengthened by having SLA with them. SLA enables the cloud users to transfer the responsibility of Data Security to the cloud providers.

ACKNOWLEDGEMENTS

This research has been funded by Fundamental Research Grant Scheme (FRGS16-022-0521) supported by Ministry of Higher Education, Malaysia.

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.
[16] J. Risher, "When to use and how to report the results of PLS-SEM," no. December, 2018.
[17] J. F. Hair, M. C. Howard, and C. Nitzl, "Assessing measurement model quality in PLS-SEM using confirmatory composite analysis," *J. Bus. Res.*, vol. 109, no. December 2019, pp. 101–110, 2020.
[18] M. Sarstedt, C. M. Ringle, and J. F. Hair, *Handbook of Market Research*, no. September. 2020.
[19] P. N. Sharma, M. Sarstedt, G. Shmueli, K. H. Kim, and K. O. Thiele, "PLS-Based Model Selection: The Role of Alternative Explanations in Information Systems Research," *J. Assoc. Inf. Syst.*, no. April, pp. 346–397, 2019.
[20] F. Ali, S. M. Rasoolimanesh, and C. Cobanoglu, "Applying Partial Least Squares in Tourism and Hospitality Research," *Appl. Partial Least Squares Tour. Hosp. Res.*, no. September, pp. 1–264, 2018.

AUTHORS

The Author is a researcher and also a lecturer at UNIKL Business School Kampung Baru. She is a PhD candidate in Kulliyyah of Information and Communication Technology (KICT), IIUM and in the process to submit her thesis.

The Co-Author is currently an Associate Professor at the Department of Information Systems, Kulliyyah of Information and Communication Technology (KICT), IIUM.